THÉORIE DE LA MESURE ET INTÉGRATION

FRANÇOIS CHAPON

Université de Toulouse

2024–2025

Table des matières

1. Introduction

L'idée intuitive de l'intégrale d'une fonction continue f définie sur un segment $[a,b]$ est l'aire sous la courbe de f.

L'approche de l'intégrale de Riemann consiste à subdiviser l'intervalle de départ $[a,b]$ en petits sous-intervalles, et d'approximer f par une fonction en escalier, c'est-à-dire une fonction constante sur ces sous-intervalles. On peut choisir pour cela sur chacun de ces sous-intervalles sa borne inf ou sa borne sup, et on obtient ainsi les sommes de Darboux inférieure et supérieur :

$$I^-(f,\sigma) = \sum_{k=1}^{n} (x_k - x_{k-1}) \inf_{[x_{k-1},x_k]} f$$
$$I^+(f,\sigma) = \sum_{k=1}^{n} (x_k - x_{k-1}) \sup_{[x_{k-1},x_k]} f,$$

où σ désigne la subdivision $\{a = x_0 < x_1 < \cdots < x_n = b\}$. Ces deux sommes correspondent alors à la somme des aires des rectangles de base la longueur $x_k - x_{k-1}$ et de hauteur $\inf_{[x_{k-1},x_k]} f$ et $\sup_{[x_{k-1},x_k]} f$ respectivement (voir figure 1).

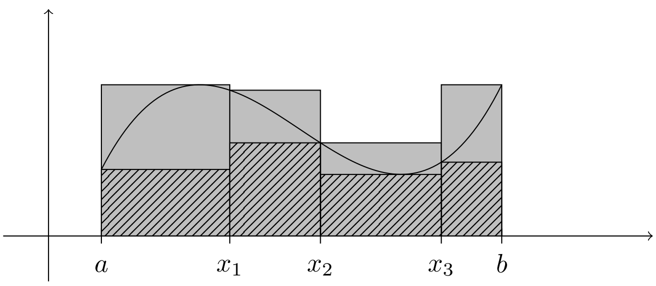

Figure 1. Intégrale de Riemann : sommes de Darboux inférieure (hachurée) et supérieure (pleine)

En prenant des subdivisions de plus en plus fines, on approche au mieux l'aire sous la courbe de f. Formellement, on définit

$$I^-(f) = \sup\{I^-(f,\sigma) \,|\, \sigma \text{ subdivision de } [a,b]\}$$
$$I^+(f) = \inf\{I^+(f,\sigma) \,|\, \sigma \text{ subdivision de } [a,b]\},$$

et on dit que f est intégrable au sens de Riemann si $I^-(f) = I^+(f)$, et l'on note $\int_a^b f(x)dx$ cette quantité. L'approche de Riemann est relativement simple et intuitive, mais assez limitée. On se contente d'intégrer des fonctions pas trop méchantes, par exemple continue ou encore continue par morceaux (et plus généralement toute limite uniforme de fonctions en escaliers) et seulement sur un intervalle borné. Un exemple de fonction qui n'est pas Riemann-intégrable est donné par la fonction indicatrice des rationnels sur $[0,1]$: $f = \mathbb{1}_{\mathbb{Q}\cap[0,1]}$, où

$$\mathbb{1}_{\mathbb{Q}\cap[0,1]}(x) = \begin{cases} 1 & \text{si } x \in \mathbb{Q}\cap[0,1], \\ 0 & \text{sinon.} \end{cases}$$

En effet, soit $\sigma = \{0 = x_0 < x_1 < \cdots < x_n = 1\}$ une subdivision quelconque de $[0,1]$, et considérons les sommes de Darboux :

$$I^-(f,\sigma) = \sum_{k=1}^n (x_k - x_{k-1}) \inf_{[x_{k-1},x_k]} f$$

$$I^+(f,\sigma) = \sum_{k=1}^n (x_k - x_{k-1}) \sup_{[x_{k-1},x_k]} f.$$

Quelque soit l'intervalle $[x_{i-1}, x_i]$, on peut trouver un rationnel q_i et un irrationnel r_i dans $[x_{i-1}, x_i]$, donc $\sup_{[x_{i-1},x_i]} f = 1$ et $\inf_{[x_{i-1},x_i]} f = 0$ pour tout i, ainsi $I^+(f,\sigma) = 1$ et $I^-(f,\sigma) = 0$, et donc les sommes de Darboux ne convergent pas vers la même limite. Ainsi, f n'est pas intégrable au sens de Riemann.

Une autre limitation de la théorie de Riemann concerne les problèmes d'inversions de limite et d'intégrale, ou de somme et d'intégrale, questions qui se posent assez naturellement. Dans le cadre de la théorie de Riemann, on ne possède pas de bons théorèmes de convergence (une limite de fonctions Riemann intégrables n'est pas forcément Riemann intégrable par exemple).

Le cadre de l'intégration de Lebesgue va pallier à ces différents problèmes : on va intégrer une classe beaucoup plus large de fonctions et ceci sur des espaces bien plus généraux que des intervalles, et ainsi disposer de bons théorèmes de convergences. L'idée de Lebesgue consiste à faire essentiellement la même chose que Riemann mais non pas sur l'espace de départ de f mais sur son espace d'arrivée, c'est-à-dire qu'on va encadrer f à partir des valeurs de $f(x)$ et non des valeurs de x. On considère donc cette fois une subdivision de l'image de f en des points $y_0 < y_1 < \ldots < y_n$ et on considère les images réciproques

$$f^{-1}([y_{k-1}, y_k[) = \{x \in [a,b] \mid y_{k-1} \leq f(x) < y_k\}.$$

On approxime alors (voir figure 2) l'aire sous la courbe de f par des rectangles de hauteur les y_{k-1} et de base la « longueur » des $f^{-1}([y_{k-1}, y_k[)$. Si $f^{-1}([y_{k-1}, y_k[)$ est une union d'intervalles, il n'y a pas de problème pour définir sa longueur, mais la difficulté ici est que $f^{-1}([y_{k-1}, y_k[)$ n'a aucune raison d'être une union d'intervalles et peut même être une partie assez tordue… Il nous faut donc au préalable étendre la notion de longueur (et en dimension supérieure la notion d'aire et de volume) à des ensembles plus généraux. Et c'est là que la théorie de la mesure va entrer en jeu.

On verra que bien heureusement, pour une fonction à la fois intégrable au sens de Riemann et intégrable au sens de Lebesgue, les deux notions coïncident. Mais on aura gagner en généralité et en souplesse de manipulation sans perdre les notions vues précédemment ! Commençons par quelques rappels, essentiellement de théorie des ensembles.

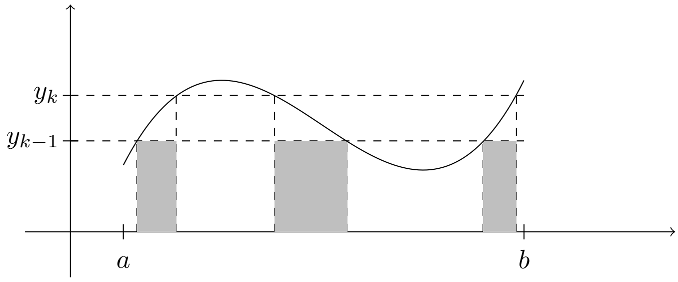

FIGURE 2. Intégrale de Lebesgue : on découpe l'espace *d'arrivée* de f pour approximer l'aire sous sa courbe

2. Rappels

2.1. **Terminologie.** Soit E un ensemble. Un sous-ensemble A de E, i.e. $A \subset E$, sera aussi appelée *une partie* de E. On note $\mathcal{P}(E)$ l'ensemble des parties de E. Un sous-ensemble $\mathcal{A}$ de $\mathcal{P}(E)$ sera appelée *une classe* (ou parfois une famille) de parties de E.

2.2. **Rappels ensembliste.** Si $(A_n)_n$ est une suite de sous-ensemble de E, croissante, au sens où $A_n \subset A_{n+1}$, pour tout n, on définit la limite de $(A_n)_n$ comme étant la réunion de tous les A_n,

$$\lim_n A_n = \bigcup_n A_n.$$

Si $(A_n)_n$ est une suite de sous-ensemble de E, décroissante, au sens où $A_{n+1} \subset A_n$, pour tout n, on définit la limite de $(A_n)_n$ comme étant l'intersection de tous les A_n,

$$\lim_n A_n = \bigcap_n A_n.$$

Pour n'importe quelle suite $(A_n)_n$ de sous-ensemble de E, on définit alors les deux parties de E suivantes :

$$\liminf_n A_n = \bigcup_n \bigcap_{k \geq n} A_k \quad \text{et} \quad \limsup_n = \bigcap_n \bigcup_{k \geq n} A_k$$

Notons que comme la suite $\left(\bigcap_{k\geq n} A_k\right)_n$ est croissante, et que la suite $\left(\bigcup_{k\geq n} A_k\right)_n$ est décroissante, les notions de liminf et limsup sont cohérentes avec les notions de limites ci-dessus. Les liminf et limsup d'une partie s'interprètent de la façon suivante (et c'est vraiment comme ça qu'il faut les comprendre) :

$$\begin{aligned} x \in \liminf_n A_n &\Leftrightarrow \exists n, \forall k \geq n,\ x \in A_k \\ &\Leftrightarrow \text{à partir d'un certain rang, } x \text{ est dans tous les } A_n, \end{aligned}$$

et

$$\begin{aligned} x \in \limsup_n A_n &\Leftrightarrow \forall n, \exists k \geq n,\ x \in A_k \\ &\Leftrightarrow x \text{ appartient à une infinité de } A_n. \end{aligned}$$

On a bien entendu que $\liminf_n A_n \subset \limsup_n A_n$.

Proposition 2.1. *Soit $f\colon E \to F$ une fonction.*

Pour tout $A \in E$, on a

$$A \subseteq f^{-1}(f(A)),$$

avec égalité si et seulement si f est injective.

Pour tout $B \in F$, on a

$$f(f^{-1}(B)) \subseteq B,$$

avec égalité si et seulement si f est surjective. (En général, on a seulement $f(f^{-1}(B)) = B \cap \mathrm{Im}(f)$).

Démonstration. Soit $x \in A$. Par définition,

$$f^{-1}(f(A)) = \{x \in E | f(x) \in f(A)\},$$

c'est-à-dire l'ensemble des antécédents de l'image de A. Donc clairement, $A \subset f^{-1}(f(A))$. Si f est injective, soit $x \in f^{-1}(f(A))$. Alors $f(x) \in f(A)$, donc par définition de $f(A)$, il existe $y \in A$ tel que $f(x) = f(y)$. Comme f est injective, $x = y$, et donc $x \in A$. Réciproquement, supposons que pour tout A, $A = f^{-1}(f(A))$. En particulier, $x = f^{-1}(f(x))$. Soit alors $x, y \in E$, tel que $f(x) = f(y)$. Alors, $f^{-1}(f(x)) = f^{-1}(f(y))$, et donc $x = y$, i.e. f est injective.

Par définition,

$$f(f^{-1}(B)) = \{f(x) | x \in f^{-1}(B)\} = \{f(x) | f(x) \in B\},$$

c'est-à-dire l'image des antécédents de B. Donc clairement, $f(f^{-1}(B)) \subset B$. Si f est surjective, soit $y \in B$. Alors, il existe $x \in E$, tel que $y = f(x)$, et donc $x \in f^{-1}(B)$. Par suite, $y \in f(f^{-1}(B))$. Réciproquement supposons que pour tout B, $f(f^{-1}(B)) = B$. En particulier, pour $B = \{y\}$, on a donc $y \in f(f^{-1}(y))$ non vide, et donc il existe x, tel que $y = f(x)$, c'est-à-dire f surjective. □

Proposition 2.2 (Formules de Hausdorff)**.** *Soit $f\colon E \to F$ une fonction. Soit $(A_i)_{i\in I}$ une famille (quelconque) de sous-ensembles de E et $(B_j)_{j\in J}$ une famille (quelconque) de sous-ensembles de F. Alors,*

$$f^{-1}\left(\bigcup_{j\in J} B_j\right) = \bigcup_{j\in J} f^{-1}(B_j)$$

$$f^{-1}\left(\bigcap_{j\in J} B_j\right) = \bigcap_{j\in J} f^{-1}(B_j)$$

$$f\left(\bigcup_{i\in I} A_i\right) = \bigcup_{i\in I} f(A_i)$$

$$f\left(\bigcap_{i\in I} A_i\right) \subseteq \bigcap_{i\in I} f(A_i),$$

avec égalité si f est injective. Si $B \in F$, alors

$$^c\left(f^{-1}(B)\right) = f^{-1}(^cB).$$

Démonstration. Exo ! □

Remarque 2.1. L'image directe n'a aucune raison de bien se comporter vis à vis du complémentaire (exo !).

2.3. **Cardinaux.** Le cardinal d'un ensemble fini E, que l'on note $\mathrm{Card}(E)$, désigne son nombre d'éléments, e.g. si $E = \{1, \ldots, n\}$, $\mathrm{Card}(E) = n$. On généralise cette notion aux ensembles infinis de la sorte. (Remarque : un ensemble E est infini si il existe $x \in E$ et une injection de E dans $E \setminus \{x\}$. Par exemple pour $\mathbb{N}$, l'application $\mathbb{N} \to \mathbb{N} \setminus \{0\}$, $n \mapsto n+1$ est une injection).

Définition 2.1. *Deux ensembles E et F sont dits équipotents, ou encore de même cardinal, ou encore de même puissance, s'il existe une bijection de l'un sur l'autre. On note alors* $\mathrm{Card}(E) = \mathrm{Card}(F)$.

On notera $\mathrm{Card}(E) \leq \mathrm{Card}(F)$ *s'il existe une injection de E dans F, ou de façon équivalente s'il existe une bijection de E sur une partie de F.*

Théorème 2.1 (Théorème de Cantor-Bernstein)**.** *Si* $\mathrm{Card}(E) \leq \mathrm{Card}(F)$ *et* $\mathrm{Card}(F) \leq \mathrm{Card}(E)$*, alors* $\mathrm{Card}(E) = \mathrm{Card}(F)$*. Autrement dit, s'il existe une injection de E dans F et une injection de F dans E, alors il existe une bijection entre E et F.*

Démonstration. Admis. $\square$

Définition 2.2. *Un ensemble E est dit dénombrable s'il s'injecte dans* $\mathbb{N}$*. Il est alors soit fini, soit infini.*

Un ensemble E est dit non-dénombrable si $\mathrm{Card}(E) > \mathrm{Card}(\mathbb{N})$ *(donc si il n'existe pas d'injection de E dans* $\mathbb{N}$*).*

Exemple 2.1.
- Les ensembles $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{N}^p$ pour tout $p \geq 1$ sont infinis dénombrables.
- $\mathbb{R}$ est non-dénombrable : voir plus bas.
- $\mathcal{P}(E)$ et $\{0,1\}^E$ sont équipotents : en effet l'application qui à $A \in \mathcal{P}(E)$ associe la fonction indicatrice $\mathbb{1}_A$ est une bijection. (exo !)
- l'intervalle $]0,1[$ est équipotent à $\mathbb{R}$. (exo !)
- l'intervalle $[0,1]$ est équipotent à $\mathbb{R}$. On remarquera qu'une bijection de $[0,1]$ dans $\mathbb{R}$ ne peut pas être continue (l'image d'un compact par une application continue est compact).

2.4. **Quelques résultats.** On commence par une propriété que l'on utilisera souvent :

Proposition 2.3. *Toute réunion dénombrable d'ensembles dénombrables est dénombrable.*

Démonstration. Soit $E = \bigcup_{n\geq 0} E_n$, où les ensembles E_n sont dénombrables. Ainsi, pour tout $n \geq 0$, il existe une injection $\varphi_n \colon E_n \to \mathbb{N}$. Soit $x \in E$, et définissons :

$$N(x) = \min\{n \geq 0 \,|\, x \in E_n\} \quad \text{(qui est bien fini).}$$

L'application $\varphi \colon E \to \mathbb{N} \times \mathbb{N}$ définie par

$$\varphi(x) = (N(x), \varphi_{N(x)}(x)),$$

est alors injective, et comme $\mathbb{N} \times \mathbb{N}$ est dénombrable, E est bien dénombrable. $\square$

Proposition 2.4. *Tout produit cartésien fini d'ensembles dénombrables est dénombrable.*

Démonstration. Exo ! $\square$

Proposition 2.5. *Soit X un ensemble. Alors* $\mathrm{Card}(\mathcal{P}(X)) > \mathrm{Card}(X)$.

Démonstration. L'ensemble X s'injecte bien évidemment dans $\mathcal{P}(X)$, il s'agit donc de montrer qu'il n'existe pas de surjection de X dans $\mathcal{P}(X)$. Soit $\Phi \colon X \to \mathcal{P}(X)$, et supposons que Φ est surjective. Soit $A \in \mathcal{P}(X)$ défini par

$$A = \{x \in X \,|\, x \notin \Phi(x)\},$$

et soit x_0 un antécédent de A, i.e. $\Phi(x_0) = A$. Alors :
- soit $x_0 \in A$, et par définition de A, $x_0 \notin \Phi(x_0)$, donc, comme $\Phi(x_0) = A$, $x \notin A$, ce qui est impossible
- soit $x_0 \notin A$, et comme $\Phi(x_0) = A$, $x_0 \notin \Phi(x_0)$, donc $x_0 \in A$, ce qui n'est pas possible non plus.

L'application Φ ne peut donc pas être surjective. $\square$

Remarque 2.2. Soit X un ensemble fini de cardinal n. Alors $\mathrm{Card}(\mathcal{P}(X)) = 2^n$ (exo !)

Proposition 2.6. *Tout produit cartésien infini dénombrable d'ensembles contenant au moins deux éléments est non-dénombrable.*

Démonstration. Soit $E_i, i \geq 0$, des ensembles tel que $\mathrm{Card}(E_i) \geq 2$. Soit alors $\varphi_i \colon \{0,1\} \to E_i$ une injection. L'application

$$\begin{array}{rccc} \varphi\colon & \{0,1\}^{\mathbb{N}} & \longrightarrow & \prod_{i\geq 0} E_i \\ & (x_1, x_2, \dots) & \longmapsto & (\varphi_1(x_1), \varphi_2(x_2), \dots) \end{array}$$

est alors injective, et donc $\mathrm{Card}\left(\prod_{i\geq 0} E_i\right) \geq \mathrm{Card}\left(\{0,1\}^{\mathbb{N}}\right) = \mathrm{Card}\,\mathcal{P}(\mathbb{N}) > \mathrm{Card}(\mathbb{N})$, ce qui montre que $\prod_{i\geq 0} E_i$ est non-dénombrable. □

Théorème 2.2. *L'ensemble des réels* $\mathbb{R}$ *est non dénombrable, et on a* $\mathrm{Card}(\mathbb{R}) = \mathrm{Card}(\mathcal{P}(\mathbb{N}))$.

Démonstration. Voir TD. □

Définition 2.3. *Un ensemble équipotent à* $\mathbb{R}$ *est dit avoir la puissance du continu.*

2.5. **La droite réelle achevée.** La droite réelle achevée, notée usuellement $\overline{\mathbb{R}}$, désigne l'ensemble des nombres réels auxquels sont adjoints deux éléments supplémentaires : un plus grand élément, $+\infty$, et un plus petit élément, $-\infty$. On la note aussi $[-\infty, +\infty]$. On définit un ordre total sur $\overline{\mathbb{R}}$ en prolongeant celui de $\mathbb{R}$, on a donc :

$$-\infty \leq x \leq +\infty,$$

pour tout $x \in \overline{\mathbb{R}}$. Les opérations algébriques usuelles sont étendues de façon évidente, en particulier

$$\infty - \infty \quad \text{et} \quad \frac{\infty}{\infty}$$

ne sont pas définies. On prendra par contre souvent comme convention (assez dangereuse, il faut bien l'avouer) que $0 \times \infty = 0$. La droite achevée est un espace métrique muni (par exemple) de la distance suivante : pour tous $x, y \in \overline{\mathbb{R}}$,

$$\delta(x, y) = |f(x) - f(y)|,$$

où f est l'homéomorphisme de $\mathbb{R}$ à $]-1, 1[$ définie par

$$f(x) = \frac{x}{\sqrt{x^2 + 1}},$$

prolongée à $\overline{\mathbb{R}}$ par $f(\pm\infty) = \pm 1$. Ainsi, $\overline{\mathbb{R}}$ est un espace compact, et la topologie induite par δ sur $\mathbb{R}$ coïncide avec la topologie usuelle de $\mathbb{R}$. Une base d'ouverts de $\overline{\mathbb{R}}$ est composée des intervalles de la forme $]a, +\infty]$, ou $[-\infty, b[$ ou $]a, b[$. Notons qu'il n'y a aucun moyen canonique de prolonger la distance usuelle de $\mathbb{R}$ à $\overline{\mathbb{R}}$.

3. Tribus

La notion de tribu sur un ensemble, va nous permettre de définir quels sont les ensembles que l'on peut « mesurer ».

3.1. Définition et exemples.

Définition 3.1. *Soit E un ensemble et $\mathcal{A} \subset \mathcal{P}(E)$ une classe de parties de E. On dit que $\mathcal{A}$ est une tribu, ou une σ-algèbre, si*

(i) elle contient E : $E \in \mathcal{A}$;

(ii) elle est stable par passage au complémentaire : pour tout $A \in \mathcal{A}$, ${}^cA \in \mathcal{A}$;

(iii) elle est stable par réunion dénombrable : pour toute famille dénombrable $(A_n)_{n\geq 0}$ d'éléments de $\mathcal{A}$, $\displaystyle\bigcup_{n\geq 0} A_n \in \mathcal{A}$.

L'espace $(E, \mathcal{A})$ est appelé un espace mesurable, et les éléments de $\mathcal{A}$ sont dit des ensembles mesurables.

Quelques conséquences immédiates de la définition de tribu :

(i) $\varnothing \in \mathcal{A}$ car $\varnothing = {}^cE$;

(ii) elle est stable par intersection dénombrable car $\bigcap_n A_n = {}^c\left(\bigcup_n {}^cA_n\right)$;

(iii) elle est stable par différence car $A \setminus B = A \cap {}^cB$;

(iv) elle est stable par différence symétrique car $A \Delta B = (A \setminus B) \cup (B \setminus A)$;

(v) elle est stable par limite inférieure car $\liminf_n A_n = \bigcup_n \bigcap_{k\geq n} A_k$;

(vi) elle est stable par limite supérieure car $\liminf_n A_n = \bigcap_n \bigcup_{k\geq n} A_k$.

Exemple 3.1. Quelques exemples de tribus :

— $\{\varnothing, E\}$, appelée tribu grossière ;

— $\mathcal{P}(E)$, appelée tribu triviale ou tribu discrète ;

— si $A \subset E$, $\{\varnothing, A, {}^cA, E\}$ est la plus petite tribu contenant A ;

— si $(A_n)_{n\geq 0}$ est une partition dénombrable de E, alors

$$\left\{ \bigcup_{i\in I} A_i \,\middle|\, I \subset \mathbb{N} \right\}$$

est une tribu (exo !). En particulier, si $(A_i)_{i=1,\ldots,n}$ est une partition finie de E, alors $\{ \cup_{i\in I} A_i \mid I \subset \{1, \ldots, n\} \}$ est une tribu de cardinal 2^n.

3.2. Tribu engendrée.

Proposition 3.1. *Soit $\mathcal{C}$ une classe de parties d'un ensemble E. L'intersection de toutes les tribus contenant $\mathcal{C}$ est une tribu, appelée tribu engendrée par $\mathcal{C}$, et notée $\sigma(\mathcal{C})$:*

$$\sigma(\mathcal{C}) = \bigcap_{\mathcal{A} \text{ tribu}, \mathcal{C}\subset\mathcal{A}} \mathcal{A}.$$

C'est la plus petite tribu contenant $\mathcal{C}$.

Démonstration. Une intersection (quelconque, au sens pas forcément dénombrable) de tribus est en effet une tribu. On a clairement $\varnothing \in \sigma(\mathcal{C})$, car $\varnothing \in \mathcal{A}$ pour toute tribu $\mathcal{A}$. Montrons la stabilité par passage au complémentaire. Soit $A \in \sigma(\mathcal{C})$. Alors $A \in \mathcal{A}$ pour toute tribu $\mathcal{A}$ contenant $\mathcal{C}$, donc ${}^cA \in \mathcal{A}$ pour toute $\mathcal{A}$, et ${}^cA$ est bien dans $\sigma(\mathcal{C})$. Montrons la stabilité par union dénombrable. Soit $(A_n)_n$ une suite d'éléments de $\sigma(\mathcal{C})$. Alors pour tout n, $A_n \in \mathcal{A}$, pour toute tribu $\mathcal{A}$ contenant $\mathcal{C}$. Donc $\bigcup_n A_n \in \mathcal{A}$ pour toute $\mathcal{A}$, et $\bigcup_n A_n \in \sigma(\mathcal{C})$. Finalement, $\sigma(\mathcal{C})$ contient clairement $\mathcal{C}$, et est bien la plus petite (au sens de l'inclusion) tribu contenant $\mathcal{C}$, car toute tribu contenant $\mathcal{C}$ contient $\sigma(\mathcal{C})$. □

Exemple 3.2. Soit E un ensemble et $\mathcal{C}$ la classe des singletons de E. Alors,

$$\sigma(\mathcal{C}) = \{A \in \mathcal{P}(E) \mid A \text{ dénombrable ou } {}^cA \text{ dénombrable}\}.$$

Pour décrire $\sigma(\mathcal{C})$, on peut raisonner de la sorte. Elle contient la classe $\mathcal{C}$, c'est-à-dire tous les singletons de E. Elle doit être stable par union dénombrable, donc elle contient les unions dénombrables de singletons, donc les parties dénombrables. Elle doit être stable par passage au complémentaire, donc elle contient les complémentaires de parties dénombrables. On a donc un candidat naturel pour $\sigma(\mathcal{C})$:

$$\{A \in \mathcal{P}(E) \mid A \text{ dénombrable ou } {}^cA \text{ dénombrable}\}.$$

Appelons $\mathcal{B}$ cette classe, et montrons que $\mathcal{B}$ est une tribu. Comme $\mathcal{B}$ contient $\mathcal{C}$ et est contenu dans $\sigma(\mathcal{C})$ par construction, et que $\sigma(\mathcal{C})$ est la plus petite tribu contenant $\mathcal{C}$, on aura forcément $\sigma(\mathcal{C}) = \mathcal{B}$. On a $\varnothing \in \mathcal{B}$, car $\varnothing$ est dénombrable. Clairement $\mathcal{B}$ est stable par passage au complémentaire. Soit $(A_n)_n$ une suite de $\mathcal{B}$. Si pour tout n, A_n est dénombrable, alors $\bigcup_n A_n$ est dénombrable comme union dénombrable de parties dénombrables, donc $\bigcup_n A_n \in \mathcal{B}$. Si il existe n_0 tel que ${}^cA_{n_0}$ soit dénombrable, alors

$${}^c\left(\bigcup_n A_n\right) = \bigcap_n {}^cA_n \subset {}^cA_{n_0}$$

qui est dénombrable, donc ${}^c\left(\bigcup_n A_n\right)$ aussi, et $\bigcup_n A_n \in \mathcal{B}$. On a bien montré que $\mathcal{B}$ est une tribu, et donc $\mathcal{B} = \sigma(\mathcal{C})$.

Remarque : Cette méthode ne marche malheureusement pas à tous les coups, et il est en général très difficile de décrire explicitement une tribu...

Remarque 3.1. Si $\mathcal{C}_1 \subset \mathcal{C}_2$, on a $\sigma(\mathcal{C}_1) \subset \sigma(\mathcal{C}_2)$. En effet, si $\mathcal{C}_1 \subset \mathcal{C}_2$, comme $\mathcal{C}_2 \subset \sigma(\mathcal{C}_2)$, on a $\mathcal{C}_1 \subset \sigma(\mathcal{C}_2)$, et comme $\sigma(\mathcal{C}_1)$ est la plus petite tribu contenant $\mathcal{C}_1$, on a $\sigma(\mathcal{C}_1) \subset \sigma(\mathcal{C}_2)$.

Remarque 3.2. La réunion de 2 tribus n'est en général pas une tribu. Par exemple, sur $\{a, b, c\}$, considérons les tribus $\mathcal{A}_1 = \{\varnothing, \{a\}, \{b, c\}, \{a, b, c\}\}$ (i.e. la tribu engendrée par $\{a\}$) et $\mathcal{A}_2 = \{\varnothing, \{b\}, \{a, c\}, \{a, b, c\}\}$ (i.e. la tribu engendrée par $\{b\}$). Alors,

$$\mathcal{A}_1 \cup \mathcal{A}_2 = \{\varnothing, \{a\}, \{b, c\}, \{b\}, \{a, c\}, \{a, b, c\}\}$$

n'est pas une tribu, il manque par exemple $\{a, b\} = \{a\} \cup \{b\}$.

3.3. **Tribu réciproque, tribu image, tribu trace.**

Notation 3.1. Soit $\mathcal{C}$ une classe de parties d'un ensemble Y, et f une fonction de X dans Y. On notera

$$f^{-1}(\mathcal{C}) = \{f^{-1}(C) \mid C \in \mathcal{C}\}$$

la classe des parties de X des images réciproques des éléments de $\mathcal{C}$.

Proposition 3.2. *Soit $f\colon X \to Y$ et $\mathcal{B}$ une tribu sur Y. Alors*

$$f^{-1}(\mathcal{B}) = \{f^{-1}(B) \mid B \in \mathcal{B}\}$$

est une tribu sur X, appelée tribu image réciproque de $\mathcal{B}$ par f. On l'appelle aussi la tribu engendrée par f, et la note aussi $\sigma(f)$.

Les propriétés de l'image réciproque (formules de Hausdorff) assure que $f^{-1}(\mathcal{B})$ est bien une tribu (exo !).

Proposition 3.3. *Soit $f\colon X \to Y$ et $\mathcal{A}$ une tribu sur X. Alors*

$$\mathcal{B} = \{B \subset Y \mid f^{-1}(B) \in \mathcal{A}\}$$

est tribu sur Y, appelée tribu image de $\mathcal{A}$ par f

Là encore, les propriétés de l'image réciproque assure que $\mathcal{B}$ est bien une tribu (exo !).

Remarque 3.3. Attention, la tribu image n'est pas $f(\mathcal{A})$ qui n'est en général pas une tribu. Par exemple, soit $X = \{a, b, c\}$ et la tribu sur X donnée par $\mathcal{A} = \{\varnothing, \{a\}, \{b, c\}, \{a, b, c\}\}$ (c'est-à-dire la tribu engendrée par $\{a\}$). Soit $Y = \{1, 2\}$ et f telle que $f(a) = f(c) = 1$ et $f(b) = 2$. Alors, $f(\mathcal{A}) = \{\varnothing, \{1\}, \{1, 2\}\}$, qui n'est pas une tribu (il manque $\{2\}$ le complémentaire de $\{1\}$).

Proposition 3.4. *Soit $(E, \mathcal{A})$ un espace mesuré. Soit $X \in \mathcal{P}(E)$ et soit $\mathcal{B}$ la classe de parties de X définie par $\mathcal{B} = \{A \cap X \mid A \in \mathcal{A}\}$. Alors, $\mathcal{B}$ est une tribu sur X, appelée tribu trace de $\mathcal{A}$ sur X.*

Démonstration. Il suffit de remarquer que $\mathcal{B}$ est la tribu image réciproque de $\mathcal{A}$ par l'injection canonique $i\colon X \hookrightarrow E$, car pour tout $A \in \mathcal{A}$, $i^{-1}(A) = A \cap X$. □

3.4. **Tribu de Borel.**

Définition 3.2. *On appelle tribu borélienne (ou tribu de Borel) sur $\mathbb{R}$ la tribu engendrée par les ouverts de $\mathbb{R}$. On la note $\mathcal{B}(\mathbb{R})$.*

Plus généralement, soit $(X, \mathcal{T})$ un espace topologique. On appelle tribu borélienne sur X, et on la note $\mathcal{B}(X)$ la tribu engendrée par les ouverts de X.

Plutôt que d'utiliser l'adjectif mesurable, on appellera les éléments de la tribu borélienne des *boréliens.*

Rappelons (voir cours de topologie), que tout ouvert de $\mathbb{R}$ s'écrit comme une réunion dénombrable d'intervalles ouverts, que l'on peut supposer disjoints. En effet, soit U un ouvert et $x \in U$. Soit $C(x)$ l'union de tous les intervalles ouverts contenu dans U et contenant x. Alors $C(x)$ est non-vide car U est ouvert et $C(x)$ est un intervalle ouvert car réunion d'intervalles ouverts, c'est même le plus grand intervalle contenant x (remarque : c'est la composante connexe de x dans U). Ansi,

$$U = \bigcup_{x \in U} C(x).$$

Soit $y \in U$. Si $y \in C(x)$, alors $C(x)$ est un intervalle ouvert contenant y, donc $C(x) \subset C(y)$. Donc $x \in C(y)$, et par suite $C(y) \subset C(x)$. Finalement, $C(x) = C(y)$.

Donc, si $C(x) \cap C(y) \neq \varnothing$, alors il existe $z \in C(x) \cap C(y)$. Donc $C(z) = C(x)$ et $C(z) = C(y)$, et donc $C(x) = C(y)$. Finalement, les $C(x)$ sont soit égaux, soit disjoints.

Par densité de $\mathbb{Q}$ dans $\mathbb{R}$, il existe un rationnel $q_x \in C(x)$ pour chaque $x \in U$, et donc, les $C(x)$ étant disjoints, l'union est dénombrable.

On a donc que la tribu borélienne $\mathcal{B}(\mathbb{R})$ est engendrée par :

— les intervalles ouverts $]a, b[$ (avec $a < b$ dans $\mathbb{R}$)

— les intervalles fermés $[a, b]$ (avec $a < b$ dans $\mathbb{R}$)

— les intervalles ouverts $]a, +\infty[$ (avec $a \in \mathbb{R}$)

— les intervalles ouverts $]-\infty, a[$ (avec $a \in \mathbb{R}$)

— les intervalles ouverts $]a, b[$ avec $a < b$ dans $\mathbb{Q}$

— les intervalles $]a, b]$, $a < b$

— ...

Détaillons la première assertion, les autres sont laissées à titre d'exercice. Notons $\mathcal{C}$ la classe des intervalles ouverts :

$$\mathcal{C} = \left\{]a, b[\mid a, b \in \overline{\mathbb{R}}, a < b \right\}.$$

On veut donc montrer que $\mathcal{B}(\mathbb{R}) = \sigma(\mathcal{C})$. On a évidemment
$$\mathcal{C} \subset \{\text{ouverts de } \mathbb{R}\},$$
donc
$$\sigma(\mathcal{C}) \subset \mathcal{B}(\mathbb{R}).$$
Par le rappel ci-dessus, tout ouvert de $\mathbb{R}$ peut s'écrire comme une réunion *dénombrable* d'intervalles ouverts, donc
$$\{\text{ouverts de } \mathbb{R}\} \subset \sigma(\mathcal{C}),$$
et comme $\mathcal{B}(\mathbb{R})$ est la plus petite tribu contenant la classe des ouverts de $\mathbb{R}$, on a
$$\mathcal{B}(\mathbb{R}) \subset \sigma(\mathcal{C}).$$
Ainsi, $\mathcal{B}(\mathbb{R}) = \sigma(\mathcal{C})$.

De même, la tribu borélienne $\mathcal{B}(\mathbb{R}^d)$ est par exemple engendrée par les pavés
$$\prod_{i=1}^{d}]a_i, b_i[, \quad \text{avec } a_i, b_i \text{ dans } \overline{\mathbb{R}}.$$

Remarque 3.4. Il est possible de montrer que $\mathcal{B}(\mathbb{R})$ est équipotente à $\mathbb{R}$ (i.e. de même cardinal). Ceci se fait par récurrence transfinie, et dépasse largement les notions abordées dans ce cours. Notons tout de même que comme $\mathrm{Card}(\mathcal{P}(\mathbb{R})) > \mathrm{Card}(\mathbb{R})$, il existe (beaucoup !) d'ensembles non-boréliens. Il n'est pourtant pas aisé d'en exhiber un… Nous reviendrons là-dessus un peu plus tard.

4. Mesures

4.1. Définition.

Définition 4.1. *Soit $(E,\mathcal{A})$ un espace mesurable. Une mesure (positive) sur $(E,\mathcal{A})$ est une application $\mu\colon \mathcal{A}\to[0,+\infty]$ telle que*

(i) $\mu(\varnothing)=0$ *;*

(ii) elle est σ-additive : pour toute suite $(A_n)_n$ d'éléments de $\mathcal{A}$ deux à deux disjoints, on a

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

On dit alors que $(E,\mathcal{A},\mu)$ est un espace mesuré.

Avant de donner quelques exemples, voici quelques propriétés immédiates que vérifie une mesure.

Proposition 4.1. *Soit $(E,\mathcal{A},\mu)$ est un espace mesuré. Alors,*

1) additivité finie : pour tout $(A_i)_{i\in I}$ mesurables, avec I fini, et disjoints deux à deux, $\mu(\bigcup_{i\in I}A_i)=\sum_{i\in I}\mu(A_i)$;

Pour tout $A,B\in\mathcal{A}$,

2) $\mu(A)=\mu(A\setminus B)+\mu(A\cap B)$ *;*

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

4) sous-additivité finie : $\mu(A\cup B)\leq\mu(A)+\mu(B)$; et plus généralement, si $A_1,\dots,A_n$ sont des ensembles mesurables, alors

$$\mu\left(\bigcup_{k=1}^n A_k\right)\leq\sum_{k=1}^n\mu(A_k);$$

5) croissance : si $A\subset B$, $\mu(A)\leq\mu(B)$.

Démonstration. 1) Découle de la σ-additivité.

2) On a $A=(A\setminus B)\cup(A\cap B)$ avec $A\setminus B$ et $A\cap B$ disjoints. Il suffit alors d'utiliser le point précédent.

3) $A\cup B=(A\setminus B)\cup(B\setminus A)\cup(A\cap B)$ et les ensembles $A\setminus B$, $B\setminus A$ et $A\cap B$ sont deux à deux disjoints. Par le point précédent,

$$\begin{aligned}\mu(A)+\mu(B)&=\mu(A\setminus B)+\mu(A\cap B)+\mu(B\setminus A)+\mu(A\cap B)\\&=\mu(A\cup B)+\mu(A\cap B).\end{aligned}$$

Remarque : on ne peut pas à priori retrancher $\mu(A\cap B)$ qui pourrait être infini.

4) Si $\mu(A)+\mu(B)=+\infty$, il n'y a rien à montrer. Si $\mu(A)+\mu(B)<+\infty$, alors par le point précédent $\mu(A\cap B)<+\infty$, donc on peut le retrancher dans l'égalité précédente, et donc

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

Par récurrence immédiate, on a $\mu(\bigcup_{i=1}^n A_i)\leq\sum_{i=1}^n\mu(A_i)$ pour tous ensembles mesurables $A_1,\dots,A_n$.

5) Si $A\subset B$, alors $A\cap B=A$, et par l'égalité du point 2),

$$\mu(B)=\mu(B\setminus A)+\mu(A)\geq\mu(A).$$ □

Outre l'exemple trivial de la mesure nulle, donnons quelques exemples de mesures :

Exemple 4.1. Soit $(X, \mathcal{A})$ un espace mesurable. On définit la *mesure de comptage* m sur $(X, \mathcal{A})$ par : pour tout ensemble mesurable $A \in \mathcal{A}$,

$$m(A) = \begin{cases} \mathrm{Card}(A) & \text{si } A \text{ est fini} \\ +\infty & \text{sinon.} \end{cases}$$

Exemple 4.2. Soit $(X, \mathcal{A})$ un espace mesurable, et $a \in X$. On définit la *mesure de Dirac* au point a, notée δ_a, sur $(X, \mathcal{A})$ par : pour tout ensemble mesurable $A \in \mathcal{A}$,

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

On a donc $\delta_a(A) = \mathbb{1}_A(a)$.

À titre d'exercice, vérifiez que les deux exemples précédents définissent bien des mesures.

Quelques définitions supplémentaires :

Définition 4.2. *Soit μ une mesure sur un espace mesurable $(E, \mathcal{A})$.*

- *On dit que μ est finie, ou de masse finie, si $\mu(E) < \infty$. On a donc que $\mu(A) < \infty$, pour tout $A \in \mathcal{A}$. Le nombre $\mu(E)$ est appelée la masse de μ.*
- *On dit que μ est une (mesure de) probabilité si sa masse vaut 1.*
- *On dit que μ est σ-finie si il existe un recouvrement dénombrable de E par des ensembles de mesures finies : il existe $(E_n)_n$ avec $E_n \in \mathcal{A}$ pour tout n, tel que $E = \bigcup_n E_n$ et $\mu(E_n) < \infty$.*

Les mesures de probabilité nous donnent alors d'autres exemples de mesure. Par exemple, la loi binomiale de paramètres n, p est la mesure μ sur l'espace $\Omega = \{0, \dots, n\}$ muni de la tribu $\mathcal{P}(\Omega)$ tel que

$$\mu(\{n\}) = \binom{n}{k} p^k (1-p)^{n-k},$$

pour tout $k = 0, \dots, n$. La théorie moderne des probabilités, axiomatisée par Kolmogorov dans les années 1930, est de fait basée sur la théorie de la mesure.

La proposition suivante donne une autre caractérisation de mesure.

Proposition 4.2. *Une application $\mu \colon \mathcal{A} \to [0, +\infty]$ est une mesure si et seulement si :*

(i) $\mu(\varnothing) = 0$

(ii) μ est finiment additive : pour tout ensembles mesurables $A_i, i \in I$ avec I fini, deux à deux disjoints, $\mu\left(\bigcup_{i \in I} A_i\right) = \sum_{i \in I} \mu(A_i)$.

(iii) μ est "continue à gauche" : pour toute suite croissante $(A_n)_{n \geq 0}$ d'ensembles mesurables, $\mu(\lim_n A_n) = \lim_n \mu(A_n)$.

Démonstration. Soit μ une mesure. La σ-additivité entraîne l'additivité finie, donc seul le dernier point reste à montrer. Soit $(A_n)_n$ une suite croissante, i.e. $A_n \subset A_{n+1}$, pour tout n. Posons $B_0 = A_0$, $B_1 = A_1 \setminus A_0$, …, $B_n = A_n \setminus A_{n-1}$, … Les ensembles B_n sont bien mesurables, et deux à deux disjoints. De plus, $\bigcup_{k=0}^n B_k = A_n$, et

$$\lim_n A_n = \bigcup_n A_n = \bigcup_n B_n.$$

Alors, par σ-additivité, on a $\mu\left(\bigcup_n B_n\right) = \sum_n \mu(B_n)$, ainsi,

$$\mu\left(\bigcup_n A_n\right) = \mu\left(\bigcup_n B_n\right) = \sum_n \mu(B_n) = \lim_n \sum_{k=0}^n \mu(B_k) = \lim_n \mu(\bigcup_{k=0}^n B_k) = \lim_n \mu(A_n).$$

Montrons la réciproque. Il s'agit donc de montrer la σ-additivité. Soit $(A_n)_n$ une suite d'ensembles mesurables deux à deux disjoints. Posons $B_n = \bigcup_{k=0}^n A_k$. Les B_n sont mesurables, et on a $B_n \subset B_{n+1}$, pour tout n. Comme $\bigcup_n A_n = \bigcup_n B_n$, on a

$$\mu\left(\bigcup_n A_n\right) = \mu\left(\bigcup_n B_n\right) = \lim \mu(B_n) = \lim_n \mu(\bigcup_{k=0}^n A_k) = \lim_n \sum_{k=0}^n \mu(A_k),$$

où la dernier égalité utilise l'hypothèse d'additivité finie. Ainsi,

$$\mu\left(\bigcup_n A_n\right) = \sum_n \mu(A_n),$$

et μ est bien une mesure sur $\mathcal{A}$. □

Corollaire 4.1. *Toute mesure μ est sous-σ-additive, c'est-à-dire que pour toute suite d'ensembles mesurables $(A_n)_n$,*

$$\mu\left(\bigcup_n A_n\right) \leq \sum_n \mu(A_n).$$

Démonstration. Soit $(A_n)_n$ mesurables, posons $B_n = \bigcup_{k=0}^n A_k$. Alors la suite $(B_n)_n$ est mesurable, croissante, et on a $\bigcup_n B_n = \bigcup_n A_n$. En appliquant la continuité à gauche de la proposition précédente, on a alors,

$$\mu\left(\bigcup_n A_n\right) = \mu\left(\bigcup_n B_n\right) = \lim_n \mu(B_n) = \lim_n \mu\left(\bigcup_{k=0}^n A_k\right).$$

Or par sous-additivité *finie*, on a $\mu\left(\bigcup_{k=0}^n A_k\right) \leq \sum_{k=0}^n \mu(A_k)$, ainsi

$$\mu\left(\bigcup_n A_n\right) \leq \lim_n \sum_{k=0}^n \mu(A_k) = \sum_n \mu(A_n).$$ □

La continuité à droite pour une mesure n'est en général pas vérifiée. Mais on a le résultat suivant :

Proposition 4.3. *Soit $(A_n)_n$ une suite décroissante d'ensembles mesurables telle qu'il existe n_0 tel que $\mu(A_{n_0}) < \infty$. Alors*

$$\lim_n \mu(A_n) = \mu(\lim_n A_n).$$

Démonstration. Comme la suite $(A_n)_n$ est décroissante, on a que pour tout $n \geq n_0$, $\mu(A_n) < \infty$, car $A_n \subset A_{n_0}$ et par croissance d'une mesure. Posons $B_n = A_{n_0} \setminus A_n$. Alors, comme $(A_n)_n$ est décroissante, la suite $(B_n)_n$ est croissante et converge vers $\bigcup_n B_n = A_{n_0} \setminus \bigcap_n A_n$, donc

$$\mu(A_{n_0}) - \mu\left(\bigcap_n A_n\right) = \mu\left(\bigcup_n B_n\right) = \lim_n \mu(B_n) = \mu(A_{n_0}) - \lim_n \mu(A_n),$$

ce qui conclut la preuve. □

Remarque. La condition finie est nécessaire, en effet considérons la mesure de comptage m sur $\mathbb{N}$, et les ensembles $A_n = \{n, n+1, \ldots\}$. Alors la suite $(A_n)_n$ est décroissante et converge vers $\bigcap_n A_n = \varnothing$, donc $m\left(\bigcap_n A_n\right) = 0$, mais $m(A_n) = +\infty$ pour tout n, et donc ne converge pas.

Proposition 4.4. *Soit μ une mesure sur $\mathcal{A}$, et $(A_n)_n$ une suite d'éléments de $\mathcal{A}$. Alors,*

$$\mu(\liminf_n A_n) \leq \liminf_n \mu(A_n).$$

Si de plus, μ est une mesure finie, alors

$$\mu(\liminf_n A_n) \leq \liminf_n \mu(A_n) \leq \limsup_n \mu(A_n) \leq \mu(\limsup_n A_n).$$

Démonstration. Rappelons que $\liminf_n A_n = \bigcup_n \bigcap_{k\geq n} A_k$. Posons $B_n = \bigcap_{k\geq n} A_k$. Alors $(B_n)_n$ est une suite croissante d'ensembles mesurables, et par continuité à gauche de μ, on a

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

Or, comme $\bigcap_{k\geq n} A_k \subset A_k$, pour tout $k \geq n$, on a $\mu(\bigcap_{k\geq n} A_k) \leq \mu(A_k)$, pour tout $k \geq n$, et donc

$$\mu(B_n) \leq \inf_{k\geq n} \mu(A_k).$$

Finalement, $\mu\left(\liminf_n A_n\right) \leq \sup_n \inf_{k\geq n} \mu(A_k) = \liminf_n \mu(A_n)$.

L'inégalité sur la $\limsup$ se montre de manière analogue par continuité à droite sous l'hypothèse que μ est une mesure finie. □

Proposition 4.5. *Soit $(\mu_n)_n$ une suite croissante de mesures, i.e. pour tout n, $\mu_n(A) \leq \mu_{n+1}(A)$ pour tout $A \in \mathcal{A}$. Alors l'application μ définie par $\mu(A) = \lim_n \mu_n(A)$ est une mesure sur $\mathcal{A}$. En particulier, toute combinaison linéaire dénombrable à coefficients positifs de mesures est une mesure.*

Démonstration. On vérifie aisément que μ est une mesure en appliquant la proposition 4.2 : l'additivité pour une famille finie d'ensembles mesurables est claire, et pour $(A_n)_n$ une suite d'éléments de $\mathcal{A}$, on a

$$\mu\left(\bigcup_k A_k\right) = \lim_{n\to\infty} \mu_n\left(\bigcup_k A_k\right) = \lim_{n\to\infty}\lim_{k\to\infty} \mu_n(A_k) = \lim_{k\to\infty}\lim_{n\to\infty} \mu_n(A_k) = \lim_{k\to\infty} \mu(A_k),$$

l'inversion des deux limites étant possible par croissance. Ainsi, μ est continue à gauche, et finalement est bien une mesure. Soit $(\alpha_n)_n$ une suite de coefficients positifs. Alors,

$$\nu_n(A) = \sum_{k=1}^{n} \alpha_k \mu_k(A)$$

définit clairement une suite croissante de mesures $(\nu_n)_n$, et sa limite $\sum_n \alpha_n \mu_n$ est alors bien une mesure. □

Exemple 4.3. Soit E un ensemble dénombrable. Alors la mesure

$$\sum_{x\in E} \delta_x$$

est la mesure de comptage sur $(E, \mathcal{P}(E))$.

Autre exemple, sur $(\mathbb{N}, \mathcal{P}(\mathbb{N}))$, la mesure

$$\sum_{n\geq 0} \frac{\lambda^n}{n!} e^{-\lambda} \delta_n$$

est la mesure de probabilité bien connue sous le nom de loi de Poisson de paramètre λ (avec $\lambda > 0$).

4.2. **Mesure de Lebesgue.** La mesure de Lebesgue est la mesure sur $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ qui donne un sens mathématique à la notion de volume en dimension quelconque (donc de longueur pour $d = 1$ et d'aire pour $d = 2$).

Théorème 4.1. *Il existe une unique mesure sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ telle que la mesure de tout intervalle $]a, b[$ soit égale à $b - a$ (avec $a < b$). On l'appelle la mesure de Lebesgue sur $\mathbb{R}$ et on la note λ.*

Plus généralement, il existe une unique mesure sur $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ ($d \geq 1$) telle que la mesure de tout pavé $\prod_{i=1}^d]a_i, b_i[$ soit égale au produit $\prod_{i=1}^d (b_i - a_i)$. On l'appelle la mesure de Lebesgue sur $\mathbb{R}^d$ et on la note λ_d ou λ si il n'y a pas d'ambiguïté sur la dimension.

La propriété suivante est fondamentale :

Propriété. (Invariance par translation)**.** *La mesure de Lebesgue sur* $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ *est invariante par translation : pour tout borélien* $A \in \mathcal{B}(\mathbb{R}^d)$ *et tout* $x \in \mathbb{R}^d$*,*

$$\lambda(A + x) = \lambda(A),$$

où $A + x = \{a + x \,|\, a \in A\}$*.*

Pour le moment, on va admettre l'existence et l'unicité de la mesure de Lebesgue. On verra plus loin comment montrer que deux mesures sont égales, et en particulier l'unicité de la mesure de Lebesgue, mais on admettra sa construction (pas évidente et assez longue).

Montrons alors quelques propriétés supplémentaires.

Le mesure de Lebesgue sur $\mathbb{R}$ ne charge pas les points, c'est-à-dire que pour tout $x \in \mathbb{R}$, $\lambda(\{x\}) = 0$ (on dit qu'elle est *diffuse*). En effet, la suite des intervalles $]x - \frac{1}{n}, x + \frac{1}{n}[$ est décroissante et converge vers

$$\bigcap_n \left]x - \frac{1}{n}, x + \frac{1}{n}\right[= \{x\}.$$

De plus, $\lambda\left(]x - \frac{1}{n}, x + \frac{1}{n}[\right) = \frac{2}{n}$, donc on peut appliquer la proposition 4.3, et on a donc

$$\lambda(\{x\}) = \lim_n \lambda\left(\left]x - \frac{1}{n}, x + \frac{1}{n}\right[\right) = \lim_n \frac{2}{n} = 0.$$

On en déduit donc que :

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

Autre conséquence, tout ensemble dénombrable est de mesure de Lebesgue nulle puisqu'un ensemble dénombrable s'écrit comme l'union dénombrable disjointe de ses singletons. En particulier, $\mathbb{N}$ et $\mathbb{Q}$ sont de mesure de Lebesgue nulle.

Autre propriété : la mesure de Lebesgue est σ-finie. En effet, on peut recouvrir $\mathbb{R}$ par

$$\mathbb{R} = \bigcup_{n \geq 1}]-n, n[,$$

et la mesure de Lebesgue de l'intervalle $]-n, n[$ est $\lambda(]-n, n[) = 2n$ qui est finie pour tout $n \geq 1$. De plus, $\lambda(\mathbb{R}) = +\infty$.

Les ensembles dénombrables ne sont pas les seules à être de mesure de Lebesgue nulle. Par exemple :

Exemple 4.4 (**Ensemble triadique de Cantor**)**.** On part de l'intervalle $C_0 = [0, 1]$. On lui retire son tiers du milieu. On obtient ainsi $C_1 = [0, \frac{1}{3}] \cup [\frac{2}{3}, 1]$. On réitère le procédé : on retire à chacun des deux intervalles de C_1 leur tiers du milieu. On obtient $C_2 = [0, \frac{1}{9}] \cup [\frac{2}{9}, \frac{1}{3}] \cup [\frac{2}{3}, \frac{7}{9}] \cup [\frac{8}{9}, 1]$. Et on réitère... On obtient la construction de la figure 3.

Formellement, on définit par récurrence les ensembles C_n, par $C_0 = [0, 1]$, et pour tout $n \geq 1$,

$$C_n = \frac{C_{n-1}}{3} \cup \left(\frac{2}{3} + \frac{C_{n-1}}{3}\right),$$

c'est-à-dire que l'intervalle $[a, b]$ est envoyé sur $[a, a + \frac{b-a}{3}] \cup [b - \frac{b-a}{3}, b]$.

On vérifie alors par récurrence que C_n est la réunion de 2^n intervalles fermés, deux à deux disjoints, de longueur $\frac{1}{3^n}$ donnés par

$$\left[\sum_{k=1}^{n} \frac{x_k}{3^k}, \sum_{k=1}^{n} \frac{x_k}{3^k} + \frac{1}{3^n}\right], \quad \text{avec } x_k \in \{0, 2\}.$$

Par exemple, $C_2 = [0, \frac{1}{3^2}] \cup [\frac{2}{3^2}, \frac{2}{3^2} + \frac{1}{3^2}] \cup [\frac{2}{3}, \frac{2}{3} + \frac{1}{9}] \cup [\frac{2}{3} + \frac{2}{3^2}, 1]$.

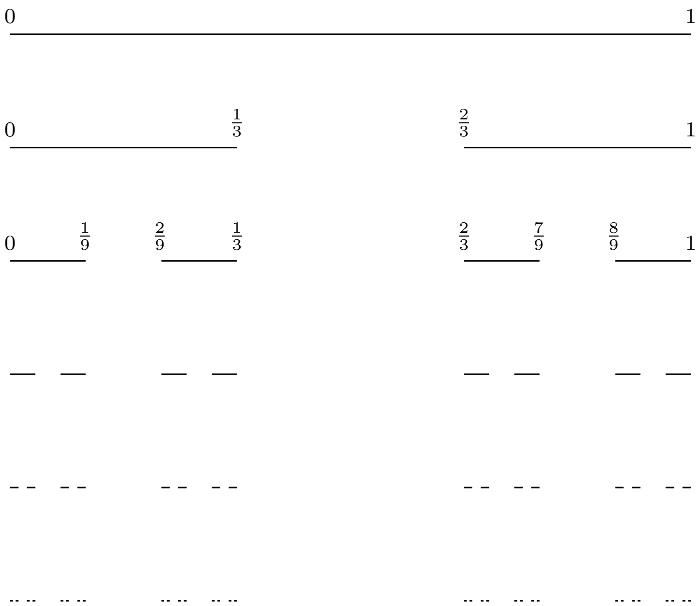

FIGURE 3. L'ensemble triadique de Cantor

On appelle alors ensemble triadique de Cantor l'ensemble :

$$C = \bigcap_{n \geq 0} C_n.$$

QUELQUES PROPRIÉTÉS DE C :

(i) il est non vide : il contient clairement 0 et 1 et même en fait toutes les bornes des intervalles définissant les C_n ;

(ii) il est compact car fermé (comme intersections de fermés, chaque C_n étant fermé car union *finie* de fermés) et borné (car contenu dans $[0,1]$) ;

(iii) il est sans point isolé : soit $x \in C$, alors $x \in C_n$, pour tout $n \geq 0$. Donc, pour tout n, x est dans un intervalle de la forme

$$\left[\sum_{k=1}^{n} \frac{x_k}{3^k}, \sum_{k=1}^{n} \frac{x_k}{3^k} + \frac{1}{3^n}\right], \quad \text{avec } x_k \in \{0,2\}.$$

avec $x_k \in \{0,2\}$. Donc x est la limite de deux suites adjacentes et ne peut être isolé ;

(iv) Il est non dénombrable, et en bijection avec $\{0,2\}^{\mathbb{N}^*}$, c'est-à-dire que C peut s'écrire sous forme

$$C = \left\{\sum_{n \geq 1} \frac{x_n}{3^n},\ x_n \in \{0,2\}\right\}.$$

$\supseteq$: Soit $(x_n)_n$ une suite de $\{0,2\}$. Posons $a_n = \sum_{k=1}^{n} \frac{x_k}{3^n}$. Alors, pour tout $n \geq 1$, $x_n \in C_n$ par définition de C_n, donc pour tout n, $x_n \in C$. Comme C est fermé, on a $\sum_{n \geq 1} \frac{x_n}{3^n} = \lim_n a_n \in C$. Remarque : cela suffit pour montrer que C est non-dénombrable, et comme $C \subset [0,1]$, qu'il est de même cardinal que $\mathbb{R}$.

$\subseteq$: Soit $x \in C$, donc pour tout $n \geq 0$, il existe un intervalle de la forme $I_n = [a_n, a_n + \frac{1}{3^n}]$ tel que $x_n \in I_n$. On a donc que $|a_n - x| \leq \frac{1}{3^n}$. Quand on passe de C_n à C_{n+1}, soit x est dans le premier tiers de I_n, auquel cas $a_{n+1} = a_n$, soit x est dans le dernier tiers de I_n, auquel cas, $a_{n+1} = a_n + \frac{2}{3^{n+1}}$. On a donc que pour tout n, $a_{n+1} = a_n + \frac{x_{n+1}}{3^{n+1}}$, avec $x_{n+1} = 0$ ou 2. Comme $a_0 = 0$, on a $a_n = \sum_{k=1}^{n} \frac{x_k}{3^k}$. Or $a_n \to \sum_{k \geq 1} \frac{x_k}{3^k}$, et $a_n \to x$, c'est donc que $x = \sum_{n \geq 1} \frac{x_n}{3^n}$.

(v) il est de mesure de Lebesgue nulle : comme $C_{n+1} \subset C_n$ et que $\lambda(C_0) = 1$, on a par limite décroissante ("continuité à droite") :

$$\lambda(C) = \lim_n \lambda(C_n).$$

Or C_n est la réunion de 2^n intervalles disjoints de longueur $\frac{1}{3^n}$, donc $\lambda(C_n) = \frac{2^n}{3^n}$ qui tend vers 0 quand $n \to \infty$. Finalement, $\lambda(C) = 0$.

(vi) C est d'intérieur vide : si ce n'était pas le cas, il contiendrait un intervalle ouvert $]a, b[$ non vide (C est non vide). On aurait donc par croissance,

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

et $\lambda(C)$ serait non nulle. Or on a montré juste au dessus que $\lambda(C) = 0$.

Remarque : On peut faire une construction analogue mais au lieu de retirer le tiers du milieu, on retire un intervalle de longueur $\frac{\varepsilon}{3^n}$ au milieu. On construit ainsi un compact d'intérieur vide, mais cette fois de mesure de Lebesgue $1-\varepsilon$, pour n'importe quel $\varepsilon \in]0, 1[$.

5. Fonctions mesurables

Notation 5.1 (TRÈS importante !)**.** Soit $f\colon E \to F$ et $A \subset F$. On utilisera la notation abrégée $\{f \in A\}$ pour désigner l'image réciproque de A, c'est-à-dire

$$\{f \in A\} = \{x \in E \,|\, f(x) \in A\} = f^{-1}(A).$$

Par exemple, si $F = \mathbb{R}$, on notera $\{f = a\} = \{x \in X \,|\, f(x) = a\}$, ou encore $\{f > a\} = \{x \in X \,|\, f(x) > a\}$, etc...

5.1. **Définition.**

Définition 5.1. *Soit $(E, \mathcal{A})$ et $(F, \mathcal{B})$ deux espaces mesurables. Une fonction $f\colon E \to F$ est dite mesurable si pour tout $B \in \mathcal{B}$,*

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

autrement dit, $f^{-1}(\mathcal{B}) \subset \mathcal{A}$.

Si $\mathcal{A}$ et $\mathcal{B}$ sont des tribus boréliennes, on parlera de fonction borélienne au lieu de fonction mesurable.

Exemple 5.1. Les (fonctions) constantes sont mesurables (exo !).

La fonction indicatrice $\mathbb{1}_A\colon (E, \mathcal{A}) \to (\{0, 1\}, \mathcal{P}(\{0, 1\}))$ est mesurable si est seulement si A est un ensemble mesurable (exo !).

Remarque 5.1. Si on se donne uniquement la tribu $\mathcal{B}$, $f^{-1}(\mathcal{B})$ est la plus petite tribu sur E qui rende f mesurable.

Si on se donne uniquement la tribu $\mathcal{A}$, la tribu image de $\mathcal{A}$ par f est la plus grande tribu sur F qui rende f mesurable.

Commençons par un résultat bien utile pour montrer la mesurabilité d'une fonction lorsque la tribu d'arrivée est engendrée par une classe de parties.

Proposition 5.1 (Lemme de transport)**.** *Soit $f\colon X \to Y$, et $\mathcal{C}$ une classe de parties de Y. Alors*

$$\sigma(f^{-1}(\mathcal{C})) = f^{-1}(\sigma(\mathcal{C})).$$

Démonstration. Montrons l'inclusion $\subset$. Comme $\mathcal{C} \subset \sigma(\mathcal{C})$, on a $f^{-1}(\mathcal{C}) \subset f^{-1}(\sigma(\mathcal{C}))$. Comme $f^{-1}(\sigma(\mathcal{C}))$ est une tribu, et que $\sigma(f^{-1}(\mathcal{C}))$ est la plus petite tribu contenant $f^{-1}(\mathcal{C})$, on obtient l'inclusion

$$\sigma(f^{-1}(\mathcal{C})) \subset f^{-1}(\sigma(\mathcal{C})).$$

Pour l'inclusion inverse, considérons

$$\mathcal{B} = \{B \subset Y \,|\, f^{-1}(B) \in \sigma(f^{-1}(\mathcal{C}))\},$$

c'est-à-dire la tribu image de $\sigma(f^{-1}(\mathcal{C}))$ par f. On a clairement que $\mathcal{B}$ contient $\mathcal{C}$, donc contient $\sigma(\mathcal{C})$, et donc

$$f^{-1}(\sigma(\mathcal{C})) \subset f^{-1}(\mathcal{B}).$$

Mais par définition de $\mathcal{B}$, on a $f^{-1}(\mathcal{B}) \subset \sigma(f^{-1}(\mathcal{C}))$, ainsi

$$f^{-1}(\sigma(\mathcal{C})) \subset \sigma(f^{-1}(\mathcal{C})).$$

□

La grande utilité de ce lemme est que pour montrer qu'une fonction est mesurable si la tribu d'arrivée est engendrée par une classe de parties $\mathcal{C}$, il suffit de montrer que pour tout $C \in \mathcal{C}$, $f^{-1}(C)$ appartient à la tribu de départ. En effet, si $f\colon (X, \mathcal{X}) \to (Y, \sigma(\mathcal{C}))$ vérifie

$$\forall C \in \mathcal{C},\ f^{-1}(C) \in \mathcal{X},$$

dit autrement, $f^{-1}(\mathcal{C}) \subset \mathcal{X}$, alors $\sigma(f^{-1}(\mathcal{C})) \subset \mathcal{X}$, et donc par le lemme de transport,

$$f^{-1}(\sigma(\mathcal{C})) \subset \mathcal{X},$$

c'est-à-dire que f est mesurable. Ceci sert en particulier pour montrer qu'une fonction est borélienne, la tribu de Borel étant engendrée par exemple par les intervalles ouverts $]a, b[$.

Remarque 5.2. Autre remarque, bien utile. Soit $f\colon (X, \mathcal{B}(X)) \to (Y, \mathcal{B}(Y))$. Alors si f est continue, f est borélienne. En effet, si A est un ouvert de Y, alors par continuité $f^{-1}(A)$ est un ouvert de X, et donc appartient à la tribu borélienne $\mathcal{B}(X)$.

5.2. **Opérations.**

Proposition 5.2. *Soit $f\colon (X, \mathcal{X}) \to (E, \mathcal{E})$ et $g\colon (E, \mathcal{E}) \to (F, \mathcal{F})$ deux fonctions mesurables. Alors, $g \circ f$ est mesurable de $(X, \mathcal{X})$ dans $(F, \mathcal{F})$.*

Démonstration. Pour tout $A \subset F$, on a $(g \circ f)^{-1}(A) = f^{-1}(g^{-1}(A))$. En effet,

$$f^{-1}\left(g^{-1}(A)\right) = \{x \in X \,|\, f(x) \in g^{-1}(A)\} = \{x \in X \,|\, g(f(x)) \in A\} = (g \circ f)^{-1}(A).$$

Donc si $A \in \mathcal{F}$, $g^{-1}(A) \in \mathcal{E}$ par mesurabilité de g, et $f^{-1}(g^{-1}(A)) \in \mathcal{X}$ par mesurabilité de f. □

Proposition 5.3. *Soient f et g deux fonctions mesurables de $(E, \mathcal{A})$ dans $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ et soit $\alpha \in \mathbb{R}$. Alors, $\alpha f + g$ et fg sont mesurables.*

Démonstration. Soit $\alpha \neq 0$ (sinon c'est trivial). Alors, pour tout $a \in \mathbb{R}$,

$$\{\alpha f < a\} = \{f < a/\alpha\} \in \mathcal{A},$$

car f est mesurable. Ainsi, αf est mesurable. Pour la somme, on écrit

$$\{f + g < a\} = \{f < a - g\} = \bigcup_{q \in \mathbb{Q}} \{f < q < a - g\} = \bigcup_{q \in \mathbb{Q}} \{f < q\} \cap \{g < a - q\}$$

qui est mesurable comme union dénombrable d'ensembles mesurables. Ainsi, $f + g$ est mesurable. Pour le produit, on commence par montrer que f^2 est mesurable :

$$\{f^2 < a\} = \{-\sqrt{a} < f < \sqrt{a}\}$$

qui est bien dans $\mathcal{A}$. Alors, comme $fg = \frac{1}{4}\left((f+g)^2 - (f-g)^2\right)$, on conclut. □

Proposition 5.4. *Soit $(f_n)_n$ une suite de fonctions mesurables de $(X, \mathcal{X})$ dans $(\overline{\mathbb{R}}, \mathcal{B}(\overline{\mathbb{R}}))$. Alors, $\sup_n f_n$, $\inf_n f_n$, $\limsup_n f_n$, $\liminf_n f_n$ sont mesurables. De plus, si $(f_n)_n$ converge simplement vers une fonction f, alors f est mesurable.*

Démonstration. Pour tout $a \in \mathbb{R}$, $\{\sup_n f_n \leq a\} = \bigcap_n \{f_n \leq a\} \in \mathcal{X}$, donc $\sup_n f_n$ est mesurable. De même, $\{\inf_n f_n \geq a\} = \bigcap_n \{f_n \geq a\} \in \mathcal{X}$, donc $\inf_n f_n$ est mesurable. On en déduit que $\limsup_n f_n$ et $\liminf_n f_n$ sont aussi mesurables. Si f_n converge simplement vers f, alors $f = \limsup_n f_n = \liminf_n f_n$, et donc f est mesurable. □

5.3. **Fonctions étagées.**

Définition 5.2. *Une fonction mesurable de $(E, \mathcal{E})$ dans $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ est dite étagée si elle ne prend qu'un nombre fini de valeurs. Autrement dit, il existe une partition finie mesurable $(A_i)_{i \in I}$ de E avec I fini et $A_i \in \mathcal{E}$ pour tout $i \in I$, et des réels α_i tel que $f = \sum_{i \in I} \alpha_i \mathbb{1}_{A_I}$.*

Remarque 5.3. On peut toujours choisir les α_i deux à deux distincts et les $A_i = \{f = \alpha_i\}$. On appelle alors forme canonique de f la décomposition :

$$f = \sum_{\alpha \in f(E)} \alpha \mathbb{1}_{\{f = \alpha\}}.$$

Exemple 5.2. La fonction indicatrice $\mathbb{1}_A$ d'un ensemble mesurable A est bien sûr étagée, elle ne prend que deux valeurs : 0 ou 1. Une fonction en escalier est une fonction étagée, mais la réciproque est fausse. La fonction indicatrice des rationnels $\mathbb{1}_{\mathbb{Q}}$ est étagée, mais n'est constante sur aucun intervalle de $\mathbb{R}$ (non réduit à un point !).

Théorème 5.1 (Lemme fondamental d'approximation)**.** *Toute fonction mesurable de $(E, \mathcal{E})$ dans $(\overline{\mathbb{R}}, \mathcal{B}(\overline{\mathbb{R}}))$ est limite simple d'une suite de fonctions étagées. Si de plus,*

(i) f est positive, on peut choisir $(f_n)_n$ positive et croissante (au sens où $f_n \leq f_{n+1}$, pour tout n) ;

(ii) f est positive et bornée, on peut choisir $(f_n)_n$ de sorte que la convergence soit uniforme.

Démonstration. Supposons f positive. L'idée est de partitionner l'espace d'arrivée $\overline{\mathbb{R}}_+$ en

$$\bigcup_{k=0,\ldots,n2^n-1} \left[\frac{k}{2^n}, \frac{k+1}{2^n}\right[\cup [n, +\infty].$$

On définit alors la suite $(f_n)_n$ de fonctions étagées par

$$f_n = \sum_{k=0}^{n2^n-1} \frac{k}{2^n} \mathbb{1}_{\{\frac{k}{2^n} \leq f < \frac{k+1}{2^n}\}} + n\mathbb{1}_{\{f \geq n\}}.$$

Les f_n sont bien positives (évident), et la suite est croissante. En effet, soit $x \in X$, on distingue alors trois cas :

— Si $f(x) \geq n+1$, $f_{n+1}(x) = n+1 > n = f_n(x)$;

— Si $f(x) \in [n, n+1[$, alors $f_n(x) = n$, et il existe $k \in \{n2^{n+1}, \ldots, (n+1)2^{n+1} - 1\}$ tel que $f(x) \in [\frac{k}{2^{n+1}}, \frac{k+1}{2^{n+1}}[$. Donc $f_{n+1}(x) = \frac{k}{2^{n+1}} \geq n = f_n(x)$;

— Si $f(x) < n$, alors il existe $k \in \{0, \ldots, n2^n - 1\}$ tel que $f(x) \in [\frac{k}{2^n}, \frac{k+1}{2^n}[= [\frac{2k}{2^{n+1}}, \frac{2(k+1)}{2^{n+1}}[$. Alors, $f_n(x) = \frac{k}{2^n}$, et $f_{n+1}(x)$ vaut soit $\frac{2k}{2^{n+1}} = \frac{k}{2^n}$, soit $\frac{2k+1}{2^{n+1}} > \frac{k}{2^n}$. Dans les deux cas, $f_{n+1}(x) \geq f_n(x)$.

Montrons maintenant la convergence.

Si $f(x) = +\infty$, $f_n(x) = n$ pour tout n, et $f_n(x) \to f(x)$ quand $n \to +\infty$.

Si $f(x) < \infty$, alors il existe n_0 (qui dépend de x) tel que $f(x) < n_0$. Pour tout $n \geq n_0$, on a $|f_n(x) - f(x)| \leq \frac{1}{2^n}$. Ainsi, $f_n(x) \to f(x)$ quand $n \to \infty$.

Si f est bornée, il existe n_0 tel que pour tout $x \in E$, $f(x) < n_0$, donc pour tout $x \in E$, pour tout $n \geq n_0$, $|f_n(x) - f(x)| \leq \frac{1}{2^n}$, et $(f_n)_n$ converge uniformément vers f.

Si f est de signe quelconque, on écrit $f = f^+ - f^-$, avec

$$f^+ = f\mathbb{1}_{f>0} \quad \text{et} \quad f^- = -f\mathbb{1}_{f<0}.$$

Les fonctions f^+ et f^- sont mesurables et positives, et on applique alors le résultat précédent. □

6. Intégrale des fonctions mesurables positives.

6.1. Intégrale des fonctions étagées positives.

Définition 6.1. *Soit $(E, \mathcal{A}, \mu)$ un espace mesuré et soit $f\colon (E, \mathcal{A}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ une fonction étagée positive. On appelle intégrale de f par rapport à μ, que l'on note $\int_E f d\mu$, l'élément de $[0, +\infty]$ défini par*

$$\int_E f d\mu := \sum_{\alpha \in f(E)} \alpha \mu \left(\{f = \alpha\}\right),$$

avec la convention que $0 \times \infty = 0$.

Notation 6.1. L'intégrale d'une fonction sera notée indifféremment :

$$\int_E f d\mu, \quad \int_E f(x) d\mu(x), \quad \int_E f(x) \mu(dx), \quad \text{ou encore } \int f d\mu$$

si l'espace sur lequel on intègre est clair. On notera aussi :

$$\int_A f d\mu = \int_E f \mathbb{1}_A d\mu.$$

Remarque 6.1. Cette définition ne dépend pas de la représentation de f. En effet, soit $f = \sum_{i \in I} \alpha_i \mathbb{1}_{A_i}$ une autre décomposition de f, avec donc I fini, et les A_i deux à deux disjoints. Alors, comme,

$$\{f = \alpha\} = \bigcup_{i \,|\, \alpha_i = \alpha} A_i,$$

on a :

$$\sum_{\alpha \in f(E)} \alpha \mu(\{f = \alpha\}) = \sum_{\alpha \in f(E)} \alpha \sum_{i \,|\, \alpha_i = \alpha} \mu(A_i) = \sum_{\alpha \in f(E)} \sum_{i \,|\, \alpha_i = \alpha} \alpha_i \mu(A_i) = \sum_{i \in I} \alpha_i \mu(A_i).$$

Exemple 6.1. L'intégrale d'une fonction étagée $f = \sum_i \alpha_i \mathbb{1}_{A_i}$ par rapport à la masse de Dirac δ_a est :

$$\int_X f \delta_a = \sum_i \alpha_i \delta_a(A_i) = \sum_i \alpha_i \mathbb{1}_{A_i}(a) = f(a).$$

L'intégrale de $\mathbb{1}_{\mathbb{Q}}$ par rapport à la mesure de Lebesgue λ donne :

$$\int_{\mathbb{R}} \mathbb{1}_{\mathbb{Q}} d\lambda = \lambda(\mathbb{Q}) = 0.$$

Rappelons encore une fois que la fonction $\mathbb{1}_{\mathbb{Q}}$ n'est pas Riemann intégrable !

Donnons les premières propriétés.

Proposition 6.1. *L'intégrale est :*

(i) linéaire : pour toutes fonctions étagées positives f et g, $\int_E (f + g) d\mu = \int_E f d\mu + \int_E g d\mu$;

(ii) positivement homogène : pour tout $\alpha \geq 0$, et toute fonction f étagée positive, $\int_E \alpha f d\mu = \alpha \int_E f d\mu$;

(iii) croissante : pour toutes fonctions étagées positives f et g, telles que $f \leq g$, on a $\int f d\mu \leq \int g d\mu$.

Démonstration. (i) Soit deux fonctions étagées $f = \sum_{i \in I} \alpha_i \mathbb{1}_{A_i}$ et $g = \sum_{j \in J} \beta_j \mathbb{1}_{B_j}$, avec $(A_i)_{i \in I}$ et $(B_j)_{j \in J}$ deux partitions finies de E. Alors $(A_i \cap B_j)_{(i,j) \in I \times J}$ est une partition finie de E, et on a

$$f + g = \sum_{(i,j) \in I \times J} (\alpha_i + \beta_j) \mathbb{1}_{A_i \cap B_j}.$$

Ainsi,

$$\begin{aligned}\int_E (f+g)d\mu &= \sum_{(i,j)\in I\times J} (\alpha_i+\beta_j)\mu(A_i\cap B_j)\\ &= \sum_{(i,j)\in I\times J} \alpha_i\mu(A_i\cap B_j) + \sum_{(i,j)\in I\times J} \beta_j\mu(A_i\cap B_j)\\ &= \sum_{i\in I}\alpha_i\sum_{j\in J}\mu(A_i\cap B_j) + \sum_{j\in J}\beta_j\sum_{i\in I}\mu(A_i\cap B_j)\\ &= \sum_{i\in I}\alpha_i\mu(A_i) + \sum_{j\in J}\beta_j\mu(B_j)\\ &= \int_E f d\mu + \int_E g d\mu.\end{aligned}$$

(ii) Facile, exo !

(iii) La fonction $g = f + g - f$ est la somme de deux fonctions étagées positives, et on applique le point (i).

□

6.2. **Intégrale des fonctions mesurables positives.**

Définition 6.2. *Soit $(E,\mathcal{A},\mu)$ un espace mesuré et soit $f\colon (E,\mathcal{A}) \to (\overline{\mathbb{R}},\mathcal{B}(\overline{\mathbb{R}}))$ une fonction mesurable positive. On appelle intégrale de f par rapport à μ, que l'on note $\int_E f d\mu$, l'élément de $[0,+\infty]$ défini par*

$$\int_E f d\mu := \sup\left\{\int_E \varphi d\mu \,\middle|\, \varphi \text{ étagée positive}, \varphi \le f\right\}.$$

Si $\int_E f d\mu < \infty$, on dit que f est intégrable.

Comme première propriété, on remarque facilement que l'intégrale est croissante :

Proposition 6.2. *Pour toutes fonctions f et g mesurables positives telles que $f \le g$, alors*

$$\int_E f d\mu \le \int_E g d\mu.$$

Démonstration. Si φ est une fonction étagée positive telle que $\varphi \le f$, alors $\varphi \le g$, et donc par définition du sup,

$$\sup\left\{\int_E \varphi d\mu \,\middle|\, \varphi \text{ étagée positive}, \varphi \le f\right\} \le \sup\left\{\int_E \varphi d\mu \,\middle|\, \varphi \text{ étagée positive}, \varphi \le g\right\}.$$ □

Théorème 6.1 (Théorème de Beppo-Levi ou Théorème de convergence monotone)**.** *Soit $(f_n)_n$ une suite* **croissante** *de fonctions mesurables positives qui converge simplement vers f. Alors*

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

Remarque 6.2. Le théorème de convergence monotone donne alors que $\int f d\mu$ est la limite de $\int f_n d\mu$ où $(f_n)_n$ est une suite arbitraire de fonctions étagées positives croissant vers f.

Démonstration. Soit $(f_n)_n$ une suite de fonctions mesurables positives qui converge en croissant vers f. Rappelons que f est alors mesurable comme limite simple de fonctions mesurables. On a alors $f_n \le f$, et par croissance de l'intégrale,

$$\int_E f_n d\mu \le \int_E f d\mu.$$

En passant à la limite (qui existe dans $[0,+\infty]$ par croissance), on obtient donc

$$\lim_n \int_E f_n d\mu \leq \int_E f d\mu.$$

Montrons l'inégalité inverse. Par définition de $\int f d\mu$, il suffit de montrer que pour toute fonction étagée positive φ telle que $\varphi \leq f$, $\int \varphi d\mu \leq \lim_n \int f_n d\mu$.

Soit alors φ une telle fonction et $a \in [0,1[$. Posons, $E_n = \{a\varphi \leq f_n\}$. Alors la suite $(E_n)_n$ est croissante (car $(f_n)_n$ est croissante), et on a $E = \bigcup_n E_n$. En effet, soit $x \in E$. Si $f(x) = 0$, alors $\varphi(x) = f_n(x) = 0$ pour tout n, et $x \in E_n$ pour tout n, et donc $x \in \bigcup_n E_n$. Si $f(x) > 0$, $f(x) \geq \varphi(x) > a\varphi(x)$, et comme $f_n(x)$ croit vers $f(x)$, il existe n assez grand tel que $f_n(x) \geq a\varphi(x)$. Donc $x \in E_n$, et finalement $E = \bigcup_n E_n$. En écrivant $\varphi = \sum_{i\in I} \alpha_i \mathbb{1}_{A_i}$, avec I fini, on a alors

$$\int_E a\varphi \mathbb{1}_{E_n} d\mu = \sum_{i\in I} a\alpha_i \mu(A_i \cap E_n) \leq \int_E f_n d\mu,$$

car $a\varphi \leq f_n$ sur E_n. Comme la somme à gauche est finie, on peut passer à la limite, ce qui donne

$$a\int_E \varphi d\mu \leq \lim_n \int_E f_n d\mu.$$

Comme a est arbitraire, en faisant tendre a vers 1, on obtient finalement

$$\int_E \varphi d\mu \leq \lim_n \int_E f_n d\mu,$$

ce qui achève la démonstration. □

Remarque 6.3. Le théorème de convergence monotone s'applique pour les suites <u>croissantes</u> de fonctions mesurables, et pas décroissantes. Par exemple, la suite $f_n = \mathbb{1}_{]n,+\infty[}$ est décroissante, et converge vers 0. Or, par rapport à la mesure de Lebesgue,

$$\int_{\mathbb{R}} \mathbb{1}_{]n,+\infty[} d\lambda = \lambda(]n,+\infty[) = +\infty$$

et donc ne converge pas vers 0. À titre d'exercice, on pourra montrer que si $(f_n)_n$ est une suite décroissante de fonctions mesurables positives qui converge simplement vers f, et telle qu'il existe n_0 tel que $\int f_{n_0} d\mu < \infty$, alors $\lim_n \int f_n d\mu = \int f d\mu$.

Remarque 6.4. Considérons la suite de fonctions boréliennes $f_n = \frac{1}{n}\mathbb{1}_{]0,n[}$. Alors, f_n converge vers 0 uniformément, car, pour tout $x \in \mathbb{R}$,

$$|f_n(x)| \leq \frac{1}{n} \to 0.$$

Or l'intégrale de f_n par rapport à la mesure de Lebesgue est

$$\int_{\mathbb{R}} f_n d\lambda = \frac{1}{n}\int \mathbb{1}_{]0,n[} d\lambda = \frac{1}{n}\lambda(]0,n[) = 1,$$

pour tout n, et donc ne converge pas vers 0 ! L'hypothèse de croissance de la suite $(f_n)_n$ ne peut pas être omise sans hypothèse supplémentaire, même si la convergence est uniforme. Cela suggère tout de même un lien entre limite de l'intégrale et intégrale de la limite, donnée par le lemme suivant :

Proposition 6.3 (Lemme de Fatou)**.** *Soit $(f_n)_n$ une suite de fonctions mesurables positives de $(E,\mathcal{A})$ dans $(\overline{\mathbb{R}}, \mathcal{B}(\overline{\mathbb{R}}))$. Alors,*

$$\int_E \liminf f_n d\mu \leq \liminf_n \int_E f_n d\mu.$$

Remarque : On remarque qu'en prenant $f_n = \mathbb{1}_{A_n}$ avec $(A_n)_n$ une suite d'ensembles mesurables, on retrouve l'inégalité

$$\mu(\liminf_n A_n) \leq \liminf_n \mu(A_n)$$

vue précédemment.

Démonstration. Commençons par rappeler que $\liminf_n f_n$ est mesurable. La suite de fonctions définies par $g_n = \inf_{k\geq n} f_k$ est croissante et croît vers $\liminf_n f_n$. Par le théorème de convergence monotone, on a donc

$$\sup_n \int_E g_n d\mu = \int_E \liminf f_n d\mu.$$

De plus, comme pour tout $k \geq n$, $g_n \leq f_k$, par croissance de l'intégrale, $\int g_n d\mu \leq \int f_k d\mu$, donc $\int g_n d\mu \leq \inf_{k\geq n} \int f_k d\mu$, et en prenant le sup,

$$\int_E \liminf f_n d\mu = \sup_n \int_E g_n d\mu \leq \sup_n \inf_{k\geq n} \int_E f_k d\mu = \liminf_n \int_E f_n d\mu. \qquad \square$$

Le théorème de convergence monotone est très commode pour montrer certaines propriétés de l'intégrale d'une fonction mesurable. Il suffit de montrer ces propriétés pour les fonctions étagées, puis de passer à la limite à l'aide du théorème. Par exemple,

Proposition 6.4. *L'intégrale est linéaire, positivement homogène et croissante.*

Exemple 6.2. (i) Soit δ_a la masse de Dirac au point a sur un espace mesurable $(E, \mathcal{A})$. Soit f une fonction mesurable de $(E, \mathcal{A})$ dans $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ et positive. Alors

$$\int_E f d\delta_a = f(a).$$

Cela se vérifie aisément pour une fonction f étagée de la forme $f = \sum_{i\in I} \alpha_i \mathbb{1}_{A_i}$, avec I fini. En effet, on a alors

$$\int_E f d\delta_a = \sum_{i\in I} \alpha_i \delta_a(A_i) = \sum_{i\in I} \alpha_i \mathbb{1}_{A_i}(a) = f(a).$$

Si f est maintenant mesurable et positive, par le théorème fondamental d'approximation, il existe une suite $(f_n)_n$ croissante de fonctions étagées positives qui converge vers f, on a donc $\lim_n f_n(a) = f(a)$, et par le théorème de convergence monotone, la suite des intégrales $\int f_n d\delta_a$ converge vers $\int f d\delta_a$. Ainsi, $\int_E f d\delta_a = f(a)$.

(ii) Soit l'espace mesuré $(\mathbb{N}, \mathcal{P}(\mathbb{N}), m)$ où m est la mesure de comptage. Soit $u\colon \mathbb{N} \to \mathbb{R}$, autrement dit une suite $(u_n)_{n\geq 0}$. Alors, u est toujours mesurable, et l'intégrale de u par rapport à m est alors :

$$\int_{\mathbb{N}} u dm = \sum_{n\geq 0} u_n.$$

Une série peut donc être vue comme une intégrale par rapport à la mesure de comptage sur $\mathbb{N}$. En particulier, tous les résultats sur l'intégrale s'applique tels quels aux séries.

(iii) De même, sur $(\mathbb{N}, \mathcal{P}(\mathbb{N}))$, considérons la mesure de Poisson $\mu = \sum_{n\geq 0} \frac{\lambda^n}{n!} e^{-\lambda} \delta_n$, avec $\lambda > 0$. Alors,

$$\int_{\mathbb{N}} f d\mu = \sum_{n\geq 0} f(n) \frac{\lambda^n}{n!} e^{-\lambda}.$$

On a donc que $\int_{\mathbb{N}} f d\mu$ est exactement l'espérance de $f(X)$ où X est une variable aléatoire suivant la loi de Poisson !

Proposition 6.5 (Inversion somme/intégrale)**.** *Pour toute suite $(f_n)_n$ de fonctions mesurables positives de $(E, \mathcal{A})$ dans $(\mathbb{R}, \mathcal{B}(\overline{\mathbb{R}}))$, et pour μ une mesure sur $(E, \mathcal{A})$, on a*

$$\int_E \left(\sum_n f_n \right) d\mu = \sum_n \int_E f_n d\mu.$$

Démonstration. Posons $g_n = \sum_{k=0}^n f_k$. Alors $(g_n)_n$ est une suite mesurable, croissante, de limite $\sum_n f_n$ (dans $[0, +\infty]$) qui est mesurable. Le théorème de convergence monotone et la linéarité de l'intégrale donne alors le résultat :

$$\int_E \left(\sum_n f_n \right) d\mu = \int_E \lim_n \sum_{k=0}^n f_k d\mu = \lim_n \int_E \sum_{k=0}^n f_k d\mu = \lim_n \sum_{k=0}^n \int_E f_k d\mu = \sum_n \int_E f_n d\mu.$$

$\square$

7. Intégrale des fonctions mesurables de signe quelconque

Notation 7.1. Soit $f\colon (E,\mathcal{A}) \to (\overline{\mathbb{R}}, \mathcal{B}(\overline{\mathbb{R}}))$ une fonction mesurable. On notera :

$$f^+ = f\mathbb{1}_{\{f>0\}} \quad \text{et} \quad f^- = -f\mathbb{1}_{\{f<0\}},$$

respectivement les parties positive et négative de f. On a alors que f^+ et f^- sont mesurables et **positives**, et on a

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

Définition 7.1. *Soit $f\colon (E,\mathcal{A}) \to (\overline{\mathbb{R}}, \mathcal{B}(\overline{\mathbb{R}}))$ une fonction mesurable. On dit que f est intégrable par rapport à μ si*

$$\int_E |f| d\mu < \infty,$$

et dans ce cas, on définit l'intégrale de f par rapport à μ par

$$\int_E f d\mu = \int_E f^+ d\mu - \int_E f^- d\mu.$$

Remarquons que $\int |f| d\mu < \infty$ équivaut à $\int f^+ d\mu < \infty$ et $\int f^- d\mu < \infty$, et donc l'intégrale de f est bien définie et est un élément de $\mathbb{R}$. On dira aussi que f admet une intégrale dans $\overline{\mathbb{R}}$ si seulement l'un de ces deux nombres est fini.

La propriété suivante s'appelle parfois l'inégalité triangulaire pour l'intégrale.

Proposition 7.1. *Pour toute fonction mesurable f,*

$$\left| \int_E f d\mu \right| \leq \int_E |f| d\mu.$$

Démonstration. C'est une conséquence simple de l'inégalité triangulaire (l'usuelle) :

$$\left| \int_E f d\mu \right| = \left| \int_E f^+ d\mu - \int_E f^- d\mu \right| \leq \int_E f^+ d\mu + \int_E f^- d\mu = \int_E |f| d\mu. \qquad \square$$

Définition 7.2. *Soit $(E,\mathcal{A},\mu)$ un espace mesuré. On notera $\mathcal{L}^1(E,\mathcal{A},\mu)$, ou en abrégé $\mathcal{L}^1(\mu)$, l'ensemble des fonctions intégrables par rapport à μ.*

On a alors :

Proposition 7.2. *L'espace $\mathcal{L}^1(\mu)$ est un $\mathbb{R}$-espace vectoriel.*

Démonstration. Soit f, g deux fonctions intégrables, et α un réel. Par l'inégalité triangulaire, $|\alpha f + g| \leq |\alpha||f| + |g|$, donc $\alpha f + g$ est intégrable par les propriétés de l'intégrale des fonctions mesurables positives. □

Proposition 7.3. *L'intégrale est une forme linéaire positive (donc en particulier croissante).*

Démonstration. On se ramène au cas fonctions positives par la décomposition $f = f^+ - f^-$ (exo !). □

Remarque 7.1. Si m est la mesure de comptage sur $\mathbb{N}$, $\mathcal{L}^1(m)$ n'est rien d'autre que l'espace $l^1(\mathbb{N})$ des suites dont la série est absolument convergente.

7.1. Propriétés de l'intégrale.

Définition 7.3. *Soit $(E,\mathcal{A},\mu)$ un espacé mesuré. On dit que $A \in \mathcal{P}(E)$ est un ensemble μ-négligeable s'il existe un ensemble mesurable $N \in \mathcal{A}$ tel que $A \subset N$ et $\mu(N) = 0$.*

Soit $P(x)$ une certaine propriété qui dépend de $x \in E$. On dira que $P(x)$ est vraie pour μ-presque tout x, ou encore que P est vraie μ-presque partout, si $\{x \in E \mid P(x) \text{est fausse}\}$ est un ensemble μ-négligeable, c'est-à-dire qu'il existe $N \in \mathcal{A}$ tel que $\mu(N) = 0$ et $\{x \in E \mid P(x) \text{est fausse}\} \subset N$.

On abrégera presque partout en p.p. et si le contexte est clair on ne précisera pas forcément pour quelle mesure.

On remarquera que dans la définition d'ensemble négligeable, on ne suppose pas l'ensemble en question mesurable, mais seulement inclus dans un ensemble mesurable de mesure nulle... On abordera (peut-être) les subtilités de ce point plus tard. On sera la plupart du temps dans la situation où l'ensemble négligeable A est lui même mesurable, et donc $\mu(A) = 0$.

Exemple 7.1. Soit $f\colon (X, \mathcal{A}, \mu) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ une fonction mesurable. Alors $f = 0$ μ-presque partout si $\{f \neq 0\}$ est μ-négligeable, et comme f est mesurable, alors $f = 0$ μ-presque partout si $\mu(\{f \neq 0\}) = 0$.

De même, $|f| < \infty$ μ-p.p. si $\mu(\{|f| = \infty\}) = 0$. Autrement dit, f est finie en dehors d'un ensemble de mesure nulle.

Sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ muni de la mesure de Lebesgue, la fonction indicatrice des rationnels $\mathbb{1}_{\mathbb{Q}}$ est nulle λ-presque partout, car $\lambda(\mathbb{Q}) = 0$.

Proposition 7.4 (Inégalité de Markov)**.** *Soit $f\colon (E, \mathcal{A}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ mesurable. Alors, pour tout $a > 0$,*

$$\mu(|f| \geq a) \leq \frac{1}{a} \int_E |f| d\mu.$$

Démonstration. On a

$$a\mathbb{1}_{\{|f| \geq a\}} \leq |f|\mathbb{1}_{\{|f| \geq a\}} \leq |f|,$$

donc par croissance de l'intégrale,

$$a\mu(|f| \geq a) \leq \int_E |f| d\mu.$$ □

Proposition 7.5. *Soit $f\colon (E, \mathcal{A}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ mesurable et* **positive***. On a*

$$\int_E f d\mu = 0 \Leftrightarrow f = 0 \ \mu\text{-}p.p.$$

Démonstration. Soit f nulle presque partout, i.e. $\mu(\{f \neq 0\}) = 0$. Soit φ étagée positive telle que $\varphi \leq f$. Comme $\{\varphi \neq 0\} \subset \{f \neq 0\}$, on a aussi $\varphi = 0$ μ-p.p. Donc $\int_E \varphi d\mu = 0$ et par définition de l'intégrale de f, on a aussi $\int_E f d\mu = 0$.

Réciproquement, soit f mesurable positive d'intégrale nulle. Par l'inégalité de Markov,

$$\mu(\{f \geq \frac{1}{n}\}) \leq n \int_E f d\mu = 0.$$

Or $\{f \neq 0\} = \bigcup_n \{f \geq \frac{1}{n}\}$, et la suite des ensembles $\{f \geq \frac{1}{n}\}$ est croissante, donc on a bien $\mu(\{f \neq 0\}) = 0$, i.e. $f = 0$ μ-p.p. □

Remarque 7.2. Soit f intégrable, et A un ensemble mesurable tel que $\mu(A) = 0$. Alors, par la proposition précédente,

$$\int |f| \mathbb{1}_A d\mu = 0,$$

car la fonction $g = |f|\mathbb{1}_A$ est positive et nulle presque partout (car $\{g \neq 0\} \subset A$).

Proposition 7.6. *Soit f et g deux fonctions intégrables telles que $f = g$ μ-presque partout. Alors,*

$$\int_E f d\mu = \int_E g d\mu.$$

Démonstration. Posons :

$$h = \max\{f, g\} - \min\{f, g\}.$$

Comme $\{f = g\} \subset \{h = 0\}$, on a $h = 0$ μ-p.p., donc par la proposition précédente $\int_E h d\mu = 0$, i.e.

$$\int_E \min\{f, g\} d\mu = \int_E \max\{f, g\} d\mu.$$

Comme $\min\{f, g\} \leq f, g \leq \max\{f, g\}$, par croissance de l'intégrale, on conclut. □

Remarque 7.3. Peu importe la valeur d'une fonction sur un ensemble négligeable pour calculer son intégrale !

Proposition 7.7. *Soit* $f\colon (E, \mathcal{A}) \to (\overline{\mathbb{R}}, \mathcal{B}(\overline{\mathbb{R}}))$ *une fonction mesurable. Alors*

$$\int_E |f| d\mu < \infty \implies f \textit{ est finie } \mu\textit{-p.p.}$$

Démonstration. Soit f μ-intégrable. On peut supposer f positive quitte à remplacer f par $|f|$. On veut montrer que f est finie μ-presque partout, i.e. $\mu(\{f = \infty\}) = 0$. Posons $A = \{f = \infty\}$. Par contraposée, supposons que $\mu(A) > 0$. Comme f s'écrit $f = \infty \mathbb{1}_A + f \mathbb{1}_{A^c}$, on a $f \geq \infty \mathbb{1}_A$. Par croissance de l'intégrale, on a donc

$$\int_E f d\mu \geq \infty \mu(A) = \infty,$$

car $\mu(A) > 0$, ce qui conclut la démonstration. □

La réciproque est bien entendue fausse, les constantes ne sont par exemple pas intégrables pour la mesure de Lebesgue ! Par contre, si μ est finie, on a la proposition suivante.

Proposition 7.8. *Soit* μ *une mesure finie et* f *une fonction mesurable. Si* f *est bornée* μ*-presque partout, alors* f *est intégrable.*

Démonstration. Soit f mesurable et vérifiant $|f| \leq M$ p.p. pour un certain $M > 0$. Comme $|f| = |f| \mathbb{1}_{|f| \leq M} + |f| \mathbb{1}_{|f| > M}$ et que $\mu(\{|f| > M\}) = 0$, on a

$$\int_E |f| d\mu = \int_E |f| \mathbb{1}_{|f| \leq M} d\mu \leq M \mu(|f| \leq M) \leq M \mu(E) < \infty,$$

car μ est finie. □

7.2. **Théorèmes de convergences.** On a déjà vu deux résultats fondamentaux de convergence, le théorème de convergence monotone et le lemme de Fatou. Énonçons maintenant un autre résultat de convergence majeur en théorie de l'intégration.

Théorème 7.1 (Théorème de convergence dominée.)**.** *Soit* $(f_n)_n$ *une suite d'éléments de* $\mathcal{L}^1(\mu)$ *et* f *une fonction mesurable telles que :*

(i) $(f_n)_n$ *converge vers* f μ*-p.p. ;*

(ii) il existe $g \in \mathcal{L}^1(\mu)$ *telle pour tout* n, $|f_n| \leq g$ μ*-p.p.*

Alors $f \in \mathcal{L}^1(\mu)$*, et on a, quand* $n \to \infty$*,*

$$\int_E |f_n - f| d\mu \to 0,$$

et en particulier $\int_E f_n d\mu \to \int_E f d\mu$.

Démonstration. Soit $A = \{\lim_n f_n = f\}$, et pour tout n, $B_n = \{|f_n| \leq g\}$. Par hypothèse, on a $\mu({}^cA) = 0$ et $\mu({}^cB_n) = 0$. Posons,

$$N = A^c \cup \bigcup_n B_n^c.$$

Alors N est μ-négligeable, et pour tout $x \notin N$, $|f(x)| \leq g(x)$, i.e. $|f| \leq g$ μ-p.p., et $f \in \mathcal{L}^1(\mu)$. Pour tout $x \notin N$, posons $h_n(x) = 2g(x) - |f_n(x) - f(x)|$ (et n'importe quoi sur N). Alors $h_n \geq 0$ μ-p.p., $h_n \to 2g$ μ-p.p. et en appliquant le lemme de Fatou à $(h_n)_n$, on obtient :

$$\int_E 2g d\mu = \int_E \liminf_n h_n d\mu \leq \liminf_n \int_E h_n d\mu = \int_E 2g d\mu + \liminf_n \int_E -|f_n - f| d\mu$$

donc

$$-\liminf_n \int_E -|f_n - f| d\mu \leq 0,$$

i.e.

$$\limsup_n \int_E |f_n - f| d\mu \leq 0,$$

et donc $\lim_n \int_E |f_n - f| d\mu = 0$. Comme $|\int_E f_n d\mu - \int_E f d\mu| \leq \int_E |f_n - f| d\mu$, on en déduit que

$$\lim_n \int_E f_n d\mu = \int_E f d\mu. \qquad \square$$

7.3. **Lien avec l'intégrale de Riemann.**

Théorème 7.2. *Soit $f\colon [a,b] \to \mathbb{R}$ une fonction intégrable au sens de Riemann et borélienne. Alors les intégrales de f au sens de Riemann et de Lebesgue coïncident.*

Démonstration. Supposons f positive, le résultat pour f de signe quelconque s'en déduira par la décomposition habituelle $f = f^+ - f^-$.

Soit $\sigma = \{a = x_0 < x_1 < \cdots < x_n = b\}$ une subdivision de $[a,b]$, et considérons les sommes de Darboux :

$$I^+(f,\sigma) = \sum_{k=1}^n (x_k - x_{k-1}) \sup_{[x_{k-1},x_k]} f$$
$$I^-(f,\sigma) = \sum_{k=1}^n (x_k - x_{k-1}) \inf_{[x_{k-1},x_k]} f.$$

Par définition de l'intégrabilité au sens de Riemann, pour tout $\varepsilon > 0$, il existe une subdivision σ_ε, telle que

$$I^+(f,\sigma_\varepsilon) - I^-(f,\sigma_\varepsilon) < \varepsilon.$$

Introduisons les fonctions étagées :

$$f_\varepsilon^+(x) = \sup_{[x_{k-1},x_k]} f \quad \text{pour } x \in [x_{k-1}, x_k]$$
$$f_\varepsilon^-(x) = \inf_{[x_{k-1},x_k]} f \quad \text{pour } x \in [x_{k-1}, x_k].$$

Par définition de l'intégrale de Lebesgue, on a

$$\int_{[a,b]} f_\varepsilon^- d\lambda = I^-(f,\sigma_\varepsilon) \quad \text{et} \quad \int_{[a,b]} f_\varepsilon^+ d\lambda = I^+(f,\sigma_\varepsilon),$$

et comme $f_\varepsilon^- \leq f \leq f_\varepsilon^+$, on a

$$I^-(f,\sigma_\varepsilon) \leq \int_{[a,b]} f d\lambda \leq I^+(f,\sigma_\varepsilon),$$

et donc les deux intégrales coïncident. $\square$

Notation 7.2. On notera alors indifféremment l'intégrale par rapport à la mesure de Lebesgue sur intervalle $[a, b]$

$$\int_{[a,b]} f d\lambda, \quad \int_{[a,b]} f(x) d\lambda(x), \quad \int_{[a,b]} f(x)\lambda(dx), \quad \int_a^b f(x)dx, \quad etc...$$

Exemple 7.2. Soit

$$I_n(\alpha) = \int_0^n \left(1 + \frac{x}{n}\right)^n e^{-\alpha x} dx.$$

On cherche la limite de $I_n(\alpha)$ quand $n \to \infty$. Posons $f_n(x) = \left(1 + \frac{x}{n}\right)^n e^{-\alpha x} \mathbb{1}_{[0,n]}(x)$. Alors, $(f_n)_n$ converge simplement vers la fonction $f(x) = e^{-(1-\alpha)x} \mathbb{1}_{[0,+\infty[}(x)$. Si $\alpha > 1$, alors f est intégrable par rapport à la mesure de Lebesgue, car :

$$\int e^{-(1-\alpha)x} \mathbb{1}_{[0,+\infty[}(x)dx = \lim_{n\to\infty} \int_0^n e^{-(1-\alpha)x} dx = \lim_{n\to\infty} \left[\frac{e^{-(1-\alpha)x}}{\alpha - 1}\right]_0^n = \frac{1}{\alpha - 1}.$$

La première égalité est une conséquence du théorème de convergence monotone, et la seconde, de l'égalité entre les intégrales de Riemann et de Lebesgue. On cherche maintenant une condition de domination par une fonction intégrable. On a, pour $x \geq 0$,

$$\begin{aligned} e^x - \left(1 + \frac{x}{n}\right)^n &= \sum_{k\geq 0} \frac{x^k}{k!} - \sum_{k=0}^n \binom{n}{k} \frac{x^k}{n^k} \\ &\geq \sum_{k=0}^n \left(\frac{1}{k!} - \frac{1}{n^k}\binom{n}{k}\right) x^k \\ &\geq 0, \end{aligned}$$

car

$$\frac{1}{n^k}\binom{n}{k} = \frac{1}{k!} \frac{n(n-1)\cdots(n-k+1)}{n^k} \leq \frac{1}{k!}.$$

Ainsi, $|f_n(x)| \leq e^{-(1-\alpha)x} \mathbb{1}_{[0,+\infty[} \in \mathcal{L}^1$. On a donc par le théorème de convergence dominée que

$$\lim_{n\to\infty} I_n(\alpha) = \int_0^\infty e^{-(1-\alpha)x} dx = \frac{1}{\alpha - 1}, \quad \text{si } \alpha > 1.$$

Si $\alpha \leq 1$, alors f n'est pas intégrable, et on a par le lemme de Fatou,

$$\infty = \int f d\lambda = \int \liminf_n f_n d\lambda \leq \liminf_n \int f_n d\lambda.$$

Ainsi,

$$\lim_{n\to\infty} I_n(\alpha) = \infty \quad \text{si } \alpha \leq 1.$$

7.4. **Mesure à densité.** Soit $(E, \mathcal{A}, \mu)$ un espace mesuré.

Proposition 7.9. *Soit $f : (E, \mathcal{A}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ une fonction mesurable positive. On pose, pour tout $A \in \mathcal{A}$,*

$$\nu(A) = \int_E f \mathbb{1}_A d\mu.$$

Alors ν est une mesure sur $(E, \mathcal{A})$ vérifiant : pour tout $A \in \mathcal{A}$,

$$\mu(A) = 0 \ \Rightarrow \ \nu(A) = 0.$$

On dit que ν est absolument continue par rapport à μ, ce que l'on note $\nu \ll \mu$. On appelle f la densité de ν par rapport à μ, et on note $d\nu = f d\mu$ (ou dérivée de Radon-Nikodym, notée souvent $\frac{d\nu}{d\mu}$).

On a de plus que ν est une mesure finie si et seulement si $f \in \mathcal{L}^1(\mu)$.

Soit $g\colon (E,\mathcal{A}) \to (\mathbb{R},\mathcal{B}(\mathbb{R}))$ *mesurable. Alors* $g \in \mathcal{L}^1(\nu)$ *si et seulement si* $gf \in \mathcal{L}^1(\mu)$*, et dans ce cas*

$$\int_E g d\nu = \int_E g f d\mu.$$

Démonstration. On vérifie aisément que ν est une mesure : clairement, $\nu(\varnothing) = 0$. Soit $(A_n)_n$ une suite d'ensembles mesurables dans $\mathcal{A}$. Alors,

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

l'avant dernière égalité se justifiant par le théorème de convergence monotone. Ainsi, ν est une mesure. Si $\mu(A) = 0$, alors $\mathbb{1}_A = 0$ μ-p.p., et donc $\nu(A) = 0$. Il est clair que ν est finie, de masse $\nu(E) = \int_E f d\mu$ si et seulement $f \in \mathcal{L}^1(\mu)$. La dernière assertion se montre en considérant d'abord g étagée positive, puis en passant à la limite par convergence monotone (comme d'habitude). □

Exemple 7.3. On peut construire ainsi de nouveaux exemples de mesures. Par exemple, par rapport à la mesure de Lebesgue dx sur $\mathbb{R}$:

(i) $d\nu(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx$, la loi normale centrée réduite ;

(ii) $d\nu(x) = \alpha e^{-\alpha x} \mathbb{1}_{]0,+\infty[}(x) dx$, la loi exponentielle de paramètre $\alpha > 0$;

(iii) $d\nu(x) = \frac{1}{\pi} \frac{1}{1+x^2} dx$, la loi de Cauchy.

Vérifiez que ces exemples sont des mesures de probabilités sur $(\mathbb{R},\mathcal{B}(\mathbb{R}))$ (au moins les deux dernières, on verra plus tard comment montrer que l'intégrale gaussienne $\int_{\mathbb{R}} e^{-x^2/2} dx = \sqrt{2\pi}$).

Remarque 7.4. La proposition précédente admet une réciproque partielle, non triviale, le théorème de Radon-Nikodym : si μ et ν sont des mesures finies, telles que ν est absolument continue par rapport à μ, alors il existe $f \in \mathcal{L}^1(\mu)$ telle que $\nu = f d\mu$.

Remarque 7.5. Sur $([0,1],\mathcal{B}([0,1])$ considérons m la mesure de comptage et λ la mesure de Lebesgue. Alors $\lambda \ll m$ car si $m(A) = 0$, alors $A = \varnothing$, et $\lambda(A) = 0$ aussi. Bien sûr, m n'est pas absolument continue par rapport à λ, puisque par exemple, la mesure de Lebesgue d'un singleton est nulle, et pas sa mesure de comptage. De plus, λ n'admet pas de densité par rapport à m bien que $\lambda \ll m$. En effet, supposons par l'absurde qu'une densité existe $f = \frac{d\lambda}{dm}$. Soit $D = \{f \neq 0\}$. Alors pas définition de la densité, on a

$$\int_{[0,1]} f dm = \lambda([0,1]) = 1.$$

De plus, pour tout $\varepsilon > 0$, on aurait donc

$$1 \geq \int_{[0,1]} f \mathbb{1}_{\{f \geq \varepsilon\}} dm \geq \varepsilon m(\{f \geq \varepsilon\}),$$

ce qui implique que $\{f \geq \varepsilon\}$ est fini. Alors, comme $D = \bigcup_{n \geq 1} \{f \geq \frac{1}{n}\}$, on a D qui est dénombrable, et donc $\lambda(D) = 0$. D'où,

$$1 = \lambda(\{f = 0\}) = \int_{[0,1]} \mathbb{1}_{\{f=0\}} f dm = 0,$$

ce qui est absurde !

7.5. **Intégrale des fonctions à valeurs complexes.**

Définition 7.4. *Soit* $f\colon (E,\mathcal{A}) \to (\mathbb{C}, \mathcal{B}(\mathbb{C}))$ *une fonction mesurable à valeurs complexes. On dit que* f *est intégrable par rapport à* μ *si le module* $|f|$ *est intégrable. Ainsi,* $\Re(f)$ *et* $\Im(f)$ *sont intégrables et on définit :*

$$\int_E f d\mu = \int_E \Re(f) d\mu + i \int_E \Im(f) d\mu.$$

Les propriétés de l'intégrale pour les fonctions à valeurs complexes se déduisent de celles pour les fonctions à valeurs réelles via la décomposition en parties réelles et imaginaires.

8. Intégrales à paramètre

Théorème 8.1 (Continuité sous le signe intégrale)**.** *Soit I in intervalle de $\mathbb{R}$ et soit $f\colon I\times E\to\mathbb{R}$ une fonction telle que pour tout $t\in I$, l'application partielle $f(t,\cdot)\colon E\to\mathbb{R}$ est mesurable. On suppose :*

(i) qu'il existe $t_0\in I$ tel que pour μ-presque tout x, $t\mapsto f(t,x)$ est continue en t_0 ;

(ii) qu'il existe $g\colon E\to\mathbb{R}$ μ-intégrable telle que pour μ-presque tout x, pour tout $t\in I$,

$$|f(t,x)|\leq g(x);$$

alors la fonction

$$t\mapsto\int_E f(t,x)d\mu(x)$$

est bien définie et continue en t_0.

Théorème 8.2 (Dérivation sous le signe intégrale)**.** *Soit I in intervalle de $\mathbb{R}$ et soit $f\colon I\times E\to\mathbb{R}$ une fonction telle que pour tout $t\in I$, l'application partielle $f(t,\cdot)\colon E\to\mathbb{R}$ est mesurable. On suppose :*

(i) que pour tout $t\in I$, $f(t,\cdot)$ est intégrable ;

(ii) que pour μ-presque tout x, $t\mapsto f(t,x)$ est dérivable sur I ;

(iii) qu'il existe $g\colon E\to\mathbb{R}$ μ-intégrable telle que pour μ-presque tout x, pour tout $t\in I$,

$$|\frac{\partial f}{\partial t}(t,x)|\leq g(x),$$

alors la fonction

$$h\colon t\mapsto\int_E f(t,x)d\mu(x)$$

est bien définie et est dérivable sur I, de dérivée :

$$h'(t)=\int_E\frac{\partial f}{\partial t}(t,x)d\mu(x),$$

pour tout $t\in I$.

Démonstration. Ces deux théorèmes sont des conséquences directes du théorème de convergence dominée. Montrons le théorème de dérivabilité, celui concernant la continuité est plus facile et laissé en exercice.

Soit $t\in I$ et $(t_n)_n$ une suite de I telle que $t_n\to t$. Posons :

$$f_n(x)=\frac{f(t_n,x)-f(t,x)}{t_n-t}$$

pour $x\in E$. Comme $t\mapsto f(t,x)$ est dérivable pour presque tout x par (i), on a $f_n(x)\to\frac{\partial f}{\partial t}(t,x)$ pour presque tout x. En particulier, $\frac{\partial f}{\partial t}(t,\cdot)$ est presque partout (!) égale à une fonction mesurable. Par l'inégalité des accroissements finis, on a

$$|f_n(x)|\leq\sup_{s\in I}\left|\frac{\partial f}{\partial t}(s,x)\right|$$

et donc,

$$|f_n(x)|\leq g(x),$$

pour μ-presque tout x par (iii). Par le théorème de convergence dominée, on a donc :

$$\int_E f_n(x)d\mu(x)\to\int_E\frac{\partial f}{\partial t}(t,x)d\mu(x)$$

Mais comme,

$$\begin{aligned}\int_E f_n(x)d\mu(x) &= \int_E \left(\frac{f(t_n,x)-f(t,x)}{t_n-t}\right) d\mu(x)\\ &= \frac{1}{t_n-t}\left(\int_E f(t_n,x)d\mu(x) - \int_E f(t,x)d\mu(x)\right)\\ &= \frac{h(t_n)-h(t)}{t_n-t},\end{aligned}$$

on conclut. □

Remarque 8.1. Les deux théorèmes précédents s'étendent immédiatement au cas d'une fonction à valeurs dans $\mathbb{C}$.

9. Compléments : lemme des classes monotones et théorème d'unicité d'une mesure

On commence par introduire deux classes de sous-ensembles satisfaisant des conditions plus faibles que celles d'une tribu.

Définition 9.1. *Une classe Λ de parties d'un ensemble E est appelée une classe monotone (ou un λ-système) si :*

(i) elle contient E ;

(ii) elle est stable par différence propre : pour tous $A, B \in \Lambda$ tel que $A \subset B$, alors $B \setminus A \in \Lambda$;

(iii) elle est stable par réunion dénombrable croissante : pour toute suite croissante $(A_n)_n$ d'éléments de Λ, alors $\bigcup_n A_n \in \Lambda$.

Exemple 9.1. Une tribu est une classe monotone. Un exemple de classe monotone qui n'est pas une tribu est : sur $E = \{1, 2, 3, 4\}$, $\Lambda = \{A \subset E \,|\, \mathrm{Card}(A) \text{ pair}\}$ (exo !).

Définition 9.2. *Une classe Π de parties d'un ensemble E est appelée un π-système si elle est stable par intersections finies : pour tous $A, B \in \Pi$, $A \cap B \in \Pi$.*

Exemple 9.2. L'ensemble des intervalles ouverts $]a, b[$ de $\mathbb{R}$ forme un π-système. Autre exemple, la classe des pavés $\{A \times B \,|\, A, B \subset E\}$ est un π-système sur $E \times E$.

Comme dans le cas des tribus, une intersection quelconque de classes monotones est encore une classe monotone, et si $\mathcal{C}$ est une classe de parties, la classe monotone engendrée par $\mathcal{C}$ est définie par :

$$\Lambda(\mathcal{C}) := \bigcap_{\substack{\Lambda \text{ classe monotone} \\ \mathcal{C} \subset \Lambda}} \Lambda,$$

qui est la plus petite classe monotone contenant $\mathcal{C}$.

Proposition 9.1. *Une classe $\mathcal{A}$ qui est à la fois une classe monotone et un π-système est une tribu.*

Démonstration. Par hypothèse, $E \in \mathcal{A}$. La stabilité par passage au complémentaire est évidente car $E \in \mathcal{A}$, et $\mathcal{A}$ est stable par différence propre (et $A^c = E \setminus A$). Montrons la stabilité par réunion dénombrable. Soit $(A_n)_n$ une suite d'éléments de $\mathcal{A}$. On pose $B_n = \bigcup_{k=0}^n A_k$. Alors la suite $(B_n)_n$ est croissante, et $\bigcup_n B_n = \bigcup_n A_n$. Il suffit donc de montrer que $B_n \in \mathcal{A}$ pour tout n, et d'appliquer la stabilité par réunion dénombrable croissante. Une classe monotone est stable par passage au complémentaire. Un π-système est stable par intersection finie. Donc, comme

$$B_n = \left(\bigcap_{k=0}^n A_k^c \right)^c,$$

B_n est bien dans $\mathcal{A}$ pour tout n. □

Théorème 9.1 (Lemme de la classe monotone)**.** *Pour tout π-système $\mathcal{C}$, la plus petite classe monotone contenant $\mathcal{C}$ est la tribu engendrée par $\mathcal{C}$, c'est-à-dire $\sigma(\mathcal{C})$.*

Avant de démontrer ce théorème (assez subtil...), donnons deux corollaires bien utiles pour montrer l'unicité d'une mesure.

Corollaire 9.1. *Soit μ et ν deux mesures finies sur $(E, \mathcal{A})$, de même masse, et qui coïncident sur un π-système engendrant $\mathcal{A}$. Alors μ et ν coïncident sur $\mathcal{A}$.*

Démonstration. Soit μ et ν deux mesures de même masse sur $\mathcal{A} = \sigma(\mathcal{C})$ où $\mathcal{C}$ est π-système. On introduit la classe

$$\Lambda = \{A \in \mathcal{A} \,|\, \mu(A) = \nu(A)\}.$$

On va montrer que Λ est une classe monotone. On a :

(i) Λ contient E, car $\mu(E) = \nu(E)$ par hypothèse ;

(ii) Soient $A, B \in \Lambda$ tels que $A \subset B$. Comme μ et ν sont finies, on a

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

et donc $B \setminus A \in \Lambda$.

(iii) Soit $(A_n)_n$ une suite de Λ croissante. Alors, par continuité à gauche d'une mesure, on a

$$\mu(\bigcup_n A_n) = \lim_n \mu(A_n) = \lim_n \nu(A_n) = \nu(\bigcup_n A_n),$$

et donc $\bigcup_n A_n \in \Lambda$.

Donc Λ est bien un λ-système. De plus, par hypothèse Λ contient $\mathcal{C}$ car μ et ν coïncident sur $\mathcal{C}$. Donc, $\Lambda(\mathcal{C}) \subset \Lambda$, et par le théorème de la classe monotone, $\Lambda(\mathcal{C}) = \sigma(\mathcal{C})$. Comme $\mathcal{A} = \sigma(\mathcal{C})$, on a $\mathcal{A} \subset \Lambda$, et donc μ et ν coïncident sur $\mathcal{A}$. $\square$

Exemple 9.3. Soit μ une mesure de probabilité sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. On rappelle que la fonction de répartition de μ est la fonction $F\colon \mathbb{R} \to [0,1]$ définie par :

$$F(t) = \mu(]-\infty, t]), \quad \text{pour tout } t \in \mathbb{R}.$$

La fonction de répartition caractérise μ. En effet, comme la classe

$$\mathcal{C} = \{]-\infty, t] \,|\, t \in \mathbb{R}\}$$

est un π-système qui engendre la tribu borélienne $\mathcal{B}(\mathbb{R})$, si deux mesures ont la même fonction de répartition, alors elles sont égales sur $\mathcal{B}(\mathbb{R})$ par le corollaire précédent.

Corollaire 9.2. *Soit μ et ν deux mesures sur $(E, \mathcal{A})$ telles que :*

(i) il existe une suite croissante d'ensembles mesurables $(E_n)_n$ telle que $E = \bigcup_n E_n$;

(ii) pour tout n, $\mu(E_n) = \nu(E_n) < \infty$;

(iii) μ et ν coïncident sur un π-système $\mathcal{C}$ engendrant $\mathcal{A}$ et contenant chaque E_n.

Alors μ et ν coïncident sur $\mathcal{A}$.

Démonstration. Pour tout n, on restreint μ et ν à E_n en posant, pour tout $A \in \mathcal{A}$:

$$\mu_n(A) = \mu(A \cap E_n) \quad \text{et} \quad \nu_n(A) = \nu(A \cap E_n).$$

On applique alors le corollaire 9.1 aux mesures μ_n et ν_n : pour tout $A \in \mathcal{A}$, $\mu_n(A)$ et $\nu_n(A)$ sont finis car $\mu(A \cap E_n) \le \mu(E_n) < \infty$, et de même pour ν_n. De plus, elles coïncident sur le π-système $\mathcal{C}$ qui engendre $\mathcal{A}$ par l'hypothèse (iii). Le corollaire 9.1 entraîne alors que μ_n et ν_n coïncident sur $\mathcal{A}$. Comme la suite $(E_n)_n$ est croissante et que $E = \bigcup_n E_n$, on a

$$\mu(A) = \mu(A \cap \bigcup_n E_n) = \lim_n \mu(A \cap E_n) = \lim_n \mu_n(A),$$

et de même pour ν. Comme $\mu_n(A) = \nu_n(A)$ pour tout $A \in \mathcal{A}$, on conclut. $\square$

Remarquons que le corollaire s'applique notamment à μ et ν σ-finies. Ce corollaire permet alors de montrer l'unicité de la mesure de Lebesgue :

Unicité de la mesure de Lebesgue sur $\mathbb{R}^d$. On considère le π-système $\mathcal{C}$ de l'ensemble des pavés ouverts de $\mathbb{R}^d$:

$$\prod_{k=1}^{d}]a_k, b_k[,$$

avec $a_k < b_k$ dans $\overline{\mathbb{R}}$. On a déjà vu que la tribu borélienne $\mathcal{B}(\mathbb{R}^d)$ est engendrée par $\mathcal{C}$. Supposons qu'il existe deux mesures μ et ν sur $\mathcal{B}(\mathbb{R}^d)$ vérifiant :

$$\mu\left(\prod_{k=1}^{d}]a_k, b_k[\right) = \nu\left(\prod_{k=1}^{d}]a_k, b_k[\right) = \prod_{k=1}^{d} (b_k - a_k).$$

Posons, pour tout $n \geq 1$, $E_n =]-n, n[^d$. Alors, la suite $(E_n)_n$ est croissante, $\bigcup_n E_n = \mathbb{R}^d$, et $\mu(E_n) = \nu(E_n) = (2n)^d$. Le corollaire 9.2 implique alors que μ et ν coïncident sur tout $\mathcal{B}(\mathbb{R}^d)$, et donne donc l'unicité de la mesure de Lebesgue sur $\mathbb{R}^d$. □

Démonstration du théorème de la classe monotone. Soit $\Lambda(\mathcal{C})$, le λ-système engendrée par $\mathcal{C}$. Montrons que $\Lambda(\mathcal{C})$ est une tribu. Par la proposition précédente, il suffit de montrer que c'est aussi un π-système.

Soit $C \in \mathcal{C}$ fixé. Soit Λ_C la classe définie par

$$\Lambda_C = \{A \in \Lambda(\mathcal{C}) \,|\, A \cap C \in \Lambda(\mathcal{C})\}.$$

Montrons que Λ_C est un λ-système contenant $\mathcal{C}$. On a bien $E \in \Lambda_C$ car $E \cap C = C \in \Lambda(\mathcal{C})$. Soit $A \subset B$ dans Λ_C. Alors,

$$(B \setminus A) \cap C = (B \cap C) \setminus (A \cap C),$$

et comme $\Lambda(\mathcal{C})$ est stable par différence propre et que $A \cap C \subset B \cap C$, on a $(B \cap C) \setminus (A \cap C) \in \Lambda(\mathcal{C})$, donc on obtient que Λ_C est stable par différence propre.

Soit $(A_n)_n$ une suite croissante d'éléments de Λ_C. Donc, pour tout n, $A_n \cap C \in \Lambda(\mathcal{C})$. Alors, comme

$$\left(\bigcup_n A_n\right) \cap C = \bigcup_n (A_n \cap C),$$

que $(A_n \cap C)$ est croissante, par stabilité par réunion dénombrable croissante de $\Lambda(\mathcal{C})$, on obtient bien que $(\bigcup_n A_n) \cap C \in \Lambda_C$. Ainsi, Λ_C est un λ-système. De plus, il contient $\mathcal{C}$ car $\mathcal{C}$ est un π-système, donc il est stable par intersection finie. Finalement, $\Lambda(\mathcal{C})$ étant le plus petit λ-système contenant $\mathcal{C}$, on en déduit que $\Lambda(\mathcal{C}) = \Lambda_C$. On a donc obtenu que

$$\forall C \in \mathcal{C},\ \forall A \in \Lambda(\mathcal{C}),\ A \cap C \in \Lambda(\mathcal{C}).$$

Reste à vérifier que l'assertion ci-dessus est en fait encore vrai pour tout $C \in \Lambda(\mathcal{C})$. Soit alors $C \in \Lambda(\mathcal{C})$, et posons encore

$$\Lambda_C = \{A \in \Lambda(\mathcal{C}) \,|\, A \cap C \in \Lambda(\mathcal{C})\}.$$

Par ce qui précède, Λ_C est un λ-système contenant la classe $\mathcal{C}$, donc on a bien $\Lambda_C = \Lambda(\mathcal{C})$. Finalement,

$$\forall C \in \lambda(\mathcal{C}),\ \forall A \in \Lambda(\mathcal{C}),\ A \cap C \in \Lambda(\mathcal{C}),$$

c'est-à-dire que $\Lambda(\mathcal{C})$ est un π-système. Ainsi, $\Lambda(\mathcal{C})$ est une tribu. Elle contient la classe $\mathcal{C}$, donc elle contient aussi la tribu engendrée par $\mathcal{C}$, donc

$$\sigma(\mathcal{C}) \subset \Lambda(\mathcal{C}).$$

Mais comme une tribu est aussi un λ-système, on a finalement $\Lambda(\mathcal{C}) = \sigma(\mathcal{C})$. □

10. Produit d'espaces mesurés

Soient $(E_1, \mathcal{A}_1)$ et $(E_2, \mathcal{A}_2)$ deux espace mesurables.

10.1. Tribu produit, mesure produit.

Définition 10.1. *On appelle tribu produit sur $E_1 \times E_2$, notée $\mathcal{A}_1 \otimes \mathcal{A}_2$, la tribu engendrée par les pavés à côtés mesurables :*

$$\mathcal{A}_1 \otimes \mathcal{A}_2 = \sigma(\mathcal{A}_1 \times \mathcal{A}_2),$$

où $\mathcal{A}_1 \times \mathcal{A}_2 = \{A_1 \times A_2 \,|\, A_1 \in \mathcal{A}_1, A_2 \in \mathcal{A}_2\}$.

L'espace $(E_1 \times E_2, \mathcal{A}_1 \otimes \mathcal{A}_2)$ est appelé espace mesurable produit.

Remarque 10.1. Il est utile de remarquer qu'un pavé s'écrit aussi $A_1 \times A_2 = (A_1 \times E_2) \cap (E_1 \times A_2)$.

En général, $\mathcal{A}_1 \times \mathcal{A}_2$ n'est pas une tribu, le complémentaire d'un pavé n'étant pas en général un pavé.

Dans toute la suite, on note $\pi_1 \colon E_1 \times E_2 \to E_1$ et $\pi_2 \colon E_1 \times E_2 \to E_2$ les projections canoniques sur E_1 et E_2 respectivement, c'est-à-dire $\pi_1(x, y) = x$ et $\pi_2(x, y) = y$, pour tout $(x, y) \in E_1 \times E_2$. Remarquons que π_1 (resp. π_2) est mesurable de $\mathcal{A}_1 \otimes \mathcal{A}_2$ dans $\mathcal{A}_1$ (resp. $\mathcal{A}_2$). En effet, pour tout $A_1 \in \mathcal{A}_1$, $\pi_1^{-1}(A_1) = A_1 \times E_2$ appartient à $\mathcal{A}_1 \times \mathcal{A}_2$, donc à $\mathcal{A}_1 \otimes \mathcal{A}_2$. Même principe pour π_2. On a en fait mieux :

Proposition 10.1. *La tribu produit est aussi la tribu engendrée par les projections canoniques, c'est-à-dire*

$$\mathcal{A}_1 \otimes \mathcal{A}_2 = \sigma\left(\pi_1^{-1}(\mathcal{A}_1) \cup \pi_2^{-1}(\mathcal{A}_2)\right),$$

où $\pi_1 \colon E_1 \times E_2 \to E_1$ et $\pi_2 \colon E_1 \times E_2 \to E_2$ sont les projections canoniques.

Démonstration. Notons $\mathcal{B}$ la tribu engendrée par π_1 et π_2, i.e.

$$\mathcal{B} = \sigma\left(\pi_1^{-1}(\mathcal{A}_1) \cup \pi_2^{-1}(\mathcal{A}_2)\right),$$

qui est par définition la plus petite tribu rendant les applications π_1 et π_2 mesurables, donc la tribu contenant tous les ensembles $\pi_1^{-1}(A_1)$ et $\pi_2^{-1}(A_2)$, pour $A_1 \in \mathcal{A}_1$ et $A_2 \in \mathcal{A}_2$. Or $\pi_1^{-1}(A_1) = A_1 \times E_2$ et $\pi_2^{-1}(A_2) = E_1 \times A_2$, donc $\mathcal{B}$ contient aussi les ensembles de la forme $(A_1 \times E_2) \cap (E_1 \times A_2) = A_1 \times A_2$. Donc $\mathcal{B}$ contient $\mathcal{A}_1 \times \mathcal{A}_2$, et donc $\mathcal{B}$ contient $\mathcal{A}_1 \otimes \mathcal{A}_2$. Mais comme $\mathcal{B} \subset \mathcal{A}_1 \otimes \mathcal{A}_2$, on a bien égalité. □

Proposition 10.2. *Soit $(E, \mathcal{A})$, $(E_1, \mathcal{A}_1)$, $(E_2, \mathcal{A}_2)$ des espaces mesurables et*

$$\begin{aligned} f \colon (E, \mathcal{A}) &\to (E_1 \times E_2, \mathcal{A}_1 \otimes \mathcal{A}_2) \\ x &\mapsto f(x) = (f_1(x), f_2(x)). \end{aligned}$$

Alors f est mesurable si et seulement si f_1 et f_2 sont mesurables.

Démonstration. On a $f_1 = \pi_1 \circ f$ et $f_2 = \pi_2 \circ f$. Donc si f est mesurable, alors f_1 et f_2 sont mesurables comme composées de fonctions mesurables. Réciproquement, soit f_1 et f_2 mesurables, et soit $A_1 \in \mathcal{A}_1$ et $A_2 \in \mathcal{A}_2$. Il suffit de vérifier la mesurabilité de f sur les pavés de la forme $A_1 \times A_2$ par le lemme de transport. Or,

$$f^{-1}(A_1 \times A_2) = f_1^{-1}(A_1) \cap f_2^{-1}(A_2) \in \mathcal{A},$$

car f_1 et f_2 sont mesurables, et donc f est mesurable. □

La tribu borélienne de $\mathbb{R}^2$ (donc la tribu engendrée par les ouverts de $\mathbb{R}^2$) est aussi la tribu produit de $\mathcal{B}(\mathbb{R})$ avec elle même :

Proposition 10.3. *On a :*

$$\mathcal{B}(\mathbb{R}^2) = \mathcal{B}(\mathbb{R}) \otimes \mathcal{B}(\mathbb{R}),$$

et plus généralement,

$$\mathcal{B}(\mathbb{R}^d) = \mathcal{B}(\mathbb{R})^{\otimes d} \quad (:= \underbrace{\mathcal{B}(\mathbb{R}) \otimes \cdots \otimes \mathcal{B}(\mathbb{R})}_{d \text{ fois}} = \mathcal{B}(\mathbb{R})^{d-1} \otimes \mathcal{B}(\mathbb{R})).$$

Démonstration. Soit $\pi_1 \colon \mathbb{R}^2 \to \mathbb{R}$ et $\pi_2 \colon \mathbb{R}^2 \to \mathbb{R}$ les projections canoniques, i.e. $\pi_1(x,y) = x$ et $\pi_2(x,y) = y$. Elles sont clairement continues (pour la topologie produit), et donc borélienne quand on munit $\mathbb{R}^2$ de la tribu $\mathcal{B}(\mathbb{R}^2)$. Comme $\mathcal{B}(\mathbb{R})^{\otimes 2}$ est la plus petite tribu rendant π_1 et π_2 mesurables, on a donc l'inclusion

$$\mathcal{B}(\mathbb{R})^{\otimes 2} \subset \mathcal{B}(\mathbb{R}^2).$$

Comme tout ouvert de $\mathbb{R}^2$ s'écrit comme la réunion *dénombrable* de pavés de la forme $U \times V$ avec U et V des ouverts de $\mathbb{R}$, on a donc que $\mathcal{B}(\mathbb{R})^{\otimes 2}$ contient les ouverts de $\mathbb{R}^2$, et donc contient la tribu borélienne $\mathcal{B}(\mathbb{R}^2)$. □

Remarque 10.2. Plus généralement, si E_1 et E_2 sont deux espaces topologiques, on a

$$\mathcal{B}(E_1) \otimes \mathcal{B}(E_2) \subset \mathcal{B}(E_1 \times E_2).$$

Et si de plus, E_1 et E_2 sont à base dénombrables d'ouverts, alors

$$\mathcal{B}(E_1 \times E_2) = \mathcal{B}(E_1) \otimes \mathcal{B}(E_2).$$

Notation 10.1. Soient $(E_1, \mathcal{A}_1)$ et $(E_2, \mathcal{A}_2)$ deux espaces mesurables. Soit $C \in \mathcal{A}_1 \otimes \mathcal{A}_2$. Pour tous $x_1 \in E_1$ et $x_2 \in E_2$, on note

$$C_{x_1} = \{x_2 \in E_2 \mid (x_1, x_2) \in C\} \quad \text{et} \quad C^{x_2} = \{x_1 \in E_1 \mid (x_1, x_2) \in C\}$$

les sections de C.

Exemples de sections : Si C est un pavé $A_1 \times A_2$, alors

$$(A_1 \times A_2)_{x_1} = \{x_2 \in E_2 \mid (x_1, x_2) \in A_1 \times A_2\} = \begin{cases} A_2 & \text{si } x_1 \in A_1, \\ \varnothing & \text{sinon.} \end{cases}$$

Même principe pour $(A_1 \times A_2)^{x_2}$. Si C est le disque $\{(x,y) \in \mathbb{R}^2 \mid x^2 + y^2 \leq 1\}$, alors

$$C_x = \{y \in \mathbb{R} \mid x^2 + y^2 \leq 1\} = \begin{cases} [-\sqrt{1-x^2}, \sqrt{1-x^2}] & \text{si } x \in [-1,1], \\ \varnothing & \text{sinon.} \end{cases}$$

Soit $f \colon (E_1 \times E_2, \mathcal{A}_1 \otimes \mathcal{A}_2) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ une fonction mesurable. On note

$$\begin{aligned} f_{x_1} \colon (E_2, \mathcal{A}_2) &\to (\mathbb{R}, \mathcal{B}(\mathbb{R})) \\ x_2 &\mapsto f(x_1, x_2) \end{aligned}$$

et

$$\begin{aligned} f^{x_2} \colon (E_1, \mathcal{A}_1) &\to (\mathbb{R}, \mathcal{B}(\mathbb{R})) \\ x_1 &\mapsto f(x_1, x_2) \end{aligned}$$

les applications partielles de f.

Proposition 10.4. *Pour tout $x_1 \in E_1$, et tout $x_2 \in E_2$,*

(i) $C_{x_1} \in \mathcal{A}_2$ *et* $C^{x_2} \in \mathcal{A}_1$.

(ii) f_{x_1} *et* f^{x_2} *sont mesurables.*

Démonstration. (i) Soit $x_1 \in E_1$ et notons $\mathcal{N}_{x_1} = \{A \in \mathcal{A}_1 \otimes \mathcal{A}_2 \,|\, A_{x_1} \in \mathcal{A}_2\}$. Alors $\mathcal{N}_{x_1}$ contient les pavés mesurables : soit $A_1 \in \mathcal{A}_1$ et $A_2 \in \mathcal{A}_2$.

$$(A_1 \times A_2)_{x_1} = \{x_2 \in E_2 \,|\, (x_1, x_2) \in A_1 \times A_2\} = \begin{cases} A_2 & \text{si } x_1 \in A_1 \\ \varnothing & \text{sinon.} \end{cases}$$

Donc $(A_1 \times A_2)_{x_1} \in \mathcal{A}_2$, et $A_1 \times A_2 \in \mathcal{N}_{x_1}$. De plus, $\mathcal{N}_{x_1}$ est une tribu. En effet, on vérifie facilement que le complémentaire d'une section est la section du complémentaire, et de même avec la réunion. Comme $\mathcal{N}_{x_1}$ contient $\mathcal{A}_1 \times \mathcal{A}_2$, c'est donc $\mathcal{A}_1 \otimes \mathcal{A}_2$, ce qui montre que $C_{x_1} \in \mathcal{A}_2$. On fait de même pour C^{x_2}.

(ii) Soit $B \in \mathcal{B}(\mathbb{R})$ un borélien. Alors,

$$f_{x_1}^{-1}(B) = \{x_2 \in E_2 \,|\, f(x_1, x_2) \in B\} = \left\{x_2 \in E_2 \,|\, (x_1, x_2) \in f^{-1}(B)\right\} = f^{-1}(B)_{x_1}$$

qui appartient à $\mathcal{A}_2$ par (i). Donc f_{x_1} est mesurable. On fait de même pour f^{x_2}.

□

Théorème 10.1. *Soit $(E_1, \mathcal{A}_1, \mu_1)$ et $(E_2, \mathcal{A}_2, \mu_2)$ deux espaces mesurés avec μ_1 et μ_2 des mesures σ-finies. Il existe une unique mesure μ sur l'espace produit $(E_1 \times E_2, \mathcal{A}_1 \otimes \mathcal{A}_2)$ telle que*

$$\mu(A_1 \times A_2) = \mu_1(A_1)\mu_2(A_2),$$

pour tous $A_1 \in \mathcal{A}_1$ et $A_2 \in \mathcal{A}_2$. On l'appelle mesure produit et on la note $\mu_1 \otimes \mu_2$. Elle est σ-finie, et pour tout $C \in \mathcal{A}_1 \otimes \mathcal{A}_2$, on a

$$\mu_1 \otimes \mu_2(C) = \int_{E_1} \mu_2(C_{x_1}) d\mu_1(x_1) = \int_{E_2} \mu_1(C^{x_2}) d\mu(x_2).$$

Démonstration. L'unicité découle du lemme des classes monotones. Soit μ et μ' deux mesures qui vérifient la propriété du théorème. Elles sont σ-finies car μ_1 et μ_2 sont σ-finies. De plus, elles coïncident sur le π-système $\mathcal{A}_1 \times \mathcal{A}_2$ qui engendre $\mathcal{A}_1 \otimes \mathcal{A}_2$, et donc coïncident sur tout $\mathcal{A}_1 \otimes \mathcal{A}_2$ par le corollaire 9.2.

Pour l'existence, définissons $m_1 \colon \mathcal{A}_1 \otimes \mathcal{A}_2 \to [0, +\infty]$ par

$$m_1(C) = \int_{E_1} \mu_2(C_{x_1}) d\mu_1(x_1).$$

Commençons par vérifier que l'application m_1 est bien définie : il s'agir de vérifier que l'application $h_C \colon x_1 \mapsto \mu_2(C_{x_1})$ est mesurable. Supposons tout d'abord que μ_2 est finie, et introduisons

$$\Lambda = \{C \in \mathcal{A}_1 \otimes \mathcal{A}_2 \,|\, h_C \text{ est } \mathcal{A}_1\text{-mesurable}\}.$$

On va montrer que Λ est une classe monotone. Comme elle contient le π-système des pavés mesurables car $h_{A_1 \times A_2} = \mu_2(A_2)\mathbb{1}_{A_1}$ qui est mesurable, par le lemme de la classe monotone elle contient $\mathcal{A}_1 \otimes \mathcal{A}_2$. On a :

(i) $E_1 \times E_2 \in \Lambda$, car $(E_1 \times E_2)_{x_1} = E_2$, et donc $h_{E_1 \times E_2} = \mu_2(E_2)$ qui est mesurable car constante.

(ii) Soit $A \subset B$ dans Λ. Donc h_A et h_B sont mesurables, et comme $h_{B \setminus A} = h_B - h_A$ car $(B \setminus A)_{x_1} = B_{x_1} \setminus A_{x_1}$ et $A_{x_1} \subset B_{x_1}$, $h_{B \setminus A}$ est mesurable comme différence de fonctions mesurables. Donc $B \setminus A \in \Lambda$.

(iii) Soit $(C^{(n)})_n$ une suite croissante d'éléments de Λ, et soit $C = \bigcup_n C^{(n)}$. Alors la suite $(C^{(n)}_{x_1})_n$ est croissante (exo !), et on a

$$h_C(x_1) = \mu_2\left(\left(\bigcup_n C^{(n)}\right)_{x_1}\right) = \mu_2\left(\bigcup_n C^{(n)}_{x_1}\right) = \lim_n \mu_2(C^{(n)}_{x_1}) = \lim_n h_{C^{(n)}}(x_1).$$

Donc, $h_C = \lim_n h_{C^{(n)}}$ est mesurable, comme limite de fonctions mesurables, et $\bigcup_n C^{(n)} \in \Lambda$.

Donc Λ est bien une classe monotone, et donc h_C est mesurable pour tout $C \in \mathcal{A}_1 \otimes \mathcal{A}_2$. Si μ_2 est σ-finie, soit $(E^{(n)})$ une suite croissante d'éléments de $\mathcal{A}_2$ telle que $E_2 = \bigcup_n E^{(n)}$ et $\mu_2(E^{(n)}) < \infty$. Alors, $\mu_2(A_2) = \lim_n \mu_2(A_2 \cap E^{(n)})$, et en appliquant ce qui précède à la mesure $\mu_2(\cdot \cap E^{(n)})$ qui est finie, on obtient que $x_1 \mapsto \mu_2(C_{x_1} \cap E^{(n)})$ est mesurable, et donc que h_C est mesurable comme limite de fonctions mesurables.

Montrons maintenant que m_1 est une mesure. Il est clair que $m_1(\varnothing) = 0$ car une section de l'ensemble vide est vide. Soit $(C^{(n)})_n$ une suite d'ensembles mesurables dans $\mathcal{A}_1 \otimes \mathcal{A}_2$ deux à deux disjoints. Comme $(\bigcup_n C^{(n)})_{x_1} = \bigcup_n C^{(n)}_{x_1}$, et que les sections $C^{(n)}_{x_1}$ sont deux à deux disjointes, on a, en utilisant le corollaire du théorème de Beppo-Levi pour échanger somme et intégrale,

$$m_1\left(\bigcup_n C^{(n)}\right) = \int_{E_1} \sum_n \mu_2(C^{(n)}_{x_1}) d\mu_1(x_1) = \sum_n \int_{E_1} \mu_2(C^{(n)}_{x_1}) d\mu_1(x_1) = \sum_n m_1(C^{(n)}),$$

donc m_1 est σ-additive. Donc m_1 est une mesure, et sur les pavés mesurables, on a

$$m_1(A_1 \times A_2) = \int_{E_1} \mu_2\left((A_1 \times A_2)_{x_1}\right) d\mu_1(x_1) = \int_{E_1} \mu_2(A_2) \mathbb{1}_{A_1}(x_1) d\mu_1(x_1) = \mu_1(A_1)\mu(A_2).$$

Il existe donc bien une mesure m_1 vérifiant $m_1(A_1 \times A_2) = \mu_1(A_1)\mu(A_2)$. On fait de même avec la mesure m_2 définie par

$$m_2(C) = \int_{E_2} \mu_1(C^{x_2}) d\mu_2(x_2),$$

ce qui conclut la preuve du théorème. □

Comme corollaire immédiat, on a :

Corollaire 10.1. *La mesure de Lebesgue λ_d sur $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ est aussi la mesure produit $\lambda^{\otimes d}$.*

Remarque. Le théorème est faux si μ_1 ou μ_2 ne sont pas σ-finies. Prendre par exemple, μ_1 la mesure de Lebesgue, et μ_2 la mesure de comptage sur $\mathbb{R}$ qui n'est pas σ-finie.

10.2. **Théorèmes de Fubini.**

Théorème 10.2 (Théorème de Fubini-Tonelli)**.** *Soit $f\colon (E_1 \times E_1, \mathcal{A}_1 \otimes \mathcal{A}_2) \to [0, +\infty]$ une fonction mesurable, positive. Alors, les fonctions*

$$\begin{array}{rcl} (E_1, \mathcal{A}_1) & \to & [0,+\infty] \\ x_1 & \mapsto & \displaystyle\int_{E_2} f(x_1,x_2) d\mu_2(x_2) \end{array} \quad \text{et} \quad \begin{array}{rcl} (E_2, \mathcal{A}_2) & \to & [0,+\infty] \\ x_2 & \mapsto & \displaystyle\int_{E_1} f(x_1,x_2) d\mu_1(x_1) \end{array}$$

sont mesurables, et on a :

$$\int_{E_1 \times E_2} f d\mu_1 \otimes \mu_2 = \int_{E_1} \left(\int_{E_2} f(x_1,x_2) d\mu_2(x_2)\right) d\mu_1(x_1) = \int_{E_2} \left(\int_{E_1} f(x_1,x_2) d\mu_1(x_1)\right) d\mu_2(x_2).$$

Remarque 10.3. Le théorème de Fubini-Tonelli assure donc que l'intégrale de f par rapport à la mesure produit $\mu_1 \otimes \mu_2$ peut se calculer en faisant deux intégrales simples successives dans l'ordre que l'on veut. Les égalités ayant lieu dans $[0, +\infty]$, il n'y a pas de précaution particulière à prendre.

Démonstration. Le théorème est vrai pour les fonctions indicatrices par le théorème 10.1 donnant l'existence de la mesure produit. Par linéarité de l'intégrale, il est vrai pour les fonctions étagées, puis par convergence monotone pour les fonctions mesurables positives. □

Théorème 10.3 (Théorème de Fubini-Lebesgue)**.** *Soit $f\colon (E_1 \times E_1, \mathcal{A}_1 \otimes \mathcal{A}_2) \to \mathbb{R}$ une fonction mesurable $\mu_1 \otimes \mu_2$-intégrable. Alors, les fonctions*

$$\begin{aligned} (E_1, \mathcal{A}_1) &\to \mathbb{R} \\ x_1 &\mapsto \int_{E_2} f(x_1, x_2) d\mu_2(x_2) \end{aligned} \quad \text{et} \quad \begin{aligned} (E_2, \mathcal{A}_2) &\to \mathbb{R} \\ x_2 &\mapsto \int_{E_1} f(x_1, x_2) d\mu_1(x_1) \end{aligned}$$

sont respectivement μ_1-intégrable et μ_2-intégrable, et on a :

$$\int_{E_1 \times E_2} f d\mu_1 \otimes \mu_2 = \int_{E_1} \left(\int_{E_2} f(x_1, x_2) d\mu_2(x_2) \right) d\mu_1(x_1) = \int_{E_2} \left(\int_{E_1} f(x_1, x_2) d\mu_1(x_1) \right) d\mu_2(x_2).$$

Remarque 10.4. Pour appliquer le théorème de Fubini-Lebesgue, il faut cette fois d'abord vérifier que f est intégrable par rapport à $\mu_1 \otimes \mu_2$, ce qui se fait en appliquant le théorème de Fubini-Tonelli à $|f|$, et en vérifiant que l'une des deux intégrales doubles est finie.

Démonstration. Le théorème se vérifie aisément en appliquant Fubini-Tonelli à f^+ et f^- (exo !). □

11. Mesure image et théorème de transfert

Définition 11.1. *Soit $f\colon (E,\mathcal{A}) \to (F,\mathcal{B})$ une fonction mesurable et μ une mesure sur $(E,\mathcal{A})$. Alors, l'application $\nu\colon \mathcal{B} \to [0,+\infty]$ définie par*

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

est une mesure sur $(F,\mathcal{B})$, appelée mesure image de μ par f. On la note parfois μ_f.

Vérifions que ν définit bien une mesure :

(i) $\nu(\varnothing) = \mu(f^{-1}(\varnothing)) = \mu(\varnothing) = 0$;

(ii) soit $(B_n)_n$ une suite d'éléments de $\mathcal{B}$, deux à deux disjoints. Alors, comme $f^{-1}(B_i \cap B_j) = f^{-1}(B_i) \cap f^{-1}(B_j)$, les $f^{-1}(B_n)$ sont deux à deux disjoints, et

$$\nu\left(\bigcup_n B_n\right) = \mu\left(f^{-1}\left(\bigcup_n B_n\right)\right) = \mu\left(\bigcup_n f^{-1}\left(B_n\right)\right) = \sum_n \mu\left(f^{-1}\left(B_n\right)\right) = \sum_n \nu(B_n).$$

Le théorème suivant est fondamental en théorie des probabilités.

Théorème 11.1 (Théorème de transfert)**.** *Soit μ_h la mesure image d'une mesure μ sur $(E,\mathcal{A})$ par une application mesurable h de $(E,\mathcal{A})$ dans $(F,\mathcal{B})$. Alors, pour f mesurable de $(F,\mathcal{B})$ dans $(\mathbb{R},\mathcal{B}(\mathbb{R}))$, on a*

$$\int_F f d\mu_h = \int_E f \circ h d\mu.$$

Démonstration. Si $f = \mathbb{1}_B$ où $B \in \mathcal{B}$, alors

$$\int_E \mathbb{1}_B \circ h(x) d\mu(x) = \int_E \mathbb{1}_{h^{-1}(B)}(x) d\mu(x) = \mu(h^{-1}(B)) = \mu_h(B) = \int_F \mathbb{1}_B d\mu_h$$

et le résultat est vrai pour les fonctions indicatrices. Par linéarité de l'intégrale, c'est encore vrai pour les fonctions étagées, puis par convergence monotone pour les fonctions mesurables positives. Le théorème pour les fonctions mesurables de signe quelconque s'en déduit en utilisant la décomposition $f = f^+ - f^-$. $\square$

Exemple 11.1. Soit λ la mesure de Lebesgue, et τ_a la translation par a, i.e. $\tau_a(x) = x+a$. Alors, par invariance par translation de la mesure de Lebesgue, on a $\lambda_{\tau_a} = \lambda$, et donc

$$\int_{\mathbb{R}} f(x+a)dx = \int_{\mathbb{R}} f(x)dx.$$

Nous admettrons le théorème suivant bien utile pour le calcul intégral, mais dont la démonstration est assez technique.

Théorème 11.2 (Formule du changement de variables)**.** *Soit U et V deux ouverts de $\mathbb{R}^d$, et $\phi\colon U \to V$ un C^1-difféomorphisme (i.e. ϕ est bijective, de classe C^1, et tel que ϕ^{-1} est C^1). Alors, pour toute fonction f positive ou intégrable,*

$$\int_U f \circ \phi(x) d\lambda(x) = \int_V f(y) |\det J_{\phi^{-1}}(y)| d\lambda(y),$$

où $J_{\phi^{-1}}$ est la matrice jacobienne de ϕ^{-1} (i.e. la matrice $d \times d$ des dérivées partielles des composantes de ϕ^{-1}). De manière équivalente, on a

$$\int_V f(y) d\lambda(y) = \int_U f \circ \phi(x) |\det J_\phi(x)| d\lambda(x).$$

Démonstration. Admis, voir [1] pour une preuve. $\square$

Remarque 11.1. En dimension 1, on retrouve la formule du changement variable usuelle :

$$\int_a^b f(x)dx = \int_{\phi^{-1}(a)}^{\phi^{-1}(b)} f(\phi(t))\phi'(t)dt.$$

Exemple 11.2. Soit $A \in GL_d(\mathbb{R})$ et $b \in \mathbb{R}^d$. Soit ϕ l'application affine définie par $\phi(x) = Ax + b$, pour tout $x \in \mathbb{R}^d$. Alors, pour tout $f \in \mathcal{L}^1(\mu)$,

$$\int_{\mathbb{R}^d} f \circ \phi(x) dx = \frac{1}{|\det A|} \int_{\mathbb{R}^d} f(x) dx.$$

En effet, la i^{e} coordonnée de $\phi(x)$ est $\sum_{j=1}^d a_{ij} x_j + b_i$, donc

$$\frac{\partial \phi(x)_i}{\partial x_j} = a_{ij},$$

et donc $J_\phi = A$.

Remarque 11.2. On rappelle l'énoncé du théorème d'inversion globale (corollaire immédiat du théorème d'inversion locale) :

Théorème 11.3. *Soient U et V deux ouverts de $\mathbb{R}^d$ et soit $\phi \colon U \to V$ une application. On suppose que ϕ est bijective, de classe C^1 et telle que la matrice jacobienne J_ϕ de ϕ est inversible. Alors ϕ est un C^1-difféomorphisme de U dans V.*

Exemple 11.3 (Coordonnées polaires)**.** Soit l'application bijective

$$\begin{aligned} \phi^{-1} \colon]0, +\infty[\times] - \pi, \pi[&\to \mathbb{R}^2 \setminus (\mathbb{R}_- \times \{0\}) \\ (r, \theta) &\mapsto (r \cos \theta, r \sin \theta). \end{aligned}$$

Sa matrice jacobienne est

$$J_{\phi^{-1}}(r, \theta) = \begin{pmatrix} \cos \theta & -r \sin \theta \\ \sin \theta & r \cos \theta \end{pmatrix}$$

de déterminant $\det J_{\phi^{-1}}(r, \theta) = r > 0$, et le théorème d'inversion globale assure que ϕ est un C^1-difféomorphisme. Ainsi, pour une fonction borélienne intégrable par rapport à la mesure de Lebesgue sur $\mathbb{R}^2$, on a

$$\begin{aligned} \int_{\mathbb{R}^2} f(x, y) dx dy &= \int_{\mathbb{R}^2 \setminus (\mathbb{R}_- \times \{0\})} f(x, y) dx dy \\ &= \int_{]0, +\infty[\times] - \pi, \pi[} f(r \cos \theta, r \sin \theta) r dr d\theta. \end{aligned}$$

En particulier, on peut calculer l'intégrale gaussienne comme suit, en appliquant le changement en coordonnées polaires à la fonction $f(x, y) = e^{-(x^2+y^2)/2}$ et le théorème de Fubini-Tonelli. On a

$$\int_{\mathbb{R}^2} e^{-(x^2+y^2)/2} dx dy = \int_{]0, +\infty[\times] - \pi, \pi[} e^{-r^2/2} r dr d\theta = 2\pi \int_{]0, +\infty[} r e^{-r^2/2} dr = 2\pi.$$

Comme

$$\int_{\mathbb{R}^2} e^{-(x^2+y^2)/2} dx dy = \left(\int_{\mathbb{R}} e^{-x^2/2} dx \right)^2,$$

encore par Fubini-Tonelli, on en déduit l'égalité bien connue :

$$\int_{\mathbb{R}} e^{-x^2/2} dx = \sqrt{2\pi}.$$

Remarque 11.3. Soient μ la mesure de densité $f \mathbb{1}_U$ par rapport à la mesure de Lebesgue λ sur $\mathbb{R}^d$, et $\phi \colon U \to V$ un C^1-difféomorphisme, alors la mesure image de μ par ϕ est la mesure de densité g par rapport à λ donnée par

$$g(y) = f \circ \phi^{-1}(y) |\det J_{\phi^{-1}}(y)| \mathbb{1}_V(y).$$

En effet, on a, pour toute fonction h mesurable positive,

$$\begin{aligned}\int h d\mu_\phi &= \int h \circ \phi(x) d\mu(x) = \int h \circ \phi(x) f(x) \mathbb{1}_U(x) d\lambda(x) \\ &= \int h(y) f \circ \phi^{-1}(y) |\det J_{\phi^{-1}}(y)| \mathbb{1}_V(y) d\lambda(y),\end{aligned}$$

par la formule du changement de variables. On obtient donc l'assertion en prenant $h = \mathbb{1}_B$.

12. Espaces L^p

Définition 12.1. *Soit $p \in [1,+\infty[$. On définit l'espace $\mathcal{L}^p(E,\mathcal{A},\mu)$, abrégé en $\mathcal{L}^p(\mu)$, l'ensemble de toutes les fonctions mesurables f à valeurs dans $\overline{\mathbb{R}}$ telles que $|f|^p$ soit intégrable par rapport à μ. Pour $f \in \mathcal{L}^p(\mu)$, on pose :*

$$\|f\|_p = \left(\int_E |f|^p d\mu\right)^p.$$

Définition 12.2. *On définit l'espace $\mathcal{L}^\infty(E,\mathcal{A},\mu)$, abrégé en $\mathcal{L}^\infty(\mu)$, l'ensemble de toutes les fonctions mesurables f à valeurs dans $\overline{\mathbb{R}}$ qui sont μ-essentiellement bornées, c'est-à-dire qu'il existe $a > 0$ tel que $\mu(\{|f| > a\}) = 0$. Pour $f \in \mathcal{L}^\infty(\mu)$, on pose*

$$\|f\|_\infty = \inf\Big\{a > 0 \,|\, \mu(\{|f| > a\}) = 0\Big\},$$

qu'on appelle supremum essentiel de f.

Le supremum essentiel est donc le plus petit des majorants de f en dehors d'un ensemble négligeable. En particulier, on a donc $|f| \leq \|f\|_\infty$ μ-p.p.

On remarquera aussi que $||f||_\infty \leq \sup_{x\in E} |f(x)|$. Si de plus $f\colon (\mathbb{R},\mathcal{B}(\mathbb{R}),\lambda) \to (\mathbb{R},\mathcal{B}(\mathbb{R}))$ est une fonction continue, alors $||f||_\infty = \sup_{x\in E} |f(x)|$ (ce qui justifie l'usage de la notion $||\cdot||_\infty$ utilisée pour le supremum habituellement). En effet, supposons que $||f||_\infty < \sup_{x\in E} |f(x)|$, alors il existe $x \in \mathbb{R}$ tel que $|f(x)| > ||f||_\infty$, donc l'ensemble $\{|f| > ||f||_\infty\}$ est non vide et ouvert par continuité de f. Il est donc de mesure de Lebesgue non nulle, c'est-à-dire $\lambda(\{|f| > ||f||_\infty\}) > 0$, ce qui contredit la définition du sup essentiel.

Remarque 12.1. Les fonctions de $\mathcal{L}^\infty(\mu)$ se caractérisent à l'aide des fonctions bornées usuelles : si $f \in \mathcal{L}^\infty(\mu)$, alors il existe une fonction g mesurable et bornée telle que $f = g$ μ-p.p. et on a $||f||_\infty = \sup_{x\in E} |g(x)|$. En effet, il suffit de poser $g = f\mathbb{1}_{\{|f|\leq ||f||_\infty\}}$. On a alors que $g = f$ μ-p.p., et que pour tout $x \in E$, $|g(x)| \leq ||f||_\infty$ (car $g(x) = 0$ sur $\{|f| > ||f||_\infty\}$). Donc, $\sup_{x\in E} |g(x)| \leq ||f||_\infty$. Et comme $g = f$ μ-p.p. implique que $||g||_\infty = ||f||_\infty$, et que $||g||_\infty \leq \sup_{x\in E} |g(x)|$, on a bien $||f||_\infty = \sup_{x\in E} |g(x)|$.

Définition 12.3. *Soit p et q dans $[1,+\infty]$. On dit que p et q sont des exposants conjugués si $\frac{1}{p} + \frac{1}{q} = 1$.*

Par exemple, 1 et $+\infty$ sont des exposants conjugués.

Proposition 12.1 (Inégalités de Hölder)**.** *Soit p, q des exposants conjugués. Alors, pour tout $f \in \mathcal{L}^p(\mu)$ et tout $g \in \mathcal{L}^q(\mu)$, on a $fg \in \mathcal{L}^1(\mu)$, et*

$$\|fg\|_1 \leq \|f\|_p \|g\|_q.$$

Démonstration. On commence par le cas $p = 1$ et $q = \infty$. Par definition du sup essentiel, on a $|g| \leq \|g\|_\infty$ μ-p.p., donc

$$|fg| \leq |f|\, \|g\|_\infty \quad \mu\text{-p.p.},$$

et en intégrant par rapport à μ, on obtient

$$\|fg\|_1 \leq \|f\|_1 \|g\|_\infty.$$

Soit maintenant $p, q \in]1,+\infty[$. Si $\|f\|_p$ ou $\|g\|_q$ est nulle, alors f ou g est nulle μ-p.p., et l'inégalité est évidente. Supposons donc que $\|f\|_p > 0$ et $\|g\|_q > 0$. Quitte à remplacer f par $\frac{f}{\|f\|_p}$ et g par $\frac{g}{\|g\|_q}$, on peut supposer que $\|f\|_p = \|g\|_q = 1$. On utilise alors le lemme suivant :

Lemme 12.1 (Inégalité de Young)**.** Soit $p, q \in]1, +\infty[$ tel que $\frac{1}{p} + \frac{1}{q} = 1$, et a et b deux réels positifs. Alors,

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

avec égalité si et seulement si $a^p = b^q$.

Finissons la preuve de la proposition. En utilisant l'inégalité de Young, on a

$$|fg| \leq \frac{|f|^p}{p} + \frac{|g|^q}{q},$$

et en intégrant par rapport à μ, on obtient, puisque $\|f\|_p = \|g\|_q = 1$ et que p et q sont conjugués,

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

ce qui démontre la proposition. □

Démonstration de l'inégalité de Young. L'inégalité est trivial si $ab = 0$, on peut donc supposer que $a, b > 0$. On utilise la concavité du logarithme :

$$\forall \alpha \in [0,1], \forall x, y > 0, \log\Big(\alpha x + (1-\alpha)y\Big) \geq \alpha \log(x) + (1-\alpha)\log(y).$$

En prenant $\alpha = \frac{1}{p}$ (et donc $1 - \alpha = \frac{1}{q}$), $x = a^p$, $y = b^q$, puis en prenant l'exponentielle, on obtient l'inégalité cherchée (le cas d'égalité provient de la stricte concavité du log). □

Proposition 12.2 (Inégalité de Minkowski)**.** *Soit $p \in [1, +\infty]$. Si $f, g \in \mathcal{L}^p(\mu)$, alors $f + g \in \mathcal{L}^p(\mu)$, et on a*

$$\|f+g\|_p \leq \|f\|_p + \|g\|_p .$$

Démonstration. Commençons par le cas $p = \infty$. Par définition de $\mathcal{L}^\infty$, on a $|f| \leq \|f\|_\infty$ μ-p.p. et $|g| \leq \|g\|_\infty$ μ-p.p., donc par l'inégalité triangulaire, $|f + g| \leq \|f\|_\infty + \|g\|_\infty$ μ-p.p., i.e. $\mu\left(\{|f+g| > \|f\|_\infty + \|g\|_\infty\}\right) = 0$. Par définition du supremum essentiel, on a alors que

$$\|f+g\|_\infty \leq \|f\|_\infty + \|g\|_\infty .$$

Le cas $p = 1$ découle de l'inégalité triangulaire : $|f + g| \leq |f| + |g|$, et donc en intégrant par rapport à μ, on obtient bien

$$\|f+g\|_1 \leq \|f\|_1 + \|g\|_1 .$$

Supposons maintenant $p > 1$. En utilisant le fait que $x \mapsto x^p$ est convexe, on a

$$|\frac{1}{2}f + \frac{1}{2}g|^p \leq \frac{1}{2}|f|^p + \frac{1}{2}|g|^p,$$

ce qui donne :

$$|f+g|^p \leq 2^{p-1}\left(|f|^p + |g|^p\right).$$

En intégrant par rapport à μ, on obtient donc que $f + g \in \mathcal{L}^p(\mu)$ car f et g sont dans $\mathcal{L}^p(\mu)$. Par l'inégalité triangulaire, on a

$$|f+g|^p = |f+g||f+g|^{p-1} \leq |f||f+g|^{p-1} + |g||f+g|^{p-1}$$

Alors, en intégrant, on a

$$\int_E |f+g|^p d\mu \leq \int_E |f||f+g|^{p-1} d\mu + \int_E |g||f+g|^{p-1} d\mu.$$

On utilise alors l'inégalité de Hölder appliqué à $|f|$ (resp. $|g|$) et $|f+g|^{p-1}$ avec les exposants conjugués p et $q = \frac{p}{p-1}$. On a bien que $|f+g|^{p-1} \in \mathcal{L}^q(\mu)$ car $f+g \in \mathcal{L}^p(\mu)$. On obtient alors :

$$\begin{aligned}\int_E |f+g|^p d\mu &\leq \left(\int_E |f|^p d\mu\right)^p \left(\int_E \left(|f+g|^{p-1}\right)^{\frac{p}{p-1}}\right)^{\frac{p-1}{p}} + \left(\int_E |g|^p d\mu\right)^p \left(\int_E \left(|f+g|^{p-1}\right)^{\frac{p}{p-1}}\right)^{\frac{p-1}{p}} \\ &= \left(\int_E |f|^p d\mu\right)^p \left(\int_E (|f+g|^p)\right)^{\frac{p-1}{p}} + \left(\int_E |g|^p d\mu\right)^p \left(\int_E \left(|f+g|^{p-1}\right)^{\frac{p}{p-1}}\right)^{\frac{p-1}{p}}\end{aligned}$$

donc,

$$\|f+g\|_p^p \leq \|f\|_p \|f+g\|_p^{p-1} + \|g\|_p \|f+g\|_p^{p-1}$$

Comme l'inégalite de Minkowski est triviale si $\|f+g\|_p = 0$, on peut toujours supposer que $\|f+g\|_p > 0$, et donc en simplifiant, on obtient

$$\|f+g\|_p \leq \|f\|_p + \|g\|_p .$$

□

Les espaces $\mathcal{L}^p(\mu)$ sont donc des $\mathbb{R}$-espaces vectoriels, puisque si $f, g \in \mathcal{L}^p(\mu)$ et que $\alpha \in \mathbb{R}$, alors $\alpha f + g \in \mathcal{L}^p(\mu)$. Ils sont de plus pas loin d'être normé pour la "norme" $\|\cdot\|_p$: l'homogénéité est évidente, et l'inégalité de Minkowski est exactement l'inégalité triangulaire pour $\|\cdot\|_p$. Par contre, $\|f\|_p = 0$ n'implique pas que $f = 0$, mais que $f = 0$ μ-presque partout ! Les espaces $\mathcal{L}^p(\mu)$ sont des espaces vectoriels dit *semi-normés*. Un procédé classique pour modifier une semi-norme en une norme est d'identifier deux fonctions vérifiant $\|f-g\|_p = 0$.

Plus précisément, soit $\sim$ la relation d'équivalence définie par

$$f \sim g \Leftrightarrow f = g \quad \mu\text{-p.p.} \quad (\Leftrightarrow \mu(\{f \neq g\}) = 0)$$

Elle est clairement réflexive ($f \sim f$), symétrique ($f \sim g \Leftrightarrow g \sim f$), et transitive ($f \sim g$ et $g \sim h$ implique que $f \sim h$). On remarque de plus que $\|f-g\|_p = 0$ équivaut à $f = g$ μ-presque partout ce qui est indépendant de p. On a alors :

Définition 12.4. *Soit $p \in [1, +\infty]$. On définit l'espace $L^p(E, \mathcal{A}, \mu)$, abrégé en $L^p(\mu)$, comme l'ensemble quotient*

$$L^p(\mu) = \mathcal{L}^p(\mu)/\sim$$

de $\mathcal{L}^p(\mu)$ par la relation d'équivalence $\sim$ définie par l'égalité μ-p.p., c'est-à-dire le sous-ensemble de $\mathcal{L}^p(\mu)$ des classes d'équivalence de $\sim$.

On a donc qu'un élément f de $L^p(\mu)$, est une classe d'équivalence $f = \{g \in L^p(\mu) \,|\, g = f \ \mu\text{-p.p.}\}$. *On identifiera toujours fonctions et classes d'équivalence.* C'est-à-dire que $f \in L^p(\mu)$ sous entend n'importe quelle fonction de $\mathcal{L}^p(\mu)$ égale à f μ-presque partout.

On a alors immédiatemment :

Théorème 12.1. *L'espace $L^p(\mu)$ est un espace vectoriel normé pour la norme $|| \cdot ||_p$.*

On a même mieux :

Théorème 12.2 (Théorème de Riesz-Fischer)**.** *L'espace $L^p(\mu)$ est un espace de Banach, c'est-à-dire un espace vectoriel normé complet.*

Remarque 12.2. Dire que $f_n \to f$ dans L^p, i.e. $||f_n - f||_p \to 0$, c'est dire que

$$\lim_{n\to\infty} \int_E |f_n - f|^p d\mu = 0.$$

Dire que $f_n \to f$ dans L^∞, i.e. $||f_n - f||_\infty \to 0$, c'est donc dire que pour tout $\varepsilon > 0$, il existe n_0 tel que pour tout $n \geq n_0$,

$$||f_n - f||_\infty < \varepsilon,$$

et donc qu'il existe un ensemble négligeable N tel que pour tout $x \in N^c$,

$$|f_n(x) - f(x)| < \varepsilon.$$

Donc (f_n) converge uniformément vers f sur N^c, et en particulier converge simplement vers f sur N^c.

Montrons d'abord deux lemmes qui serviront pour la démonstration du théorème. Le premier énonce que pour toute suite de Cauchy, on peut extraire une sous-suite qui est "rapidement" de Cauchy, et le second donne une caractérisation des espaces de Banach.

Lemme 12.2. Soit $(E, \|\cdot\|)$ un espace vectoriel normé, et $(u_n)_n$ une suite de Cauchy dans E. Alors il existe une suite extraite $(u_{n_k})_k$ tel que pour tout $k \geq 1$,

$$\left\|u_{n_{k+1}} - u_{n_k}\right\| \leq \frac{1}{2^k}.$$

Démonstration. Par définition d'une suite de Cauchy, on a que pour tout $\varepsilon > 0$, il existe $n_0 > 0$, tel que pour tout $p \geq n_0$, pour tout $q \geq n_0$,

$$\|u_p - u_q\| < \varepsilon.$$

On construit la suite $(n_k)_{k\geq 1}$ par récurrence.

Soit $\varepsilon = \frac{1}{2}$. Choisissons $n_1 \geq 0$ tel que pour tout $p, q \geq n_1$, $\|u_p - u_q\| \leq \frac{1}{2}$.

Soit maintenant $\varepsilon = (\frac{1}{2})^2$. Il existe n_2' tel que $||u_p - u_q|| \leq (\frac{1}{2})^2$, pour tout $p, q \geq n_2'$. Posons alors $n_2 = \max\{n_2', n_1 + 1\}$. Alors $n_2 > n_1$ et est tel que pour tout $p, q \geq n_2$, $\|u_p - u_q\| \leq \left(\frac{1}{2}\right)^2$.

Et ainsi de suite. Supposons n_k construit. On choisit alors $n_{k+1} > n_k$ tel que pour tout $p, q \geq n_{k+1}$, $\|u_p - u_q\| \leq \left(\frac{1}{2}\right)^{k+1}$.

On a donc bien que pour tout $k \geq 1$,

$$\left\|u_{n_{k+1}} - u_{n_k}\right\| \leq \left(\frac{1}{2}\right)^k,$$

car $n_{k+1} > n_k$. □

Lemme 12.3. Soit $(E, \|\cdot\|)$ un espace vectoriel normé. Si toute série de terme général u_n telle que $\sum_n \|u_n\| < +\infty$ est convergente (dans E !), alors E est complet.

Remarque 12.3. Attention à bien comprendre l'énoncé du lemme : si pour toute suite $(u_n)_n$ d'éléments de E vérifiant $\sum_n ||u_n|| < \infty$ (dans $\mathbb{R}$), il existe $u \in E$ telle que $\lim_{n\to\infty} || \sum_{k=0}^n u_k - u|| = 0$, alors E est complet.

Remarque 12.4. C'est en fait une équivalence, mais le sens direct, i.e. dans un espace de Banach toute série normalement convergente est convergente, est facile (exo).

Démonstration. Soit $(x_n)_n$ une suite de Cauchy dans E. Par le lemme précédent, il existe une suite extraite $(x_{n_k})_k$, telle que pour tout k, $\left\|x_{n_{k+1}} - x_{n_k}\right\| \leq (\frac{1}{2})^k$. Posons $u_k = x_{n_{k+1}} - x_{n_k}$. Alors la suite $(u_k)_k$ vérifie

$$\sum_k \|u_k\| \leq \sum_k \left(\frac{1}{2}\right)^k < \infty.$$

Par hypothèse, on a donc que la série $\sum_k u_k$ est convergente dans E, i.e. la suite des sommes partielles $\sum_{k=0}^n u_k$ converge en norme vers $\sum_k u_k$. Mais, comme

$$x_{n_k} = x_{n_0} + \sum_{j=0}^{k-1}(x_{n_{j+1}} - x_{n_j}) = x_{n_0} + \sum_{j=0}^{k-1} u_j,$$

on a donc que la suite $(x_{n_k})_k$ converge aussi dans E. La suite $(x_n)_n$ est donc une suite de Cauchy qui admet une sous-suite convergente, c'est donc qu'elle converge, et finalement E est complet. $\square$

Démonstration du théorème de Riesz-Fischer. On utilise le lemme précédent : on va montrer que toute série de terme général f_n telle que $\sum_n \|f\|_p < \infty$ est convergente (dans L^p).

Commençons par le cas $p \in [1, +\infty[$. Soit donc $(f_n)_n$ une suite de L^p telle que $\sum_n \|f_n\|_p = M < \infty$. Notons

$$h_n = \sum_{k=0}^{n} |f_k|.$$

La suite $(h_n)_n$ est positive, croissante et converge simplement vers la fonction $h = \sum_n |f_n|$. Par l'inégalité de Minkowski, on a

$$\|h_n\|_p \leq \sum_{k=0}^{n} \|f_k\|_p \leq M.$$

Donc $h_n \in L^p$. Par le théorème de convergence monotone,

$$\lim_n \int_E h_n^p d\mu = \int_E h^p d\mu \leq M^p,$$

donc $\|h\|_p \leq M$, et $h \in L^p$. Donc, h est fini μ-presque partout, i.e. pour μ-presque tout x, $\sum_n |f_n(x)| < \infty$. Ainsi, en dehors d'un ensemble négligeable, la série de terme général $f_n(x)$ est absolument convergente dans $\mathbb{R}$, et donc convergente, i.e.

$$\sum_n f_n(x) < \infty, \quad \mu\text{-presque partout.}$$

Définissons f par $f(x) = \sum_n f_n(x)$, en dehors d'un ensemble négligeable. On a donc

$$\sum_{k=0}^{n} f_k(x) \to f(x), \quad \text{pour } \mu\text{-presque tout } x,$$

et $|\sum_{k=0}^{n} f_k| \leq \sum_{k=0}^{n} |f_k| \leq h$, avec $h \in L^p$. En appliquant le théorème de convergence dominée dans L^p, on obtient donc que

$$\sum_{k=0}^{n} f_k \to f, \quad \text{en norme } \|\cdot\|_p,$$

c'est-à-dire la série de terme général f_n est convergente dans L^p. Le lemme précédent assure donc que L^p est complet.

Le cas $p = \infty$ se traite de manière analogue ou de façon plus directe (exo !).

$\square$

Pour $p = 2$, on a même mieux :

Théorème 12.3. *L'espace $L^2_{\mathbb{C}}(\mu)$ des fonctions à valeurs dans $\mathbb{C}$ de carré intégrable est un espace de Hilbert muni du produit scalaire*

$$\langle f, g \rangle = \int_E f\overline{g} d\mu.$$

Remarque 12.5. L'inégalité de Cauchy-Schwarz dans $L^2_{\mathbb{C}}(\mu)$ n'est rien d'autre que l'inégalité de Hölder avec $p = q = \frac{1}{2}$:

$$|\langle f, g \rangle| \leq \|f\|_2 \|g\|_2 .$$

13. Régularité de la mesure de Lebesgue et résultats de densité

13.1. **Régularité de la mesure de Lebesgue.** Bien qu'on ne sache pas décrire précisément la tribu borélienne, on va voir qu'on peut encadrer avec une précision arbitraire la mesure d'un borélien par la mesure d'un ouvert plus grand et d'un fermé plus petit. C'est la propriété de régularité de la mesure de Lebesgue.

Commençons par quelques rappels sur la distance à un ensemble dans un espace métrique (E, d) (on peut toujours prendre $E = \mathbb{R}$ pour se fixer les idées). La distance à un ensemble A est donnée par

$$d(x, A) = \inf\{d(x, a) \mid a \in A\}.$$

On a alors que la fonction $x \mapsto d(x, A)$ est continue sur E, elle est même 1-lipschitzienne : en effet, pour tout $a \in A$, on a

$$d(x, A) \leq d(x, a) \leq d(x, y) + d(y, a),$$

par l'inégalité triangulaire. Ainsi,

$$d(x, A) - d(x, y) \leq d(y, a),$$

pour tout $a \in A$, et donc, en prenant l'infimum sur $a \in A$, on a $d(x, A) - d(x, y) \leq d(y, A)$. D'où,

$$d(x, A) - d(y, A) \leq d(x, y).$$

Les rôles de x et de y étant symétriques, on obtient l'inégalité en valeur absolue :

$$|d(x, A) - d(y, A)| \leq d(x, y),$$

pour tout $x, y \in E$. Donc $x \mapsto d(x, A)$ est 1-lipschitzienne, et en particulier continue.

On a de plus la propriété suivante : $x \in \overline{A}$ si et seulement si $d(x, A) = 0$. En effet, $x \in \overline{A}$ signifie que pour tout $\varepsilon > 0$, il existe $a \in A$ tel que $d(x, a) < \varepsilon$, i.e. pour tout $\varepsilon > 0$, $d(x, A) < \varepsilon$, i.e. $d(x, A) = 0$.

On commence par montrer la régularité des mesures finies.

Proposition 13.1. *Soit μ une mesure finie sur $(E, \mathcal{B}(E))$, où (E, d) est un espace métrique. Alors μ est régulière, c'est-à-dire, pour tout borélien A, et pour tout $\varepsilon > 0$, il existe un ouvert U et un fermé F, tel que $F \subset A \subset U$ et*

$$\mu(U \setminus F) \leq \varepsilon.$$

Comme μ est finie, on peut aussi écrire que $\mu(U) - \mu(F) \leq \varepsilon$. Donc, comme $\mu(F) \leq \mu(A) \leq \mu(U)$, on a

$$\mu(F) \leq \mu(A) \leq \mu(F) + \varepsilon$$

et

$$\mu(U) - \varepsilon \leq \mu(A) \leq \mu(U).$$

Démonstration. On pose

$$\mathcal{A} = \{A \in \mathcal{B}(E) \mid \forall \varepsilon > 0, \exists \text{ un ouvert } U \text{ et un fermé } F, \text{ t.q. } F \subset A \subset U \text{ et } \mu(U \setminus F) \leq \varepsilon\}.$$

On va montrer que $\mathcal{A}$ est une tribu qui contient les ouverts, ainsi on aura $\mathcal{A} = \mathcal{B}(E)$.

Commençons par montrer que $\mathcal{A}$ contient les ouverts. Soit $\varepsilon > 0$ et A un ouvert. On pose $U = A$. Pour F, on définit

$$F_n = \left\{x \in A \mid d(x, A^c) \geq \frac{1}{n}\right\}.$$

La suite des ensembles F_n est clairement croissante. Comme $x \mapsto d(x, A^c)$ est continue, alors F_n est fermée. De plus, comme A^c est fermé, $d(x, A^c) > 0$ pour tout $x \in A$, et donc il existe n tel que $d(x, A^c) \geq \frac{1}{n}$. Ainsi,

$$A = \bigcup_n F_n,$$

et par "continuité à droite", la mesure μ étant supposée finie,

$$\lim_n \mu(A \setminus F_n) = \mu(\varnothing) = 0.$$

On choisit donc F de la forme F_n pour un n assez grand. Ainsi $A \in \mathcal{A}$, et donc $\mathcal{A}$ contient bien les ouverts.

Montrons alors que $\mathcal{A}$ est une tribu :

(i) elle contient clairement E, il suffit de prendre $U = F = E$;

(ii) elle stable par passage au complémentaire : si $A \in \mathcal{A}$, alors il existe U ouvert, F fermé, tel que $F \subset A \subset U$ tel que $\mu(U \setminus F) \leq \varepsilon$. Il suffit alors de poser $U' = F^c$ qui est ouvert, $F' = U^c$ qui est fermé, et on a $F' \subset A^c \subset U'$ et comme $U' \setminus F' = U \setminus F$, on a bien $\mu(U' \setminus F') \leq \varepsilon$;

(iii) elle est stable par réunion dénombrable : soit $\varepsilon > 0$. Soit pour tout n, $A_n \in \mathcal{A}$, alors il existe U_n ouvert, F_n fermé, tel que $F_n \subset A_n \subset U_n$ et tel que

$$\mu(U_n \setminus F_n) \leq \frac{\varepsilon}{2^{n+1}}.$$

On a alors que

$$\cup_n F_n \subset \cup_n A_n \subset \cup_n U_n$$

et

$$(\cup_n U_n) \setminus (\cup_n F_n) \subset \cup_n U_n \setminus F_n$$

(car si x est dans un des U_n et dans aucun des F_n, il existe n tel que $x \in U_n \setminus F_n$). Alors,

$$\mu\left((\cup_n U_n) \setminus (\cup_n F_n)\right) \leq \mu\left(\cup_n U_n \setminus F_n\right) \leq \sum_n \mu(U_n \setminus F_n) \leq \frac{\varepsilon}{2},$$

où on a utilisé la sous-σ-additivité. On pose $U' = \bigcup_n U_n$ qui est ouvert comme réunion d'ouverts. On ne peut pas faire la même chose pour F, car une union quelconque de fermés n'est pas forcément fermé... On pose alors $F'_n = \bigcup_{k \leq n} F_k$ qui par contre est bien fermé comme union finie de fermés. Comme $\bigcup_n F_n = \bigcup_n \bigcup_{k \leq n} F_k$, la suite étant croissante, et que $\mu(\bigcup_n F_n) < \infty$, il existe n_ε tel que

$$\mu\left(\cup_n F_n\right) \leq \mu\left(\cup_{k \leq n_\varepsilon} F_k\right) + \frac{\varepsilon}{2}.$$

Posons alors $F' = \cup_{k \leq n_\varepsilon} F_k$. Alors, $F' \subset A \subset U'$, et on a

$$\mu(U' \setminus F') = \mu(U') - \mu(F') = \mu(U') - \mu(\cup_n F_n) + \mu(\cup_n F_n) - \mu(F') \leq \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.$$

Ainsi, $\mathcal{A}$ est une tribu qui contient les ouverts, et donc c'est bien la tribu borélienne $\mathcal{B}(E)$. □

Proposition 13.2. *La mesure de Lebesgue est régulière, c'est-à-dire :*

(i) (régularité extérieur) pout tout $A \in \mathcal{B}(\mathbb{R}^d)$,

$$\lambda(A) = \inf\{\lambda(U) \,|\, U \text{ ouvert tel que } A \subset U\};$$

(ii) (régularité intérieur) pout tout $A \in \mathcal{B}(\mathbb{R}^d)$,

$$\lambda(A) = \sup\{\lambda(K) \,|\, K \text{ compact tel que } K \subset A\}.$$

Démonstration. On se contente de le montrer pour la mesure de Lebesgue sur $\mathbb{R}$, le cas $\mathbb{R}^d$ étant identique. Soit A un borélien, et soit $\varepsilon > 0$.

Commençons par la régularité intérieur. Soit A un borélien. Posons $E_n = [-n, n]$, pour tout $n \geq 1$. On considère la mesure trace $\lambda_n(A) = \lambda(A \cap E_n)$. Comme λ_n est finie, d'après la proposition précédente, λ_n est régulière, et donc il existe un fermé F_n tel que $F_n \subset A$, et

$$\lambda(A \cap E_n) \leq \lambda(F_n \cap E_n) + \varepsilon/2.$$

Posons $K_n = F_n \cap E_n$ qui est fermé et borné, donc compact. On a donc, comme $F_n \subset A$,

$$K_n = F_n \cap E_n \subset A \cap E_n \subset \bigcup_n A \cap E_n = A,$$

ainsi $K_n \subset A$. On a de plus que $\lambda(A) = \sup_n \lambda(A \cap E_n)$ par "continuité à gauche" de λ. Si $\lambda(A) < +\infty$, alors il existe n_ε tel que

$$\lambda(A) \leq \lambda(A \cap E_{n_\varepsilon}) + \frac{\varepsilon}{2}.$$

Ainsi, en combinant les deux inégalités ci-dessus, on obtient

$$\lambda(A) \leq \lambda(K_{n_\varepsilon}) + \varepsilon,$$

et donc $\lambda(A) = \sup\{\mu(K) \,|\, K \text{ compact tel que } K \subset A\}$. Si $\lambda(A) = +\infty$, on a toujours $\lambda(A) = \sup_n \lambda(A \cap E_n)$, et comme

$$\begin{aligned} \lambda(A \cap E_n) &= \sup \lambda(K) \,|\, K \text{ compact tel que } K \subset A \cap E_n\} \\ &\leq \sup \lambda(K) \,|\, K \text{ compact tel que } K \subset A\}, \end{aligned}$$

on conclut. Ainsi, λ est intérieurement régulière.

Même principe pour la régularité extérieur, mais en plus astucieux. On peut supposer que $\lambda(A) < \infty$, sinon le résultat est évident car $\lambda(A) \leq \lambda(U)$, pour tout ouvert U contenant A. Soit $n \geq 1$. On considère la mesure trace définie par $\lambda_n(A) = \lambda(A \cap E_n)$, où cette fois $E_n =]-n, n[$, qui est une mesure finie. Par la proposition précédent, λ_n est régulière, il existe donc U_n ouvert tel que $A \subset U_n$, et tel que

$$\lambda(U_n \cap E_n) - \frac{\varepsilon}{2^n} \leq \lambda(A \cap E_n).$$

On pose alors $U = \bigcup_n U_n \cap E_n$ qui est ouvert (en tant que réunion d'ouverts) et vérifie

$$A = \bigcup_n A \cap E_n \subset U$$

Montrons alors que $\lambda(U) - \varepsilon \leq \lambda(A)$. Posons $U_n' = \bigcup_{k=1}^n U_k \cap E_k$, et montrons par récurrence que

$$\lambda(U_n') \leq \lambda(A \cap E_n) + \sum_{k=1}^n \frac{\varepsilon}{2^k}.$$

C'est vrai à $n = 1$, par ce qu'on vient de voir. Supposons que l'inégalité soit vraie à n. Alors,

$$\begin{aligned} \lambda(U_{n+1}') &= \lambda(U_n') + \lambda(U_{n+1} \cap E_{n+1}) - \lambda(U_n' \cap U_{n+1} \cap E_{n+1}) \\ &\leq \lambda(A \cap E_n) + \sum_{k=1}^n \frac{\varepsilon}{2^k} + \lambda(A \cap E_{n+1}) + \frac{\varepsilon}{2^{n+1}} - \lambda(U_n' \cap U_{n+1} \cap E_{n+1}). \end{aligned}$$

Or $A \subset U_{n+1}$, $E_n \subset E_{n+1}$, donc $A \cap E_n \subset U_{n+1} \cap E_{n+1}$, et $A \cap E_n \subset U_n \cap E_n \subset U_n'$, ainsi $A \cap E_n \subset U_n' \cap U_{n+1} \cap E_{n+1}$, et par croissance d'une mesure,

$$\lambda(A \cap E_n) - \lambda(U_n' \cap U_{n+1} \cap E_{n+1}) \leq 0,$$

(les ensembles considérées sont de mesure finie car inclus dans E_n), donc finalement :

$$\lambda(U'_{n+1}) \leq \lambda(A \cap E_{n+1}) + \sum_{k=1}^{n+1} \frac{\varepsilon}{2^k},$$

et l'inégalité est montrée par récurrence. Finalement, comme (U'_n) croit vers U, $A \cap E_n$ croit vers A, on obtient par continuité à gauche,

$$\lambda(U) \leq \lambda(A) + \varepsilon,$$

ce qui montre la régularité extérieure. □

13.2. **Densité des fonctions continues à support compact.**

Proposition 13.3. *Pour tout $p \in [1, +\infty[$, l'ensemble des fonctions étagées intégrables est dense dans $L^p(\mu)$.*

Démonstration. Soit $f \in L^p(\mu)$. On peut supposer f positive (quitte à raisonner sur f^+ et f^-). Par le lemme fondamental d'approximation, il existe une suite $(\varphi_n)_n$ de fonctions étagées positives qui converge en croissant vers f. Il faut donc montrer que $\varphi_n \in L^1(\mu) \cap L^p(\mu)$ pour tout n, et que la convergence a lieu dans L^p. Comme $\varphi_n \leq f$, et que $f \in L^p$, on a bien $\varphi_n \in L^p$. On a donc que $\varphi_n < \infty$ p.p, et donc comme φ_n ne prend qu'un nombre fini de valeurs, on a aussi $\varphi_n \in L^1$. Comme φ_n est positive, on a $f - \varphi_n \leq f \in L^p$, et donc par le théorème de convergence dominée, la suite $(\varphi_n)_n$ converge vers f en norme $|| \cdot ||_p$. □

Remarque 13.1. Le résultat de densité précédent est encore vrai pour $p = \infty$. L'ensemble des fonctions étagées intégrables est en effet dense dans L^∞ par le lemme fondamental d'approximation.

Par contre le résultat suivant est faux pour $p = \infty$.

On rappelle la définition de fonctions à support compact.

Définition 13.1. *Soit $f \colon \mathbb{R}^d \to \mathbb{R}$. On dit que f est à support compact s'il existe un compact $K \subset \mathbb{R}^d$ tel que f soit nulle sur K^c. On définit le support de f comme l'ensemble*

$$\mathrm{supp}(f) = \overline{\{x \in \mathbb{R}^d \mid f(x) \neq 0\}}.$$

On a alors que f est à support compact si et seulement si $\mathrm{supp}(f)$ *est compact (pourquoi ?).*

Théorème 13.1. *Pour tout $p \in [1, +\infty[$, l'espace vectoriel $C_K(\mathbb{R}^d, \mathbb{R})$ des fonctions continues à support compact est dense dans $L^p(\lambda)$.*

Démonstration. On montre le résultat pour $d = 1$ pour alléger les notations, la dimension d quelconque se montre de la même façon. Rappelons que si A est dense dans B et que B est dense dans C, alors A est dense dans C. Il suffit donc de montrer que C_K est dense dans les fonctions étagées intégrables par la proposition précédente. Par linéarité, il suffit alors de montrer que pour tout borélien A de mesure finie, $\mathbb{1}_A$ est limite dans $L^p(\lambda)$ d'une suite d'éléments de C_K.

Soit $\varepsilon > 0$. Par régularité (extérieur) de la mesure de Lebesgue, et comme $\lambda(A) < \infty$, il existe un ouvert O contenant A tel que $\lambda(O \setminus A) < \varepsilon^p$, et donc $||\mathbb{1}_O - \mathbb{1}_A||_p < \varepsilon$ (car $\lambda(O) = \int \mathbb{1}_O d\lambda = \int \mathbb{1}_O^p d\lambda$ et de même pour A). Comme O est la réunion dénombrable d'intervalles ouverts I_n disjoints, en posant $E_N = \bigcup_{n=1}^N I_n$, on a comme $\lambda(O) = \lim_N \lambda(E_N)$ (limite croissante) et donc il existe N tel que $\lambda(O \setminus E_N) < \varepsilon^p$, et donc de même $||\mathbb{1}_O - \mathbb{1}_{E_N}||_p < \varepsilon$.

Or pour tout intervalle $]a, b[$ il est facile de construire une suite de fonction (φ_n) dans C_K qui converge simplement vers $\mathbb{1}_{]a,b[}$. On peut s'en convaincre facilement par un dessin et on peut prendre par exemple :

$$\varphi_n(x) = \begin{cases} 1 & \text{si } x \in]a, b[\\ 0 & \text{si } x \in \mathbb{R} \backslash]a - \frac{1}{n}, b + \frac{1}{n}[\end{cases}$$

et φ_n affine sur $]a - \frac{1}{n}, a[\cup]b, b + \frac{1}{n}[$.

Soit alors $\varphi_n^{(k)}$ dans C_K qui converge simplement vers $\mathbb{1}_{I_k}$. On posant $\psi_n = \sum_{k=1}^N \varphi_n^{(k)}$, on obtient une suite de fonctions continues à support compact qui converge simplement vers $\mathbb{1}_{E_N}$. Par convergence dominée (dans L^p), on obtient $||\mathbb{1}_{E_N} - \psi_n||_p < \varepsilon$, pour n suffisamment grand. Finalement,

$$||\mathbb{1}_A - \psi_n||_p \leq ||\mathbb{1}_A - \mathbb{1}_O||_p + ||\mathbb{1}_O - \mathbb{1}_{E_N}||_p + ||\mathbb{1}_{E_N} - \psi_n||_p < 3\varepsilon.$$

□

Remarque 13.2. L'espace C_K n'est pas dense dans L^∞. En effet, une suite de fonctions continues qui converge uniformément a pour limite une fonction continue...

On va améliorer ce résultat dans la prochaine section en introduisant le produit de convolution.

14. Produit de convolution et densité des fonctions indéfiniment dérivables

On se place dans cette section sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ muni de la mesure de Lebesgue λ. Les résultats se généralisent facilement à $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$, mais on se restreint à $d = 1$ par souci de simplicité.

14.1. **Opérateurs de translation sur les fonctions.**

Définition 14.1. *Soient $a \in \mathbb{R}^d$ et $f \colon (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d)) \to \mathbb{R}$ une application borélienne. On définit la a-translatée de f par a comme la fonction définie par :*

$$\begin{aligned} \tau_a f \colon \mathbb{R}^d &\to \mathbb{R} \\ x &\mapsto \tau_a f(x) = f(x-a). \end{aligned}$$

Proposition 14.1. *Soit $f \in L^p$, avec $p \in [1, +\infty[$. Alors, l'opérateur τ_a est continue dans L^p, c'est-à-dire que* $\lim_{a\to 0} ||\tau_a f - f||_p = 0$.

Démonstration. On commence par supposer que f est dans C_K. Soit $M = \sup_{x\in\mathbb{R}^d} |f(x)|$ et soit R tel que le support de f soit inclus dans la boule $]-R, R[$. Alors si $||a|| \leq 1$, on a

$$|\tau_a f(x) - f(x)| = |f(x-a) - f(x)| \leq M\mathbb{1}_{\{||x-a||\leq R\}} + M\mathbb{1}_{\{||x||\leq R\}} \leq 2M\mathbb{1}_{\{||x||\leq R+1\}} \in L^p.$$

Donc par le théorème de convergence dominée (dans L^p), on a $\lim_{a\to 0} ||\tau_a f - f||_p = 0$.

Soit $\varepsilon > 0$. Si maintenant $f \in L^p$, par densité des fonctions C_K dans L^p, il existe $g \in C_K$ telle que $||f - g||_p < \varepsilon/3$. Par l'argument ci-dessus, il existe $\delta > 0$ tel que si $|a| < \delta$, $||\tau_a g - g||_p < \varepsilon/3$. Alors, on écrit

$$\begin{aligned} ||\tau_a f - f||_p &\leq ||\tau_a f - \tau_a g||_p + ||\tau_a g - g||_p + ||g - f||_p \\ &= ||\tau_a g - g||_p + 2||g - f||_p, \end{aligned}$$

car $||\tau_a f - \tau_a g||_p = ||f - g||_p$ par invariance par translation de la mesure de Lebesgue. Ainsi, $||\tau_a f - f||_p < \varepsilon$, et on obtient le résultat annoncé. □

Remarque 14.1. Le résultat est faux pour $p = \infty$. Par exemple, si $f = \mathbb{1}_{[0,+\infty[}$, on a $||\tau_a f - f||_\infty = 1$ pour tout $a \neq 0$.

La preuve s'appuie fortement sur l'invariance par translation de la mesure de Lebesgue. Si on remplace la mesure de Lebesgue par la mesure δ_0, on $||\tau_a f - f||_p \to 0$ si et seulement si f est continue en 0.

14.2. **Convolution.**

On va définir une nouvelle opération sur les fonctions, et même sur les mesures, appelée *produit de convolution.*

Définition 14.2. *Soit f et g deux fonctions boréliennes sur $\mathbb{R}^d$. On définit la convolée de f et g ou le produit de convolution de f et de g par : pour tout $x \in \mathbb{R}^d$,*

$$f * g(x) = \int_{\mathbb{R}^d} f(x-y)g(y)\lambda(dy).$$

Le produit de convolution est bien définie comme fonction sur $\mathbb{R}^d$ si f et g sont toutes les deux positives, ou au moins pour les $x \in \mathbb{R}^d$ tel que $|f| * |g|(x) < \infty$.

Quand il est défini, on a alors que le produit de convolution est commutatif, i.e. $f * g = g * f$, et de plus, on a que $|f * g|(x) \leq |f| * |g|(x)$ par l'inégalité triangulaire, et que

$$||f * g||_1 \leq \int\int |f(x-y)||g(y)|dydx = ||f||_1||g||_1,$$

et dans ce cas,

$$\int_{\mathbb{R}^d} f * g(x)\lambda(dx) = \left(\int_{\mathbb{R}^d} f(x)\lambda(dx)\right)\left(\int_{\mathbb{R}^d} g(x)\lambda(dx)\right).$$

Il suffit en effet d'appliquer un changement de variables affines, et le théorème de Fubini. On a de plus que

$$\{f * g \neq 0\} \subset \{f \neq 0\} + \{g \neq 0\},$$

où $A + B = \{a + b \,|\, a \in A, b \in B\}$. En effet, si f et g sont positives, et si $x \notin \{f \neq 0\} + \{g \neq 0\}$, alors en écrivant $x = x - y + y$, on a $f(x - y) = 0$ ou $g(y) = 0$, pour tout y, donc $f(x - y)g(y) = 0$, et donc $f * g(x) = 0$, soit $x \notin \{f * g \neq 0\}$. Si f et g ne sont pas positives, il suffit de remarquer que $\{f \neq 0\} = \{|f| \neq 0\}$.

Remarque 14.2. On a donc que si f et g sont dans $L^1(\lambda)$, la convolée $f * g \in L^1(\lambda)$. On a alors que $(L^1(\lambda), +, \cdot, *)$ est une algèbre (sur $\mathbb{R}$ ou $\mathbb{C}$) associative et commutative dont la multiplication interne est donnée par le produit de convolution.

Exemple 14.1. Si $\mathbf{1}$ désigne la fonction constante égale à 1, on a : $\mathbf{1} * g(x) = \int_{\mathbb{R}^d} g(x)dx$.

Calculons $\mathbb{1}_{[0,1]} * \mathbb{1}_{[0,1]}$. Si $x \notin [0, 1] + [0, 1] = [0, 2]$, $\mathbb{1}_{[0,1]} * \mathbb{1}_{[0,1]}(x) = 0$. Si $x \in [0, 2]$, alors

$$\mathbb{1}_{[0,1]} * \mathbb{1}_{[0,1]}(x) = \int \mathbb{1}_{[0,1]}(x-y)\mathbb{1}_{[0,1]}(y)dy = \min(x, 1) - \max(0, x-1) = \begin{cases} x & \text{si } x \in [0, 1], \\ 2 - x & \text{si } x \in [1, 2]. \end{cases}$$

On remarque que le produit de convolution de la fonction discontinue $\mathbb{1}_{[0,1]}$ avec elle même donne une fonction continue ! On a plus généralement :

Proposition 14.2 (Convolution L^p-L^q)**.** *Soit $p, q \in [1, +\infty]$ des exposants conjugués, i.e. $\frac{1}{p} + \frac{1}{q} = 1$. Soient $f \in L^p$, et $g \in L^q$. Alors le produit de convolution $h = f * g$ est défini λ-p.p., et $h \in L^\infty$. De plus, h est uniformément continue.*

Démonstration. L'inégalité de Hölder donne facilement que $h \in L^\infty$. Soit $z \in \mathbb{R}^d$. Alors,

$$\begin{aligned} |h(x) - h(x - z)| &\leq \int_{\mathbb{R}^d} \Big|f(x - y) - f(x - y - z)\Big| g(y)dy \\ &\leq ||f - \tau_z f||_p ||g||_q, \end{aligned}$$

de nouveau par l'inégalité de Hölder. Ainsi,

$$\sup_{x \in \mathbb{R}^d} |h(x) - h(x - z)| \leq ||f - \tau_z f||_p ||g||_q \underset{z \to 0}{\longrightarrow} 0,$$

par continuité de l'opérateur de translation dans L^p. Ainsi, h est uniformément continue. □

Le produit de convolution permet ainsi de régulariser les fonctions, et va nous permettre d'obtenir des résultats de densité par des fonctions très régulières en introduisant la notion de suite régularisante.

Remarque 14.3. L'algèbre L^1 ne possède pas d'unité pour le produit de convolution. Supposons par l'absurde que ça soit le cas, et notons $u \in L^1$ la fonction vérifiant $f * u = f$ λ-p.p. pour tout $f \in L^1$. Prenons $f(x) = e^{-\rho x^2}$, avec $\rho > 0$. Alors comme $f \in L^\infty$, $f * u$ est continue. Comme f est continue aussi, $f * u(x)$ et $f(x)$ coïncident partout [1]. En $x = 0$, on a

$$f * u(0) = \int e^{-\rho y^2} u(y)dy \to 0 \quad \text{quand } \rho \to \infty,$$

par convergence dominée. Or $f(0) = 1$, ce qui donne une contradiction.

Pour palier le manque d'élément neutre pour le produit de convolution dans L^1, on introduit la définition suivante :

Définition 14.3. *Une suite $(\alpha_n)_n$ de L^1 est appelée une approximation de l'identité si*

1. En effet, il suffit de vérifier qu'une fonction continue nulle λ-p.p. est nulle partout (en l'appliquant à $f - g$). Soit alors f continue tel que $\lambda(\{f \neq 0\}) = 0$. Or $\{f \neq 0\}$ est un ouvert, et un ouvert non vide est de mesure de Lebesgue non nulle (pourquoi ?), donc $\{f \neq 0\}$ est vide.

(i) pour tout n, $\int \alpha_n d\lambda = 1$;

(ii) $\sup_n \int |\alpha_n| d\lambda < \infty$ *;*

(iii) pour tout $a > 0$, $\lim_{n\to\infty} \int_{\{|x|>a\}} |\alpha_n(x)|dx = 0$.

La construction d'une telle suite se fait usuellement en prenant un élément $\alpha \in L^1$ d'intégrale $\int \alpha d\lambda = 1$, et en posant :

$$\alpha_n(x) = n\alpha(nx).$$

On vérifie facilement par convergence dominée que $(\alpha_n)_n$ est alors une approximation de l'identité. Les "noyaux" usuels sont par exemple :

— noyau de Gauss : $\alpha(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$;

— noyau de Laplace : $\alpha(x) = \frac{1}{2} e^{-|x|}$;

— noyau de Cauchy : $\alpha(x) = \frac{1}{\pi}\frac{1}{1+x^2}$.

Lemme 14.1. Soit $f \in L^p$, avec $p \in [1, +\infty]$, et $g \in L^1$. Alors $f * g$ est bien définie λ-p.p., et

$$||f * g||_p \leq ||f||_p ||g||_1.$$

Démonstration. Si $p = 1$, le résultat est évident car

$$|||f| * |g|||_1 = \int |f| * |g|(x)dx = \int\int |f(x-y)||g(y)|dydx = ||f||_1||g||_1.$$

Si $p = \infty$, c'est évident aussi. Si $1 < p < \infty$, on a, en appliquant l'inégalité de Hölder à la mesure $|g(y)|dy$,

$$|f| * |g|(x) = \int |f(x-y)| \times 1\, |g(y)|dy \leq \left(\int |f(x-y)|^p |g(y)|dy\right)^{1/p} \left(\int |g(y)|dy\right)^{1/q},$$

avec $\frac{1}{p} + \frac{1}{q} = 1$. En élevant à la puissance p, on a

$$\begin{aligned} |||f| * |g|||_p^p = \int |f * g(x)|^p dx &\leq \int\int |f(x-y)|^p |g(y)|dydx \left(\int |g(y)|dy\right)^{p-1} \\ &= ||g||_1^p ||f||_p^p, \end{aligned}$$

ce qui donne le résultat. □

Proposition 14.3. *Soit $f \in L^p$, avec $p \in [1, +\infty[$ et soit (φ_n) une approximation de l'identité. Alors, $f * \varphi_n \in L^p$, et on a*

$$||f - f * \varphi_n||_p \underset{n\to 0}{\longrightarrow} 0.$$

Démonstration. Comme $f \in L^p$ et $\alpha_n \in L^1$, on a par Hölder $f * \alpha_n \in L^p$ par le lemme précédent. On écrit alors, pour λ-presque tout x,

$$f * \varphi_n(x) - f(x) = \int \Big[f(x-y) - f(x)\Big]\varphi_n(y)dy,$$

et donc

$$\begin{aligned} |f * \varphi_n(x) - f(x)| &\leq \int \Big|f(x-y) - f(x)\Big| |\varphi_n(y)|dy \\ &\leq \left(\int |f(x-y) - f(x)|^p |\varphi_n(y)|dy\right)^{1/p} \left(\int |\varphi_n(y)|dy\right)^{1-1/p}, \end{aligned}$$

en utilisant l'inégalité de Hölder appliquée à la mesure à densité $|\varphi_n(y)|dy$. Or par définition d'une approximation de l'identité, $\sup_n \int |\varphi_n| d\lambda = M < \infty$, on en déduit, en utilisant le théorème de Fubini,

$$\begin{aligned}\int |f * \varphi_n(x) - f(x)|^p dx &\leq M^{p-1} \int \left(\int |f(x-y) - f(y)|^p |\varphi_n(y)| dy \right) dx \\ &= M^{p-1} \int \left(\int |f(x-y) - f(y)|^p dx \right) |\varphi_n(y)| dy \\ &= M^{p-1} \int ||\tau_y f - f||_p^p |\varphi_n(y)| dy\end{aligned}$$

On découpe alors l'intégrale

$$\int ||\tau_y f - f||_p^p |\varphi_n(y)| dy = \int_{\{|x| \leq a\}} ||\tau_y f - f||_p^p |\varphi_n(y)| dy + \int_{\{|x| > a\}} ||\tau_y f - f||_p^p |\varphi_n(y)| dy$$

Le deuxième terme est borné par (car $||\tau_y f||_p = ||f||_p$ par invariance par translation de la mesure de Lebesgue) :

$$\begin{aligned}\int_{\{|y|>a\}} ||\tau_y f - f||_p^p |\varphi_n(y)| dy &\leq \int_{\{|y|>a\}} (||\tau_y f||_p + ||f||_p)^p |\varphi_n(y)| dy \\ &\leq 2^p ||f||_p^p \int_{\{|y|>a\}} |\varphi_n(y)| dy\end{aligned}$$

qui tend vers 0 quand $n \to \infty$ par définition d'une approximation de l'identité. Le premier terme est lui bornée par

$$\begin{aligned}\int_{\{|y| \leq a\}} ||\tau_y f - f||_p^p |\varphi_n(y)| dy &\leq \sup_{|y| \leq a} ||\tau_y f - f||_p^p \int |\varphi_n(y)| dy \\ &= M \sup_{|y| \leq a} ||\tau_y f - f||_p^p\end{aligned}$$

qui tend vers 0 quand $a \to 0$ par continuité dans L^p de l'opérateur de translation. Ceci achève la démonstration. □

Un cas particulièrement intéressant d'approximation de l'identité est le cas où la suite est C^∞. On parle alors de :

Définition 14.4. *Une suite (φ_n) de fonctions positives et C^∞ à support compact est appelée une suite régularisante si pour tout n, $\int_{\mathbb{R}} \varphi_n d\lambda = 1$, et si le support de φ_n est inclus dans l'intervalle $]-\frac{1}{n}, \frac{1}{n}[$.*

On peut toujours prendre par exemple :

$$\psi(x) = \begin{cases} \frac{1}{c} \exp\left(-\frac{1}{1-|x|^2}\right) & \text{si } |x| < 1, \\ 0 & \text{si } |x| \geq 1, \end{cases}$$

où $c = \int_{[-1,1]} \exp\left(-\frac{1}{1-|x|^2}\right) d\lambda$, et poser $\varphi_n(x) = n\psi(nx)$.

En effet, on a la propriété suivante :

Proposition 14.4. *Soit f une fonction localement intégrable, c'est-à-dire intégrable sur tout compact de $\mathbb{R}$, et φ une fonction de classe C_K^n, pour $n \in \mathbb{N}^* \cup \{+\infty\}$ (c'est-à-dire une fonction à support compact n-fois dérivable de dérivée n^e continue). Alors $f * \varphi$ est de classe C_K^n, et on a*

$$(f * \varphi)' = f * \varphi'.$$

Démonstration. C'est une simple application du théorème de dérivation sous le signe intégrale, les détails sont laissées en exercice. □

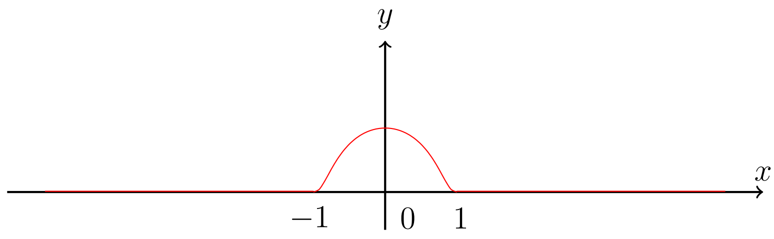

Figure 4. La fonction ψ.

On a alors le puissant théorème suivant de densité par des fonctions très régulières :

Théorème 14.1. *Soit $p \in [1, +\infty[$. L'espace C_K^∞ des fonctions indéfiniment dérivables à support compact est dense dans L^p.*

Démonstration. Soit $f \in L^p$ et soit $\varepsilon > 0$. Par densité des fonctions continues à support compact dans L^p, il existe $g \in C_K$ tel que $||f - g||_p < \varepsilon/2$. Soit (φ_n) une suite régularisante. Par les résultats précédents, $g * \varphi_n$ est dans C_K^∞ (proposition 14.4) et $||g - g * \varphi_n||_p \to 0$ quand $n \to \infty$ (proposition 14.3). Donc en choisissant n suffisamment grand de sorte que $||g - g * \varphi_n||_p < \varepsilon/2$, on obtient

$$||f - g * \varphi_n||_p \leq ||f - g||_p + ||g - g * \varphi_n||_p < \varepsilon,$$

ce qui montre la densité de C_K^∞ dans L^p. $\square$

Remarque 14.4. On peut plus généralement définir le produit de convolution entre mesures σ-finies :

Soient μ et ν deux mesures σ-finies sur $\mathcal{B}(\mathbb{R}^d)$. On appelle produit de convolution, et on note $\mu * \nu$, l'image de $\mu \otimes \nu$ par l'application somme $(x, y) \mapsto x + y$.

Pour tout borélien A, on a alors

$$\mu * \nu(A) = \int \mathbb{1}_A(x + y) d\mu \otimes \nu(x, y),$$

qui s'écrit aussi, par le théorème de Fubini,

$$\mu * \nu(A) = \int d\mu(x)\nu(A - x) = \int d\nu(x)\mu(A - x).$$

Le produit de convolution est clairement commutatif (et aussi associatif par associativité du produit de mesures, mais il faut d'abord vérifier que $\mu * \nu$ est σ-finie, ce qui n'est pas toujours le cas, comme le montre le contre-exemple $\lambda * \lambda$), et de plus on a $\mu * \nu(\mathbb{R}^d) = \mu(\mathbb{R}^d)\nu(\mathbb{R}^d)$. Son élément neutre est la mesure δ_0.

D'un point de vue probabiliste, si X est une variable aléatoire de loi μ et Y une variable aléatoire de loi ν, avec X et Y indépendantes, le produit de convolution $\mu * \nu$ n'est rien d'autre que la loi de la somme $X + Y$.

15. Transformée de Fourier

Définition 15.1. *Soit $f\colon \mathbb{R}^d \to \mathbb{C}$ une fonction intégrable par rapport à la mesure de Lebesgue sur $\mathbb{R}^d$. On définit la transformée de Fourier de f, que l'on note $\hat{f}$, la fonction $\hat{f}\colon \mathbb{R}^d \to \mathbb{C}$ définie par*

$$\hat{f}(u) = \int_{\mathbb{R}^d} e^{-i\langle u,x\rangle} f(x) d\lambda(x),$$

pour tout $u \in \mathbb{R}^d$.

Proposition 15.1 (Lemme de Riemann-Lebesgue)**.** *Soit $f \in L^1(\lambda)$. Alors $\hat{f}$ est continue, $||\hat{f}||_\infty \leq ||f||_1$ et $\hat{f}(u) \underset{|u|\to\infty}{\longrightarrow} 0$.*

Démonstration. La continuité de $\hat{f}$ découle immédiatement du théorème de continuité sous le signe intégrale. En effet, $x \mapsto f(x)e^{-itx}$ est borélienne et dominée par $|f| \in L^1$. De plus, la fonction $t \mapsto f(x)e^{-itx}$ est continue pour tout $x \in \mathbb{R}$. On a de plus,

$$|\hat{f}(t)| \leq \int_{\mathbb{R}} |f(x)| d\lambda(x) = ||f||_1,$$

pour tout $t \in \mathbb{R}$, donc $||\hat{f}||_\infty \leq ||f||_1$.

Montrons que $\hat{f}(u) \underset{|u|\to\infty}{\longrightarrow} 0$. Soit $t \neq 0$. Alors, comme $e^{i\pi} = -1$, on a

$$\hat{f}(t) = \int_{\mathbb{R}} f(x)e^{-itx}dx = -\int_{\mathbb{R}} f(x)e^{-it(x-\frac{\pi}{t})}dx = -\int_{\mathbb{R}} f\Big(y + \frac{\pi}{t}\Big)e^{-ity}dy,$$

par un changement de variable. Alors,

$$2\hat{f}(t) = \int_{\mathbb{R}} \Big(f(x) - f(x+\frac{\pi}{t})\Big) e^{-itx}dx,$$

d'où

$$|\hat{f}(t)| \leq \frac{1}{2}\int_{\mathbb{R}} |f(x) - f(x+\frac{\pi}{t})|dx = \frac{1}{2}||f - \tau_{\frac{\pi}{t}} f||_1,$$

où $\tau_{\frac{\pi}{t}}$ est l'opérateur de translation. Par continuité de $\tau_{\frac{\pi}{t}}$ dans L^1, on en déduit que $\hat{f}(u) \underset{|u|\to\infty}{\longrightarrow} 0$. □

Remarque 15.1. Le résultat est encore vrai dans $\mathbb{R}^d$ et se montre de la même manière.

La transformée de Fourier étant clairement linéaire, on en déduit alors immédiatement que :

Proposition 15.2. *L'application $\mathcal{F}\colon f \mapsto \hat{f}$ de L^1 dans C_0 est un opérateur continue.*

Définition 15.2. *Soit μ une mesure finie sur $\mathbb{R}^d$. On définit la transformée de Fourier de μ, que l'on note $\hat{\mu}$, la fonction $\hat{\mu}\colon \mathbb{R}^d \to \mathbb{C}$ définie par*

$$\hat{\mu}(u) = \int_{\mathbb{R}^d} e^{-i\langle u,x\rangle} d\mu(x),$$

pour tout $u \in \mathbb{R}^d$.

Si μ a une densité f par rapport à la mesure de Lebesgue, on a bien sûr $\hat{\mu} = \hat{f}$.

Proposition 15.3. *Soient f et g dans L^1. Alors,*

$$\widehat{f * g} = \hat{f}\hat{g}$$

Démonstration. On applique Fubini et un changement de variables

$$\begin{aligned}\widehat{f*g}(u) &= \int e^{-iux} f*g(x)dx = \int\int e^{-iux} f(x-y)g(y)dxdy \\ &= \int\int e^{-iuz}e^{-iuy} f(z)g(y)dzdy \\ &= \hat{f}(u)\hat{g}(u)\end{aligned}$$

□

Les propriétés suivantes sont faciles et laissées en exercice :

Proposition 15.4. *Lorsque tout est défini, on a :*

(i) *si* $g(x) = f(-x)$*, alors* $\hat{g}(u) = \hat{f}(-u)$ *;*

(ii) *si* $g(x) = \overline{f}(x)$*, alors* $\hat{g}(u) = \overline{\hat{f}(-u)}$ *;*

(iii) *si* $g(x) = f(x/a)$*,* $\hat{g}(u) = a^d \hat{f}(au)$ *;*

(iv) $\widehat{f*g}(t) = \hat{f}(t)\hat{g}(t)$ *;*

(v) $\widehat{f'}(t) = it\hat{f}(t)$ *;*

(vi) $\widehat{xf(x)}(t) = i(\hat{f})'(t)$ *;*

Annexe A. Un ensemble non borélien

On a dit qu'il existait des ensembles non boréliens, mais sans donner d'exemple... C'est bien le cas, et on va "exhiber" un ensemble qui n'admet pas de mesure de Lebesgue (et est donc non borélien). Pour cela, on commence par faire une petite digression sur l'axiome du choix...

A.1. **L'axiome du choix.** L'axiome du choix, en abrégé AC, est un axiome de la théorie des ensembles, formulé en 1904 par Zermelo, et qui affirme qu'il est possible de construire des ensembles en répétant un nombre infini de fois l'opération de choisir un élément dans un ensemble non vide, même si cette action n'est pas spécifiée. De façon plus formelle, l'axiome du choix énonce que pour toute famille $\mathcal{A}$ d'ensembles non-vides, il existe une fonction, appelée fonction de choix, qui à chaque ensemble A de $\mathcal{A}$ associe un élément appartenant à A :

$$\forall \mathcal{A}, \left[\varnothing \notin \mathcal{A} \Rightarrow \exists f \colon \mathcal{A} \to \bigcup_{A \in \mathcal{A}} A, \ \forall A \in \mathcal{A}, f(A) \in A \right].$$

Une formulation équivalente est alors que pour toute famille $(X_i)_{i\in I}$ d'ensembles non-vides, le produit cartésien $\prod_{i\in I} X_i$ est non-vide.

Cette formulation peut paraître intuitive (si I est fini, on a pas besoin de l'axiome du choix), mais il est possible de montrer que l'axiome du choix est équivalent à certains énoncés beaucoup moins intuitifs, tel que le lemme de Zorn (tout ensemble inductif admet un élément maximal), ou le théorème de Zermelo (tout ensemble peut être bien ordonné)... Il mêne aussi à un paradoxe bien connu, le paradoxe de Banach-Tarski, qui énonce qu'il est possible de découper une boule de $\mathbb{R}^3$ en un nombre fini de morceaux et de réassembler ces morceaux pour former deux boules identiques à la première. Mais les morceaux en question ne sont pas des ensembles mesurables...

Il a été démontré (Gödel 1938, et Cohen 1963) que l'axiome du choix est indépendant des axiomes usuels de la théorie des ensembles ZF (théorie de Zermelo-Fraenkel), on ne peut donc n'y le montrer, n'y le réfuter dans ZF.

Ce que permet l'axiome du choix, c'est alors de supposer l'existence de certains objets mais sans les construire explicitement, ce qui peut paraître insatisfaisant.

A.2. **L'ensemble de Vitali.** On construit ici, via l'axiome du choix, une partie de $\mathbb{R}$ qui n'admet pas de mesure de Lebesgue, et donc qui n'est pas borélienne. Cet exemple a été découvert par Giuseppe Vitali en 1905.

On définit la relation d'équivalence $\sim$ sur $\mathbb{R}$ par :

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

Elle est bien réflexive ($x \sim x$), symétrique ($x \sim y \Rightarrow y \sim x$) et transitive ($x \sim y$ et $y \sim z \Rightarrow x \sim z$). Par exemple, $\sqrt{2}$ et $\sqrt{2}+3$ sont dans la même classe d'équivalence, mais π et $\sqrt{2}$ non. Chaque classe d'équivalence consiste en une copie translatée de $\mathbb{Q}$, et l'union de ces classes forme une partition non-dénombrable de $\mathbb{R}$.

On "choisit" alors, via l'axiome du choix, un et un seul représentant de chaque classe d'équivalence de $\sim$ dans $[0,1]$. Notons V cette partie. Elle est non dénombrable, et nous allons montrer qu'elle n'admet pas de mesure de Lebesgue.

Supposons par l'absurde que V possède une mesure de Lebesgue. Définissons :

$$A = \bigcup_{q \in \mathbb{Q} \cap [-1,1]} (q + V),$$

où $q+V = \{q+x \mid x \in V\}$. Comme V admet une mesure de Lebesgue, chaque partie $q+V$ aussi, et l'invariance par translation de la mesure de Lebesgue donne $\lambda(q+V) = \lambda(V)$.

La partie A étant une réunion dénombrable des ensembles $q+V$, elle admet elle aussi une mesure de Lebesgue. Les parties $q+V$ sont de plus disjointes, puisque V ne contient qu'un seul représentant de chaque classe d'équivalence : soit $x \in (q_1+V) \cap (q_2+V)$ pour $q_1 \neq q_2$ deux rationnels. Alors il existe v_1 et v_2 dans V tel que $x = q_1 + v_1 = q_2 + v_2$. Donc $v_1 - v_2 = q_2 - q_1 \in \mathbb{Q}$ et donc $v_1 \sim v_2$. Ainsi, $v_1 = v_2$, et $q_1 = q_2$, une contradiction.

Par σ-additivité, $\lambda(A) = \sum_{q \in \mathbb{Q} \cap [-1,1]} \lambda(V)$, et donc $\lambda(A)$ est infinie ou nulle. Mais A est inclus dans $[-1,2]$, donc sa mesure de Lebesgue est bornée par 3, donc c'est forcément 0.

Observons maintenant que A contient $[0,1]$. Soit $x \in [0,1]$, et notons $a(x)$ le représentant dans $[0,1]$ de x pour la relation $\sim$. On a donc $x - a(x) = q$ pour un certain $q \in \mathbb{Q}$. Mais x et $a(x)$ sont tous deux dans $[0,1]$, donc $x - a(x) \in [-1,1]$ et donc $q \in \mathbb{Q} \cap [-1,1]$, et ainsi $x \in (q+V)$. Ainsi A contient $[0,1]$ et sa mesure de Lebesgue est donc minorée par 1.

On obtient donc une contradiction, et V ne peut admettre de mesure de Lebesgue.

Références

Ces notes de cours sont essentiellement basées sur :

[1] Briane, Marc et Pagès, Gilles, *Théorie de l'intégration : licence de mathématiques ; cours et exercices.*, Vuibert, 2000.