Intégration pour les probabilités

François Le Maître

La théorie des probabilités vue en terminale S nous apprend comment modéliser des phénomènes aléatoires discrets, par exemple un nombre entier aléatoire entre 1 et 10. Modéliser un phénomène aléatoire discret, c'est se donner un ensemble dénombrable Ω appelé **univers**, et une fonction $\mathbf{P} : \Omega \to [0,1]$ telle que :

$$\sum_{\omega\in\Omega} \mathbf{P}(\omega) = 1.$$

Pour modéliser un nombre entier aléatoire uniforme X entre 1 et 10, on prend par exemple $\Omega = \{1, ..., 10\}$, et on pose, pour $n \in \{1, ..., 10\}$, $\mathbf{P}(n) = \frac{1}{10}$ la densité de probabilité en n. La fonction $\mathbf{P}(n)$ correspond à la probabilité d'obtenir $X = n$ comme résultat. Si l'on veut calculer la probabilité d'un évènement plus compliqué, comme par exemple l'évènement "X est pair", il faut considérer l'ensemble $A = \{n \in \Omega : n \text{ est pair}\}$ et calculer sa probabilité $\mathbb{P}(A)$ en écrivant

$$\mathbb{P}(A) = \sum_{n\in A} \mathbf{P}(n). \tag{1}$$

On obtient ainsi une fonction $\mathbb{P} : \mathcal{A} \to [0,1]$ où $\mathcal{A} = \mathcal{P}(\Omega)$ est l'ensemble des parties de Ω. On peut oublier la fonction $\mathbf{P}$, pour raisonner directement sur $\mathbb{P}$ définie par la formule (1).

Exercice 1. Vérifier que $\mathbb{P}$ satisfait les axiomes suivants.

(i) $\mathbb{P}(\emptyset) = 0$,

(ii) $\mathbb{P}(\Omega) = 1$,

(iii) pour tout A évènement, $\mathbb{P}(\Omega \setminus A) = 1 - \mathbb{P}(A)$,

(iv) pour toute famille dénombrable $(A_n)_{n\in\mathbb{N}}$ d'évènements disjoints, $\mathbb{P}(\bigsqcup_{n\in\mathbb{N}} A_n) = \sum_{n\in\mathbb{N}} \mathbb{P}(A_n)$

Montrer qu'à partir d'une fonction satisfaisant ces hypothèses, on peut retrouver une densité de probabilité $\mathbb{P}$.

Les choses se compliquent dès lors que l'on veut modéliser un réel aléatoire, mettons entre 0 et 1. On a un analogue de notre densité de probabilité $\mathbf{P}$, en partant d'une fonction $\mathbf{p} : [0,1] \to [0,1]$ intégrable, d'intégrale 1. Pour un réel entre 0 et 1 "uniforme", $\mathbf{p}(x) = x$, et on peut par exemple calculer la probabilité que notre variable aléatoire atterrisse dans l'intervalle $[a,b] \subseteq [0,1]$, par la formule :

$$\mathbb{P}([a,b]) = \int_a^b \mathbf{p}(x)dx.$$

On rencontre maintenant la question de l'ensemble de définition $\mathcal{A}$ de $\mathbb{P}$. Tout d'abord, si on connaît $\mathbb{P}(A)$, on doit connaître $\mathbb{P}([0,1] \setminus A) = 1 - \mathbb{P}(A)$. De même, il semble naturel de pouvoir calculer $\mathbb{P}(A \cup B)$ ou $\mathbb{P}(A \cap B)$. Ensuite, on voudrait pouvoir traiter d'évènements du type "avoir une décomposition ternaire qui commence par un nombre fini de 0 puis un 1. Cet évènement A s'écrit $A = \bigsqcup_{n\in\mathbb{N}^*} A_n$ où $A_n = [\frac{1}{3^n}, \frac{2}{3^n}[$, et on veut écrire

$$\mathbb{P}(A) = \sum_{n\in\mathbb{N}^*} \mathbb{P}(A_n) = \frac{1}{2}.$$

Mais se pose le problème qu'une telle écriture n'est pas nécessairement unique (par exemple rien n'interdit de couper en deux $A_1 = [1/3, 1/2[\sqcup [1/2, 2/3[$), et il n'est donc pas sûr que $\mathbb{P}(A)$ soit bien défini. En résumé, l'ensemble $\mathcal{A}$ des parties de $[0,1]$ sur lesquelles $\mathbb{P}$ est définie doit satisfaire les hypothèses suivantes :

(a) $\mathcal{A}$ est stable par complémentaire,

(b) $\mathcal{A}$ est stable par union dénombrable disjointe,

(c) $\mathcal{A}$ est stable par intersection finie,

(d) $\mathcal{A}$ contient $\emptyset, [0,1]$.

Exercice 2. Montrer que $\mathcal{A} \subseteq \mathcal{P}([0,1])$ satisfaisant les hypothèses précédentes sera également stable par intersection dénombrable. Montrer que les hypothèses précédentes impliques que $\mathcal{A}$ est stable par union dénombrable quelconque. Montrer que les axiomes précédents équivalent aux suivants :

(a') $\mathcal{A}$ est stable par complémentaire

(b') $\mathcal{A}$ est stable par union dénombrable

(c') $\mathcal{A}$ contient $\emptyset$.

Pour définir un début de théorie des probabilités, il faut donc à la fois trouver un ensemble de parties pas trop grand sur lesquelles on peut espérer définir $\mathbb{P}$ (notion de **σ-algèbre**), puis définir $\mathbb{P}$ (on utilisera notamment le **théorème de Carathéodory**).

1 σ-algèbres et fonctions mesurables

Définition 1.1. Soit Ω un ensemble. Une **σ-algèbre**, ou **tribu** sur Ω est un ensemble $\mathcal{A} \subseteq \mathcal{P}(\Omega)$ satisfaisant les hypothèses suivantes :

(a) $\mathcal{A}$ est stable par complémentaire : pour tout $A \in \mathcal{A}$, $\Omega \setminus A \in \mathcal{A}$,

(b) $\mathcal{A}$ est stable par union dénombrable : pour toute suite $(A_n)_{n\in\mathbb{N}} \in \mathcal{A}^{\mathbb{N}}$, $\bigcup_{n\in\mathbb{N}} A_n \in \mathcal{A}$,

(c) $\mathcal{A}$ contient $\emptyset$.

Le couple $(\Omega, \mathcal{A})$ sera appelé **espace mesurable**.

Exercice 3. Montrer que ces axiomes impliquent que $\mathcal{A}$ est stable par intersection dénombrable.
Soit $\mathbb{B} \subseteq \mathcal{P}(\Omega)$. Montrer qu'il existe une plus petite (pour l'inclusion) σ-algèbre contenant $\mathbb{B}$. On l'appelle la **σ-algèbre engendrée** par $\mathbb{B}$, notée $\sigma(\mathbb{B})$.

Remarque. D'un point de vue plus informaticien, on peut voir $\mathbb{B}$ comme l'ensemble des propriétés "définissables" de base (par exemple les intervalles dans $\mathbb{R}$, ou les singletons dans $\mathbb{N}$). $\sigma(\mathbb{B})$ est ensuite défini par induction comme l'ensemble des propriétés définissables du type $\exists n : P_n(x)$ (union dénombrable), ou du type $\neg P(x)$ (complémentaire), où P, P_n sont des propriétés définissables. La stabilité par intersection dénombrable s'interprète également comme le fait que si (P_n) sont définissables, alors $\forall n P_n(x)$ l'est aussi.

Donnons tout de suite un exemple fondamental : la tribu des **boréliens** de $\mathbb{R}$. On la définit comme $\mathcal{B}(\mathbb{R}) = \sigma\{[a,b] : a < b\}$.

Exercice 4. Montrer que pour tout $a \in \mathbb{R}$, $\{a\} \in \mathcal{B}(\mathbb{R})$ [1]. Montrer que $\mathcal{B}(\mathbb{R}) = \sigma(\{[a,b[: a < b\}) = \sigma(\{]a,b[: a < b\}$.

La notion de la tribu des boréliens fait sens dès qu'on a un espace topologique (X,τ) : c'est la σ-algèbre engendrée par les ouverts. Si vous ne savez pas ce qu'est un espace topologique, retenez que dans le cas des espaces métriques séparables, la tribu des boréliens est la σ-algèbre engendrée par les boules ouvertes. Montrons qu'il y a bien équivalence entre ces définitions :

Proposition 1.2. *Soit (X,d) un espace métrique séparable, alors $\sigma(\{ouverts\}) = \sigma(\{\text{ boules ouvertes}\})$, et donc on peut définir au choix les boréliens comme la tribu engendrée par les ouverts, ou comme la tribu engendrée par les boules ouvertes.*

Démonstration. Tout d'abord, comme toute boule ouverte est ouverte, $\mathcal{A} \subseteq \mathcal{B}$. Pour conclure, il suffit de montrer que les ouverts sont dans $\mathcal{A}$. Pour cela, on les écrit comme réunion dénombrable de boules ouvertes, en utilisant une suite $(x_n)_{n\in\mathbb{N}}$ dense (hypothèse de séparabilité). Soit donc U un ouvert de X, soit $N_U = \{n \in \mathbb{N} : x_n \in U\}$. Alors pour tout $n \in N_U$, comme U est ouvert et $x_n \in U$, on a $r_n = d(x_n, X \setminus U) > 0$. Alors $U = \bigcup_{n\in N_U} B(x_n, r_n/2)$. En effet, l'inclusion de droite à gauche est immédiate, et réciproquement si $x \in U$, soit $r = d(x, X \setminus U > 0$. Par densité de (x_n), il existe $n \in \mathbb{N}$ tel que $d(x,x_n) < r/2$. Mais alors par inégalité triangulaire $d(x_n, X \setminus U) \geqslant r/2 + r/2 = r$, donc $r_n \geqslant r$, et donc $x \in B(x_n, r_n/2)$. $\square$

On note $\mathcal{B}(X)$ la tribu des boréliens de (X,d) métrique séparable.

Exercice 5. Montrer que les ensembles suivants sont dans $\mathcal{B}(\mathbb{R})$:

1. $\{x \in \mathbb{R} : \text{ la partie entière de } x \text{ est paire}\}$,
2. $\{x \in \mathbb{R} : \sin(x) < 1/3\}$
3. $\{x \in \mathbb{R} : \text{ l'écriture ternaire de } x \text{ contient un nombre fini de } 2\}$

Comme on peut le voir, la démonstration du fait qu'un ensemble est borélien peut être vite pénible, d'où l'intérêt de la notion de fonction mesurable que voici.

Définition 1.3. Soient $(X,\mathcal{A})$ et $(Y,\mathcal{B})$ [2] deux espaces mesurables. Une application $f : X \to Y$ est dite **mesurable** si pour tout $B \in \mathcal{B}$, $f^{-1}(B) \in \mathcal{A}$. Si X et Y sont munis de tribus boréliennes, on dit aussi que f est **borélienne**

1. Indication : on peut définir un réel à partir des rationnels.
2. On passe à la notation standard en théorie de la mesure : Ω reviendra lors du cours de probabilités.

Proposition 1.4. *Soient (X, d_X) et (Y, d_Y) deux espaces métriques séparables. Soit $f : X \to Y$ telle que pour toute boule ouverte B de Y, $f^{-1}(B) \in \mathcal{B}(X)$. Alors f est borélienne.*

La preuve est laissée au lecteur, en suivant cet exercice qui illustre bien les preuves utilisant les σ algèbres (vous pourrez constater que l'on frôle vite l'abstract nonsense !).

Exercice 6. Montrer tout d'abord que si $f : X \to Y$ est une application quelconque, alors pour tout suite $(A_n)_{n\in\mathbb{N}}$ de sous-ensembles de Y, on a :

$$f^{-1}\left(\bigcup_{n\in\mathbb{N}} A_n\right) = \bigcup_{n\in\mathbb{N}} f^{-1}(A_n).$$

Montrer également que pour tout $A \subseteq Y$, $f^{-1}(Y \setminus A) = X \setminus f^{-1}(A)$. En déduire que, si $\mathcal{B}$ est une σ-algèbre sur Y, $f^{-1}(\mathcal{B}) := \{f^{-1}(B) : B \in \mathcal{B}\}$ est une σ-algèbre sur X [3]. Montrer que si $\mathbb{B} \subseteq \mathcal{P}(Y)$, $f^{-1}(\sigma(\mathbb{B})) = \sigma(f^{-1}(\mathbb{B}))$:

1. Pour $\supseteq$, utiliser la question précédente
2. Pour $\subseteq$, considérer $\mathcal{A} = \{B \in \sigma(\mathbb{B}) : f^{-1}(B) \in \sigma(f^{-1}(B))\}$ et montrer que $\mathcal{A} = \sigma(\mathbb{B})$.

Conclure quant à la preuve de la proposition précédente.

Corollaire 1.5. *Une application continue est borélienne.*

1.1 Espace produit

Soient $(X, \mathcal{A})$ et $(Y, \mathcal{B})$ deux espaces mesurables. On définit sur $X \times Y$ la **tribu produit** comme la tribu engendrée par les $A \times B$, où $A \in \mathcal{A}$ et $B \in \mathcal{B}$. On la note $\mathcal{A} \otimes \mathcal{B}$.

Exercice 7.
1. Vérifier que π_X, π_Y sont mesurables.
2. Montrer que $\mathcal{A} \otimes \mathcal{B}$ est la plus petite tribu telle que π_X, π_Y soient mesurables.
3. Montrer que $f : (Z, \mathcal{C}) \to (X \times Y, \mathcal{A} \otimes \mathcal{B})$ est mesurable ssi $\pi_X \circ f$ et $\pi_Y \circ f$ sont mesurables.

Exercice 8. Soient (X, d_X) et (Y, d_Y) deux espaces métriques séparables. On munit $X \times Y$ de la métrique $d((x_1, y_1), (x_2, y_2)) = \max(d_X(x_1, x_2), d_Y(y_1, y_2))$.

1. Montrer que $f : (Z, d_Z) \to (X \times Y, d)$ est continue ssi $\pi_X \circ f$ et $\pi_Y \circ f$ sont continues.
2. Vérifier que les boules de $X \times Y$ sont des produits de boules, et que $X \times Y$ est séparable. Montrer que les produits de boules sont ouverts.
3. Montrer que $\mathcal{B}(X \times Y) \subseteq \mathcal{B}(X) \otimes \mathcal{B}(Y)$.
4. Soit U ouvert de Y. Montrer que $\{A \in \mathcal{B}(X) : A \times U \in \mathcal{B}(X \times Y)\}$ est une σ-algèbre qui contient les ouverts de X, en déduire que pour tout borélien A de X, $A \times U \in \mathcal{B}(X \times Y)$.

3. Piège gratuit : auriez-vous oublié de montrer que $\emptyset \in f^{-1}(\mathcal{B})$?

5. Soit A un borélien de X, montrer que $\{B \in \mathcal{B}(Y) : A \times B \in \mathcal{B}(X \times Y)\}$ est une σ-algèbre qui contient les ouverts de Y, et en déduire que $\mathcal{B}(X \times Y) = \mathcal{B}(X) \otimes \mathcal{B}(Y)$.
6. Supposons $\mathcal{A} = \sigma(\mathbb{A})$ et $\mathcal{B} = \sigma(\mathbb{B})$. Montrer qu'alors $\mathcal{A} \otimes \mathcal{B} = \sigma(\{A \times B : A \in \mathbb{A}, B \in \mathbb{B})\}$. On pourra s'inspirer des deux questions précédentes. En déduire une preuve plus simple de $\mathcal{B}(X \times Y) = \mathcal{B}(X) \otimes \mathcal{B}(Y)$.
7. Soient $(X, \mathcal{A})$, $(Y, \mathcal{B})$ et $(Z, \mathcal{C})$ des espaces mesurables. Montrer que $\mathcal{A} \otimes (\mathcal{B} \otimes \mathcal{C}) = (\mathcal{A} \otimes \mathcal{B}) \otimes \mathcal{C} = \sigma(\{A \times B \times C : A \in \mathcal{A}, B \in \mathcal{B}, C \in \mathcal{C}\}$. Définir un produit fini arbitraire d'espaces mesurables, et montrer que

$$\bigotimes_{i=1}^{n} \mathcal{A}_i \otimes \bigotimes_{j=1}^{m} \mathcal{A}_{j+n} = \bigotimes_{i=1}^{n+m} \mathcal{A}_i.$$

Soient $(X_n, d_n)_{n\in\mathbb{N}}$ séparables, on munit $X = \prod_{n\in\mathbb{N}} X_n$ de la distance [4]

$$d((x_n), (y_n)) = \sum_{n\in\mathbb{N}} \frac{1}{2^n} \frac{d_n(x_n, y_n)}{1 + d_n(x_n, y_n)}.$$

Exercice 9. 1. Vérifier que d est une distance, et que les projections π_n sont continues.

2. Montrer que $f : (Z, d_z) \to X$ est continue ssi $\pi_{X_n} \circ f$ est continue pour tout $n \in \mathbb{N}$, et montrer que (X, d) est séparable.
3. Montrer que toute boule de $\prod_{n\in\mathbb{N}} X_n$ s'écrit comme union dénombrable de $\prod_n B_n$ où $B_n = X_n$ sauf pour un nombre fini de n, auquel cas B_n est une boule de X_n.

Soient $(X_n \mathcal{A}_n)_{n\in\mathbb{N}}$ mesurables, on définit la **tribu produit** $\bigotimes_{n\in\mathbb{N}} \mathcal{A}_n$ par

$$\bigotimes_{n\in\mathbb{N}} \mathcal{A}_n = \sigma(\{\pi_n^{-1}(A_n) : A_n \in \mathcal{A}_n, n \in \mathbb{N}\}).$$

Exercice 10. Soient $(X_n, \mathcal{A}_n)_{n\in\mathbb{N}}$ mesurables.

1. Vérifier que $\bigotimes_{n\in\mathbb{N}} \mathcal{A}_n$ est la plus petite tribu telle que les π_{X_n} soient mesurables.
2. Montrer que si pour tout n, $A_n \in \mathcal{A}_n$, alors $\prod_{n\in\mathbb{N}} A_n \in \bigotimes_{n\in\mathbb{N}} \mathcal{A}_n$.
3. Montrer que si (U_n) est une suite d'ouverts de (X_n, d_n) espaces métriques, $\prod_n U_n$ peut ne pas être ouvert. Trouver une CNS pour qu'il le soit.
4. Montrer que $f : (Z, \mathcal{C}) \to \prod X_n$ est mesurable ssi pour tout $n \in \mathbb{N}$, $\pi_{X_n} \circ f$ est mesurable.
5. Montrer que $\mathcal{B}(\prod_{n\in\mathbb{N}} X_n) = \bigotimes_{n\in\mathbb{N}} \mathcal{B}(X_n)$. Pour $\supseteq$, on pourra se référer à l'exercice 8, questions 4 et 5.

4. On n'aurait pas besoin de quotienter par $1 + d_n(x_n, y_n)$ si on avait supposé les distances uniformément bornées.

Exercice 11. Soient $(X_n, \mathcal{A}_n)_{n\in\mathbb{N}}$ mesurables. On définit un **cylindre** comme un produit
$$C = \prod_{n\in\mathbb{N}} A_n$$
où $A_n = X_n$ *sauf en un nombre fini* de $n \in \mathbb{N}$. Par le question 2 de l'exercice précédent, les cylindres sont mesurables. Montrer que l'intersection d'un nombre fini de cylindres est un cylindre. Montrer que les cylindres engendrent une algèbre dont chaque élément peut s'écrire comme union disjointe de cylindres [5].

1.2 Abécédaire des ensembles et applications mesurables

La notion d'espace produit nous permet de traiter de la mesurabilité de nombreuses applications ; on a voulu ici lister toutes sortes de propriétés de mesurabilité, pour les ensembles comme pour les applications. Commençons par définir la droite réelle étendue :

Définition 1.6. On note $\bar{\mathbb{R}} = \{-\infty \cup \mathbb{R} \cup +\infty\}$ la droite réelle étendue, munie de la tribu borélienne $\mathcal{B}(\bar{\mathbb{R}}) = \sigma(\{[b, +\infty]\})$.

Exercice 12. Montrer que cette tribu est la tribu des boréliens pour $\bar{\mathbb{R}}$ muni de la distance
$$d_{\bar{\mathbb{R}}}(x, y) = |\arctan(x) - \arctan(y)|$$
où l'on pose $\arctan(-\infty) = -\pi/2$ et $\arctan(+\infty) = \pi/2$. Montrer que cette tribu induit sur $\mathbb{R}$ les boréliens usuels et que les singletons $\{-\infty\}$ et $\{+\infty\}$ sont boréliens.

On remarque que la distance $d_{\bar{\mathbb{R}}}$ est compatible avec les notions de "tendre vers $\pm\infty$" pour une suite réelle.

Proposition 1.7. *Soit $(X, \mathcal{A})$ un espace mesurable, soient $f_n : X \to \bar{\mathbb{R}}$ des fonctions mesurables, alors les fonctions suivantes, à valeur dans $\bar{\mathbb{R}}$ muni de sa tribu borélienne, sont mesurables :*

1. $x \mapsto \sup_{n\in\mathbb{N}} f_n(x)$,
2. $x \mapsto \limsup_{n\to\infty} f_n(x)$,
3. $x \mapsto \inf_{n\in\mathbb{N}} f_n(x)$,
4. $x \mapsto \liminf_{n\to\infty} f_n(x)$,

De plus, les ensembles suivants sont mesurables :

a. $\{x : f_n(x)$ *converge*$\}$,
b. $\{x : f_n(x) \to -\infty\}$,
c. $\{x : f_n(x) \to +\infty\}$.

Exercice 13. Montrer que étant donné une suite de fonctions positives mesurables (f_n), leur somme $\sum_{n\in\mathbb{N}} f_n$ est mesurable.

5. Une algèbre contient $\emptyset$, est stable par complémentaire et réunion finie.

Exercice 14. Soit $(X, \mathcal{A})$ un espace mesurable, et soit une suite $(A_n)_{n\in\mathbb{N}}$ d'éléments de $\mathcal{A}$. On définit $\limsup_{n\in\mathbb{N}} A_n$ comme l'ensemble des $x \in X$ tels que x appartient à une infinité de A_n.

1. Vérifier que $\limsup_{n\to\infty} A_n = \bigcap_{m\in\mathbb{N}} \bigcup_{n\geqslant m} A_n$. En déduire qu'un tel ensemble est mesurable.
2. On définit $\liminf_{n\to\infty} A_n = \bigcup_{m\in\mathbb{N}} \bigcap_{n\geqslant m} A_m$. Exprimer le complémentaire d'une $\liminf$.

2 Mesures

Définition 2.1. Soit $(X, \mathcal{A})$ un espace mesurable. Une **mesure** sur $(X, \mathcal{A})$ est une application $\mu : \mathcal{A} \to [0, +\infty]$ telle que :

1. $\mu(\emptyset) = 0$
2. Pour toute suite $(A_n)_{n\in\mathbb{N}}$ d'éléments disjoints de $\mathcal{A}$,
$$\mu(\bigsqcup_{n\in\mathbb{N}} A_n) = \sum_{n\in\mathbb{N}} \mu(A_n).$$

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

Exercice 15. Vérifier que la mesure de Dirac en $x \in X$ définie par $\delta_x(A) = \chi_A(x)$ est une mesure. Montrer que toute mesure finie sur un ensemble fini est combinaison linéaire de mesures de Dirac. Quelles sont les mesures finies sur un ensemble dénombrable X muni de la tribu $\mathcal{P}(X)$?

Exercice 16. Soit $(X, \mathcal{A}, \mu)$ un espace mesuré.

1. Soit $B \in \mathcal{A}$, montrer que si $A \in \mathcal{A}$ est inclus dans B, $\mu(B) = \mu(A) + \mu(B \setminus A)$. En particulier $\mu(A) \leqslant \mu(B)$.
2. Soit (A_n) une suite *croissante* d'éléments de $\mathcal{A}$, montrer que
$$\mu(\bigcup_{n\in\mathbb{N}} A_n) = \lim_{n\to\infty} \mu(A_n).$$
On pourra se ramener à une union disjointe en posant $B_n = A_n \setminus \big(\bigcup_{m<n} A_m\big)$.
3. Soi (A_n) une suite *décroissante* d'éléments de $\mathcal{A}$, avec $\mu(A_0) < +\infty$. Montrer que
$$\mu(\bigcap_{n\in\mathbb{N}} A_n) = \lim_{n\to\infty} \mu(A_n).$$
Trouver un contre-exemple à cette formule dans le cas où A_0 est de mesure infinie.

La technique exposée dans la question 2 de l'exercice 16 est très souvent usitée, ce qui paraît raisonnable vu que la “compatibilité avec la réunion disjointe” est essentiellement la seule propriété des mesures ! Passons maintenant à la notion de classe monotone, qui nous permettra notamment de montrer l'unicité de certaines mesures.

2.1 Théorème de la classe monotone

Définition 2.2. Soit X un ensemble. Un π**-système** sur X est ensemble de parties de X stable par intersection finie, contenant X.

Définition 2.3. Soit X un ensemble. Une **classe monotone** est un ensemble $\mathcal{C}$ de parties de X tel que

1. $\emptyset \in \mathcal{C}$,
2. Si $A, B \in \mathcal{C}$ et $A \subseteq B$, alors $B \setminus A \in \mathcal{C}$,
3. Si $(A_n)_{n\in\mathbb{N}}$ est une suite croissante d'éléments de $\mathcal{C}$, $\bigcup_{n\in\mathbb{N}} A_n \in \mathcal{C}$.

Exercice 17. Montrer que si $\mathbb{B} \subseteq \mathcal{P}(X)$, il existe une plus petite classe monotone contenant $\mathbb{B}$, que l'on appelle **classe monotone engendrée** par $\mathbb{B}$.

Théorème 2.4 (Lemme de la classe monotone)**.** *Soit $\mathbb{B}$ un π-système. Alors $\sigma(\mathbb{B})$ est égal à la classe monotone engendrée par $\mathbb{B}$.*

La preuve est laissée au lecteur, à travers l'exercice suivant :

Exercice 18. On note $\mathcal{B}$ la classe monotone engendrée par $\mathbb{B}$.

1. Montrer que $\mathcal{B} \subseteq \sigma(\mathbb{B})$. Il suffit alors de montrer que $\mathcal{B}$ est une σ-algèbre.
2. Montrer qu'il suffit de savoir que $\mathcal{B}$ est stable par intersection finie.
3. Soit $B \in \mathbb{B}$. Montrer que l'ensemble des $A \in \mathcal{B}$ tels que $A \cap B \in \mathcal{B}$ est une classe monotone. En déduire que pour tout $A \in \mathcal{B}$, $A \cap B \in \mathbb{B}$
4. Soit $A \in \mathcal{B}$. Montrer que l'ensemble des $B \in \mathcal{B}$ tels que $A \cap B \in \mathcal{B}$ est une classe monotone. Conclure.

Corollaire 2.5. *Soient μ et ν deux mesures de probabilités sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ qui coïncident sur $\{]-\infty, a] : a \in \mathbb{R}\}$. Alors $\mu = \nu$.*

Démonstration. Soit $\mathbb{B} = \{]-\infty, a] : a \in \mathbb{R}\} \cup \{\mathbb{R}\}$, μ et ν coïncident sur le π-système $\mathbb{B}$ par hypothèse ; on note alors $\mathcal{B} = \{B \in \mathcal{B}(\mathbb{R}) : \mu(B) = \nu(B)\}$. On sait que $\mathcal{B}(\mathbb{R}) = \sigma(\{]-\infty, a] : a \in \mathbb{R}\})$, il suffit donc de montrer que $\mathcal{B}$ est une classe monotone pour appliquer le lemme de classe monotone et conclure. Que $\mathcal{B}$ soit une classe monotone est conséquence des résultats de l'exercice 16. □

Nous pouvons désormais montrer l'unicité d'une mesure μ sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ telle que pour tout $a \leqslant b$, $\mu([a,b]) = b - a$. Soit donc ν une autre mesure telle que pour tout $a \leqslant b$, $\mu([a,b]) = b - a$. Tout d'abord, remarquons que les intervalles fermés $[a,b]$ forment un π-système, et comme $\mathbb{R} = \bigcup_{n\in\mathbb{N}}[-n,n]$, on a $\mu(\mathbb{R}) = \nu(\mathbb{R}) = +\infty$. Par contre l'ensemble des $A \in \mathcal{B}(\mathbb{R})$ tels que $\mu(A) = \nu(A)$ n'est pas une classe monotone a priori, essentiellement parce que $\mu(\mathbb{R}) = \nu(\mathbb{R}) = +\infty$, ce qui empêche d'avoir la stabilité par $A \subseteq B \mapsto B \setminus A$. Mais on peut considérer $\mathcal{C}_n = \{B \in \mathcal{B}([-n,n]) : \mu(B) = \nu(B)\}$[6]. Il contient le π-système $\mathbb{B}_n = \{[a,b] : -n \leqslant a \leqslant b \leqslant n\}$, et comme sa mesure est finie c'est une classe monotone, qui par le lemme de classe monotone

6. Remarquons que $\mathcal{B}([-n,n]) = \{A \cap [-n,n] : A \in \mathcal{B}(\mathbb{R})$

contient $\mathcal{B}([-n,n])$. Soit maintenant $A \in \mathcal{B}(\mathbb{R})$, on écrit $A = \bigcup_{n\in\mathbb{N}} A \cap [-n,n]$, qui est une union croissante. Alors par l'exercice 16,

$$\mu(A) = \lim_{n\to\infty} \mu(A \cap [-n,n]) = \lim_{n\to\infty} \nu(A \cap [-n,n]) = \nu(A).$$

Ainsi on a bien l'unicité d'une mesure μ sur les boréliens de $\mathbb{R}$ satisfaisant $\mu([a,b]) = b - a$ pour $a \leqslant b$. Son existence est admise :

Théorème 2.6. *Il existe une unique mesure λ sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ telle que pour tout $a < b$, $\lambda([a,b]) = b - a$.*

Exercice 19. Montrer que pour tout $x \in \mathbb{R}$, $\mu(\{x\}) = 0$.

Il nous faut maintenant apprendre à intégrer des fonctions par rapport à λ, avant de pouvoir constater que l'on a généralisé l'intégrale de Riemann.

3 Intégration

3.1 Intégration de fonctions positives

Soit $(X, \mathcal{A})$ un espace mesurable. Les **fonctions simples** sur X sont les fonctions $s : X \to \mathbb{R}^+$ mesurables qui prennent un nombre fini de valeur. Une telle fonction s'écrit alors de manière unique

$$s = \sum_{i=1}^{n} \alpha_i \chi_{A_i},$$

où $0 \leqslant \alpha_1 < \cdots < \alpha_n$ et les A_i sont mesurables disjoints. On a nécessairement $A_i = \{x \in X : s(x) = \alpha_i\}$. Soit μ une mesure sur $\mathcal{A}$. On peut alors définir

$$\int_X s d\mu = \sum_{i=1}^{n} \alpha_i \mu(A_i).$$

Supposons que l'on ait une partition $(B_j)_{j=1}^m$ de X où $s_{\restriction B_j}$ est constante égale à β_j. On vérifie alors que $\int_X s d\mu = \sum_{j=1}^m \beta_j \mu(B_j)$ (utiliser l'additivité de μ).

Soient maintenant s, t deux fonctions simples, en utilisant le fait précédent, on montre que

$$\int_X (s+t) d\mu = \int_X s d\mu + \int_X t d\mu.$$

On en déduit $\int_X s d\mu \leqslant \int_X t d\mu$ dès lors que $s \leqslant t$. Nous sommes alors en mesure de définir l'intégrale d'une fonction positive mesurable $f : X \to \mathbb{R}^+$:

$$\int_X f d\mu = \sup_{s \leqslant f \text{ simple}} \int_X s d\mu.$$

Remarquons que cette définition est compatible avec la définition pour les fonctions simples, puisque si s est simple, c'est la plus grande fonction simple $\leqslant s$!. Le théorème suivant, fondamental, est aussi appelé théorème de Beppo-Levi.

Théorème 3.1 (Théorème de convergence monotone)**.** *Soit $(X, \mathcal{A}, \mu)$ un espace mesuré, et soit (f_n) une suite croissante de fonctions mesurables positives (on autorise $f_n(x) = +\infty$). Alors*

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

La proposition suivante nous donne une bonne occasion d'utiliser la convergence monotone :

Proposition 3.2. *Soit $f : (X, \mathcal{A}) \to \mathbb{R}^+ \cup \{+\infty\}$ mesurable. Il existe une suite croissante de fonctions simples f_n telle que $f_n(x) \to f(x)$ $[n \to +\infty]$ pour tout $x \in X$.*

Corollaire 3.3. *Soient f et g deux fonctions mesurables positives. Alors*

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

Démonstration. Soient par la proposition précédente f_n et g_n simples, croissantes, convergeant simplement respectivement vers f et g. Alors $f_n + g_n$ est une suite croissante de fonctions simples qui converge simplement vers $f + g$. Mais pour les fonctions simples $\int_X (f_n + g_n) d\mu = \int_X f_n d\mu + \int_X g_n d\mu$. On applique alors trois fois le théorème de convergence monotone pour conclure. □

Un cas particulier de l'énoncé ci-dessous est le fait que si $0 \leqslant f \leqslant g$, $\int_X f \leqslant \int_X g$. En utilisant encore la convergence dominée et le corollaire précédent, on obtient la généralisation suivante :

Corollaire 3.4. *Soient $f_n : X \to [0, +\infty]$ mesurables, alors*

$$\int_X \sum_{n\in\mathbb{N}} f_n d\mu = \sum_{n\in\mathbb{N}} \int_X f d\mu$$

Soit $f : (X, \mathcal{A}, \mu) \to [0, +\infty]$ mesurable, on définit la mesure μ_f de densité f sur X par

$$\mu_f(A) = \int_A f d\mu = \int_X \chi_A f d\mu.$$

Pour voir que c'est une mesure, il faut appliquer le corollaire précédent. μ_f est appelée la mesure de **densité** f par rapport à μ.

Exemple 3.5. Soit $f : \mathbb{R} \to [0, 1]$ définie par $f(x) = \frac{1}{\sqrt{\pi}} e^{-x^2}$. Elle est d'intégrale 1, donc est la densité d'une mesure de probabilités sur $\mathbb{R}$: la **mesure gaussienne**.

Finissons cette section avec un "lemme" tout aussi fondamental :

Théorème 3.6 (Lemme de Fatou)**.** *Soit une suite de fonctions $f_n : X \to [0, +\infty]$ mesurables. Alors*

$$\int_X \liminf f_n d\mu \leqslant \liminf \int_X f_n d\mu$$

Démonstration. Soit $g_n(x) = \inf_{m \geqslant n} f_m(x)$, alors les g_n sont croissantes, et leur limite est par définition $\liminf f_n$. On a donc

$$\lim_{n\to\infty} \int_X g_n = \int_X \liminf f_n.$$

Mais pour tout n, $g_n \leqslant f_n$ donc $\int_X g_n \leqslant \int_X f_n$, et en passant à la liminf, $\liminf \int g_n \leqslant \liminf \int_X f_n$, mais $\int g_n$ converge vers $\int \liminf f_n$, donc au final $\int_X \liminf f_n \geqslant \int_X \liminf f_n$. □

3.2 Intégration de fonctions réelles

Soit $f : X \to \bar{\mathbb{R}}$ mesurable, alors $f_+ = \max(f, 0)$ et $f_- = \max(-f, 0)$ sont mesurables, on a $f = f_+ - f_-$ et $|f| = f_+ + f_-$. On dit que f est **intégrable** si $\int_X |f| < +\infty$, alors $\int_X f_+ < +\infty$ et $\int_X f_- < +\infty$, donc on peut définir l'**intégrale** de f :

$$\int_X f d\mu = \int_X f_+ d\mu - \int_X f_- d\mu.$$

Remarquons que si f est positive, et $a \geqslant 0$, $\int af = a \int f$ puisque $s \leqslant f$ ssi $as \leqslant af$ et que la formule est claire pour les fonctions simples.

Proposition 3.7. *L'intégrale est une forme linéaire sur l'espace vectoriel des fonctions intégrables.*

Démonstration. Il faut tout d'abord voir que les fonctions intégrables forment un espace vectoriel. C'est le cas car on a $|f + g| \leqslant |f| + |g|$, et $|af| = |a|\,|f|$ donc on peut appliquer la remarque précédente.. Ensuite, si $h = f + g$, on a $h = h_+ - h_-$, $f = f_+ - f_-$ et $g = g_+ - g_-$, ce qui donne $h_+ - h_- = f_+ - f_- + g_+ - g_-$, et donc

$$h_+ + f_- + g_- = h_- + f_+ + g_+.$$

Reste alors à appliquer le corollaire 3.3. □

Exercice 20. Vérifier que $\left|\int f\right| \leqslant \int |f|$.

Nous pouvons désormais énoncer un théorème crucial : le théorème de convergence dominée, dont la preuve fait l'objet du prochain exercice.

Théorème 3.8. *Soit (f_n) une suite de fonctions $X \to \bar{\mathbb{R}}$ telle que pour tout n, $|f_n| \leqslant g$ où g est intégrable. On suppose en outre que pour tout $x \in X$, $f_n(x) \to f(x)$. Alors*

$$\int_X f_n d\mu \to \int_X f d\mu.$$

Exercice 21. On cherche à prouver le théorème de convergence dominée.

1. Montrer qu'il suffit d'avoir $\limsup_{n\to\infty} \int_X |f_n - f| = 0$
2. En utilisant le lemme de Fatou sur $2g - |f_n - f|$, conclure.

3.3 Le presque partout

Définition 3.9. Soit $(X, \mathcal{A}, \mu)$ un espace mesuré. On dit qu'un ensemble mesurable $A \subseteq X$ est **négligeable** (sous entendu par rapport à μ) si $\mu(A) = 0$. On dit qu'il est **de mesure pleine** si $X \setminus A$ est négligeable. On dira d'une propriété $P(x)$ qu'elle vraie μ **presque partout** si l'ensemble qui la définit est de mesure pleine.

Exemple 3.10. Soient $f, g : (X, \mathcal{A}, \mu) \to (Y, \mathcal{B})$ mesurables. On dit que $f = g$ presque partout si $\mu(\{x \in X : f(x) = g(x)\}) = 0$ (on pourra vérifier qu'un tel ensemble est mesurable).

L'intérêt principal de la notion de presque partout est que c'est le bon "filtre" pour regarder les fonctions à partir de leurs intégrales. On a par exemple les deux propositions suivantes :

Proposition 3.11. *Soit $f : (X, \mathcal{A}, \mu) \to \bar{\mathbb{R}}$ mesurable. Alors si f est intégrable, on a$-\infty < f < +\infty$ presque partout.*

Démonstration. Dans le cas contraire, on aurait $|f| = +\infty$ sur un ensemble A de mesure non nulle, mais alors $\int_X |f| \geqslant \int_A |f| = \mu(A) \cdot +\infty = +\infty$, et donc f n'est pas intégrable. $\square$

Proposition 3.12. *Soit $f : (X, \mathcal{A}, \mu) \to [0, +\infty]$ mesurable. Alors $\int_X f d\mu = 0$ ssi $f = 0$ μ-presque partout.*

Démonstration. Soit pour $n \in \mathbb{N}^*$, $A_n = \{x \in X : f(x) \geqslant 1/n\}$. Alors $\int_X f \geqslant \int_{A_n} f \geqslant \mu(A_n)/n$, donc chaque A_n est de mesure nulle, donc leur union est de mesure nulle, mais cette union est précisément l'ensemble des $x \in X$ tels que $f(x) > 0$. Ainsi $f = 0$ presque partout. $\square$

Corollaire 3.13. *Soient f et g deux fonctions, si $f = g$ μ presque partout, alors $\int_X f d\mu = \int_X g d\mu$.*

Les deux propositions précédentes utilisent le lemme suivant, qu'il convient d'isoler tant son usage est fréquent :

Lemme 3.14 (Lemme de Tchebychev)**.** *Soit $f : (X, \mathcal{A}, \mu) \to [0, +\infty]$ mesurable. Alors pour tout $t > 0$, si on note $A_t = \{x \in X : f(x) \geqslant t\}$:*

$$\mu(A_t) \leqslant \frac{\int_X f d\mu}{t}$$

Finissons cette section avec un exercice mettant en avant des ensembles négligeables "usuels" des réels.

Exercice 22.

1. Soit $A \subseteq \mathbb{R}$ dénombrable, montrer que A est négligeable pour la mesure de Lebesgue. (on rappelle qu'un singleton est de mesure nulle).
2. (Ensembles de Cantor) Pour $n \in \mathbb{N}$, on identifie 2^n à l'ensemble $\{0,1\}^{\{0,...,n-1\}}$ des applications de $\{0, ..., n-1\}$ dans $\{0, 1\}$. On a une application naturelle de restriction

$$\begin{aligned} 2^{n+1} &\to 2^n \\ k &\mapsto k_{\restriction\{0,...,n-1\}}, \end{aligned}$$

que l'on notera encore $k \mapsto k_{\restriction n}$.
Soit (λ_n) une suite d'éléments de $[0,1[$.on définit par récurrence sur $n \in \mathbb{N}$ une famille d'intervalles $(I_n^k)_{k \in 2^n, n \in \mathbb{N}}$ qui s'écrivent $I_n^k = [a_n^k, b_n^k]$, de longueur $l_n = b_n^k - a_n^k$ indépendante de k.
On commence par $I_n^0 = [0,1]$, puis étant donné $(I_n^k)_{k \in 2^n}$ on pose, pour $k \in 2^{n+1}$:

- Si $k(n) = 0$, $I_{n+1}^k = \left[a_n^{k_{\restriction n}}, a_n^{k_{\restriction n}} + \dfrac{l_n(1-\lambda_n)}{2}\right]$
- Si $k(n) = 1$, $I_{n+1}^k = \left[b_n^{k_{\restriction n}} - \dfrac{l_n(1-\lambda_n)}{2}, b_n^{k_{\restriction n}}\right]$

On vérifie aisément par récurrence que à n fixé, de tels intervalles sont disjoints, et que pour tout $k \in 2^{n+1}$ on a $I_k^{n+1} \subseteq I_{k_{\restriction n}}^n$. Il est également clair que $l_{n+1} = \frac{l_n(1-\lambda_n)}{2}$. Soit alors $K_n = \bigcup_{k \in 2^n} I_n^k$.

(a) Vérifier que les K_n sont décroissants et mesurables.

(b) Calculer la mesure de leur intersection en fonction de la suite (λ_n). Cette intersection est appelée un **ensemble de Cantor**, et est homéomorphe à $\{0,1\}^{\mathbb{N}}$.

(c) Montrer qu'il existe des ensembles négligeables ayant la cardinalité de $\mathbb{R}$.

4 Mesure produit et théorèmes de Fubini

4.1 Mesure produit

Dans toute cette section, $(X, \mathcal{A})$ et $(Y, \mathcal{B})$ seront deux espaces mesurables.

Définition 4.1. Soit $A \subseteq X \times Y$, alors pour tout $x \in X$, on définit la section de A selon x

$$A_x = \{y \in Y : (x,y) \in A\}.$$

De même on définit pour $y \in Y$ la section de A selon y par

$$A^y = \{x \in X : (x,y) \in A\}.$$

Exercice 23. Montrer que pour tout $A \in \mathcal{A} \otimes \mathcal{B}$, et pour tout $(x,y) \in X \times Y$, A_x ainsi que A^y sont mesurables.

Définition 4.2. Une mesure μ sur $(X, \mathcal{A})$ est dite **finie** si $\mu(X) < +\infty$, elle est dite **σ-finie** si il existe une suite croissante (X_n) de sous ensembles mesurables de X tels que $\mu(X_n) < +\infty$ et $\bigcup_{n \in \mathbb{N}} X_n = X$.

Exercice 24. Montrer que la mesure de Lebesgue est σ-finie. Montrer que dans la définition de σ-finie, il est inutile de demander les X_n croissants.

On suppose dorénavant X muni d'une mesure μ, Y d'une mesure ν, qui sont toutes deux σ-finies. On va construire une mesure $\mu \otimes \nu$ sur $\mathcal{A} \otimes \mathcal{B}$ en intégrant les mesures des sections (rappelons que celles-ci sont mesurables par l'exercice 23), autrement dit on voudrait poser :

$$\mu \otimes \nu(C) = \int_X \nu(C_x) d\mu(x) \tag{2}$$

Exercice 25. Cet exercice détaille la construction de la mesure produit.

1. Vérifions tout d'abord que la formule (2) a un sens, ce qui revient à voir que $x \mapsto \nu(C_x)$ est mesurable pour tout $C \in \mathcal{A} \otimes \mathcal{B}$.
   (a) On se restreint tout d'abord au cas où ν est finie, on note $\mathcal{R} = \{A \times B : A \in \mathcal{A}, B \in \mathcal{B}\}$ le π système des rectangles mesurables. Soit $\mathcal{C} = \{C \in \mathcal{A} \otimes \mathcal{B} : x \mapsto \nu(C_x) \text{ est mesurable}\}$. Montrer que $\mathcal{C}$ est une classe monotone contenant $X \times Y$ ainsi que $\mathcal{R}$, et conclure.
   (b) Dans le cas où ν est σ-finie, exprimer $\nu(C_x)$ en fonction de $\nu(C_x \cap Y_n)$ où les Y_n sont croissants d'union Y, et de mesure finie. Conclure.
2. Montrer que la formule (2) définit une mesure sur $\mathcal{A} \otimes \mathcal{B}$.
3. Il l'agit maintenant de voir l'unicité d'une mesure $\mu \otimes \nu$ telle que pour tout $A \in \mathcal{A}, B \in \mathcal{B}$,
$$\mu \otimes \nu(A \times B) = \mu(A)\nu(B).$$
   (a) Vérifier que la mesure construite précédemment satisfait cette équation.
   (b) En s'inspirant de ce qui a été fait pour l'unicité de la mesure de Lebesgue (cf. paragraphe avant le théorème 2.6), montrer l'unicité.
   (c) Montrer que la mesure produit $\mu \otimes \nu$, désormais uniquement définie, satisfait également
$$\mu \otimes \nu(C) = \int_Y \mu(C^y) d\nu(y).$$

En résumé, on a montré le théorème suivant :

Théorème 4.3 (Définition de la mesure produit)**.** *Soient $(X, \mathcal{A}, \mu)$ et $(Y, \mathcal{B}, \nu)$ deux espaces mesurés σ-finis. Alors il existe une unique mesure $\mu \otimes \nu$ sur $\mathcal{A} \otimes \mathcal{B}$ telle que pour tout $A \in \mathcal{A}, B \in \mathcal{B}$,*
$$\mu \otimes \nu(A \times B) = \mu(A)\nu(B).$$
De plus, cette mesure satisfait l'équation suivante, pour tout $C \in \mathcal{A} \otimes \mathcal{B}$:
$$\mu \otimes \nu(C) = \int_X \nu(C_x) d\mu(x) = \int_Y \mu(C^y) d\nu(y).$$
*On l'appelle **mesure produit** de μ et ν.*

4.2 Théorème de Fubini-Tonelli

Exercice 26. Soient X, Y et Z trois espaces mesurables, on munit $X \times Y$ de la tribu produit. Soit $f : X \times Y \to Z$ mesurable, montrer que à x fixé, $y \mapsto f(x, y)$ est mesurable, et que à y fixé, $x \mapsto f(x, y)$ l'est également.

Théorème 4.4. *Soient $(X, \mathcal{A}, \mu)$ et $(Y, \mathcal{B}, \nu)$ deux espaces mesurés σ-finis. Alors pour toute fonction $f : X \times Y \to [0, +\infty]$ qui est $\mathcal{A} \otimes B$ mesurable, on a que $x \mapsto \int_Y f(x, y) d\nu(y)$ et $y \mapsto \int_X f(x, y) d\mu(x)$ sont mesurables, et*
$$\int_{X \times Y} f(x, y) d\mu \otimes \nu(x, y) = \int_X \int_Y f(x, y) d\nu(y) d\mu(x) = \int_Y \int_X f(x, y) d\mu(x) d\mu(y).$$

Démonstration. Vérifions tout d'abord la mesurabilité de $x \mapsto \int_Y f(x,y)d\nu(y)$. Par symétrie, on aura également la mesurabilité de $y \mapsto \int_X f(x,y)d\mu(x)$. Tout d'abord, si $f = \chi_C$ est une fonction caractéristique d'ensemble mesurable, le théorème 4.3 nous assure que $x \mapsto \int_Y \chi_C(x,y)d\nu(y) = \nu(C_x)$ est mesurable. Par linéarité de l'intégrale et mesurabilité de l'addition, $x \mapsto \int_Y f(x,y)d\nu(y)$ est mesurable pour f simple. Enfin, si f est positive quelconque, on utilise la proposition 3.2 pour écrire f comme limite croissante de fonctions simples f_n. Alors par théorème de convergence monotone, pour tout $x \in X$ on a $\int_Y f(x,y)d\nu(y) = \lim_{n\to\infty} \int_Y f_n(x,y)d\nu(y)$. Comme les f_n sont simples, $x \mapsto \int_Y f_n(x,y)d\nu(y)$ est mesurable, donc sa limite $x \mapsto \int_Y f(x,y)d\nu(y)$ l'est également.

La formule à prouver est vraie pour les fonctions caractéristiques par le théorème 4.3, donc vraie par linéarité pour les fonctions simples. Pour f positive quelconque, on écrit encore $f = \lim_{n\to\infty} f_n$ où les f_n sont simples croissantes en n. Alors en appliquant trois fois le théorème de convergence monotone, et le fait que l'intégrale croît avec les fonctions, on obtient

$$\begin{aligned}\int_{X\times Y} f(x,y)d\mu\otimes\nu(x,y) &= \lim_{n\to\infty}\int_{X\times Y} f_n(x,y)d\mu\otimes\nu(x,y)\\ &= \lim_{n\to\infty}\int_X\int_Y f_n(x,y)d\nu(y)d\mu(x)\\ &= \int_X \lim_{n\to\infty}\int_Y f_n(x,y)d\nu(y)d\mu(x))\\ &= \int_X\int_Y f(x,y)d\nu(y)d\mu(x).\end{aligned}$$

Le fait que $\int_{X\times Y} f(x,y)d\mu\otimes\nu(x,y) = \int_Y\int_X f(x,y)d\mu(x)d\nu(y)$ s'obtient de la même manière. □

4.3 Théorème de Fubini-Lebesgue

Commençons par revenir sur la notion de fonction mesurable : si on a $A \subseteq X$ de mesure pleine, et $f : A \to \mathbb{R}$ mesurable, on dira quand même que f est mesurable. Et elle sera **intégrable** si $\int_A |f| < +\infty$. Cela revient à identifier f à la fonction $\tilde{f} : x \mapsto f(x)$ si $x \in A$, et 0 sinon. Le bon cadre pour parler de ces notions est celui des fonctions L^1, mais nous avons choisi de ne pas l'évoquer par souci de simplicité.

Théorème 4.5. *Soient $(X, \mathcal{A}, \mu)$ et $(Y, \mathcal{B}, \nu)$ deux espaces mesurés σ-finis. Soit $f : X \times Y \to \bar{\mathbb{R}}$ une fonction $\mathcal{A}\otimes B$ mesurable et intégrable pour la mesure produit, i.e. $\int_{X\times Y} |f(x,y)|\, d\mu\otimes\nu(x,y) < +\infty$. Alors les fonctions $x \mapsto \int_Y f(x,y)d\nu(y)$ et $y \mapsto \int_X f(x,y)d\mu(x)$ sont respectivement μ et ν presque partout définies, mesurables et intégrables. De plus, on a la formule*

$$\int_{X\times Y} f(x,y)d\mu\otimes\nu(x,y) = \int_X\int_Y f(x,y)d\nu(y)d\mu(x) = \int_Y\int_X f(x,y)d\mu(x)d\mu(y).$$

Démonstration. Par le théorème de Fubini-Tonelli, on a

$$\int_{X\times Y} |f(x,y)|\, d\mu\otimes\nu(x,y) = \int_X\int_Y |f(x,y)|\, d\nu(y)d\mu(x) = \int_Y\int_X |f(x,y)|\, d\mu(x)d\mu(y) < +\infty.$$

Soient X' et Y' sur lesquels $\int_Y |f(x,y)|\, d\nu(y) < +\infty$ et $\int_X |f(x,y)|\, d\mu(x) < +\infty$ respectivement, par la proposition 3.11 ces ensembles sont de mesure pleine. En restreignant au produit de ces ensembles, qui est de mesure pleine [7], on voit qu'on peut supposer $\int_Y |f(x,y)|\, d\nu(y) < +\infty$, et $\int_X |f(x,y)|\, d\mu(x) < +\infty$ pour tous $(x,y) \in X \times Y$. Toujours par le théorème de Fubini-Tonelli, on a $x \mapsto \int_Y f_+(x,y) d\nu(y)$ et $y \mapsto \int_X f_-(x,y) d\mu(x)$ qui sont mesurables, et alors $x \mapsto \int_Y f(x,y) d\nu(y) = \int_Y f_+(x,y) d\nu(y) - \int_Y f_-(x,y) d\nu(y)$ est mesurable. Le même argument donne que $y \mapsto \int_X f(x,y) d\mu(x)$ est mesurable, et alors la formule du théorème de Fubini Tonelli et le fait que $\int_{X\times Y} |f(x,y)|\, d\mu \otimes \nu(x,y) < +\infty$ nous donne que $x \mapsto \int_Y f(x,y) d\nu(y)$ et $y \mapsto \int_X f(x,y) d\mu(x)$ sont intégrables (on utilise également $\left|\int_Y f(x,y) d\nu(y)\right| \leqslant \int_Y |f(x,y)|\, d\nu(y)$...).

Reste alors à écrire $f = f_+ - f_-$ et à utiliser la formule de Fubini-Tonelli pour obtenir la formule de Fubini-Lebesgue. □

Ainsi, l'emploi du théorème de Fubini-Lebesgue est justifié dès lors que $\int_{X\times Y} |f(x,y)|\, d\mu \otimes \nu(x,y) < +\infty$. Pour justifier un tel fait, il faut souvent utiliser le théorème de Fubini-Tonelli afin de calculer cette intégrale de fonction positive. Attention donc à ne pas confondre les deux.

4.4 Produit infini de mesures

Il peut être bon de relire la fin de la section 1.1 avant d'aborder ce théorème.

Théorème 4.6. *Soit $(X_n, \mathcal{A}_n, \mu_n)$ une suite d'espaces de probabilités (autrement dit des espaces mesurés où $\mu_n(X_n) = 1$). Il existe une unique mesure μ sur $(\prod_n X_n, \bigotimes_{n\in\mathbb{N}} \mathcal{A}_n$ telle que pour tout cylindre $C = A_1 \times \cdots \times A_n \times \prod_{m>n} X_m$,*

$$\mu(C) = \prod_{i=1}^{n} \mu_i(A_i)$$

Exercice 27. Montrer l'unicité de μ.

L'existence est admise et fait appel au théorème de Carathéodory. Celui-ci est en fait incontournable dès lors qu'il s'agit de définir la mesure de Lebesgue sur $[0,1]$, ou la mesure produit standard sur $\{0,1\}^{\mathbb{N}}$ qui est $(1/2\delta_0 + 1/2\delta_1)^{\otimes\mathbb{N}}$.

5 Compléments

5.1 Mesure image

Définition 5.1. Soit $f : (X, \mathcal{A}, \mu) \to (Y, \mathcal{B})$ mesurable. La **mesure image** par f de μ, notée $f_*\mu$, est définie par : pour tout $B \in \mathcal{B}$,

$$f_*\mu(B) = \mu(f^{-1}(B)).$$

Exercice 28. 1. Vérifier que $f_*\mu$ est bien définie, et que c'est une mesure.

7. Ce qui ne change donc pas la valeur des intégrales dans la formule

2. Montrer que si μ est une mesure de probabilités, alors $f_*\mu$ aussi.
3. Est-ce-que si μ est σ-finie, alors $f_*\mu$ également ?

Définition 5.2. On dit que $f : (X, \mathcal{A}, \mu) \to (Y, \mathcal{B}, \nu)$ **préserve la mesure** si $f_*\mu = \nu$.

Exercice 29. Donnons un exemple non trivial d'application préservant la mesure.

1. Soit $f : (X, \mathcal{A}, \mu) \to (Y, \mathcal{B})$ mesurable. Montrer que $\{B \in \mathcal{B} : \mu(f^{-1}(B)) = \nu(B)\}$ est une σ-algèbre [8].
2. En utilisant une partie génératrice bien choisie de la σ-algèbre de $[0,1]$, montrer que la fonction
$$f : \begin{array}{rcl} \left(\{0,1\}^{\mathbb{N}}, (1/2\delta_0 + 1/2\delta_1)^{\otimes\mathbb{N}}\right) & \to & ([0,1], \lambda) \\ (x_n)_{n\in\mathbb{N}} & \mapsto & \sum_{n\in\mathbb{N}} \frac{1}{2^{n+1}} x_n \end{array}$$
préserve la mesure.

Remarque. Cet exercice montre qu'on aurait pu d'abord construire la mesure produit infini pour obtenir la mesure de Lebesgue. Cependant, le point difficile dans l'existence de la mesure de Lebesgue est le théorème de Carathéodory, qui joue également un rôle crucial dans la preuve de l'existence d'une mesure produit infini.

Théorème 5.3. *Soit $f : (X, \mathcal{A}, \mu) \to (Y, \mathcal{B})$ mesurable. Alors pour tout $g : Y \to \mathbb{R}$ intégrable, on a*
$$\int_X g(f(x)) d\mu(x) = \int_Y g(y) df_*\mu(y).$$
De plus, g est intégrable ssi $g \circ f$ l'est.

Autrement dit, pour intégrer $g \circ f$, on a juste besoin de connaître g et $f_*\mu$. Ce théorème est important en probabilités, il dit que pour calculer l'espérance d'une fonction dépendant d'une variable aléatoire, il suffit de connaître la loi de celle-ci... La preuve suit un schéma classique (vrai pour les fonctions caractéristiques, donc pour les fonctions simples, donc pour les fonctions positives, donc pour les fonctions intégrables). Écrire le détail de cette parenthèse est un bon exercice [9].

5.2 Espace L^1

Soit $(X, \mathcal{A}, \mu)$ un espace mesuré, on note $\mathcal{L}^1(X, \mathcal{A}, \mu)$ l'espace des fonctions intégrables sur X. On le munit de la prénorme [10] N définie par
$$N(f) = \int_X |f| \, d\mu.$$
La vérification du fait que ce soit une prénorme est laissée en exercice. On peut alors quotienter par l'espace vectoriel $\{x : N(x) = 0\}$, pour obtenir l'espace vectoriel normé
$$\mathrm{L}^1(X, \mathcal{A}, \mu) = \mathcal{L}^1(X, \mathcal{A}, \mu) / \{f : \int_X |f| \, d\mu = 0\}.$$

8. Pour la stabilité par union dénombrable, écrire l'union dénombrable comme union disjointe
9. Si ça ne vous paraît pas clair, relisez les preuves des théorèmes de Fubini.
10. Une norme sans $N(x) = 0 \Rightarrow x = 0$ (axiome de séparation)

On le note parfois $\mathrm{L}^1(X)$ ou $\mathrm{L}^1(X,\mu)$. La norme induite est notée $\|\cdot\|_1$. Notons que par la proposition 3.12, on peut considérer que l'on a quotienté $\mathcal{L}^1(X)$ par la relation d'équivalence $f\mathcal{R}_\mu g$ ssi $f=g$, μ presque partout. On identifie souvent une fonction et sa classe d'équivalence dans $\mathrm{L}^1(X)$, par exemple on écrira directement

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

En particulier, dans le théorème de Fubini-Lebesgue, on aurait pu directement dire que la fonction $x \mapsto \int_Y f(x,y)d\nu(y)$ est dans $\mathrm{L}^1(X)$.

Pour la preuve du théorème suivant, qui est au programme de l'agrégation, on pourra consulter [Rud98], théorème 3.11. Il ne nous servira pas pour la suite, on l'a inclus par souci de complétude.

Théorème 5.4. $\mathrm{L}^1(X)$ *est un espace complet.*

5.3 Convergence de fonction intégrables

La norme que nous avons introduite sur $\mathrm{L}^1(X)$ en fait un espace vectoriel normé, en particulier on a une notion naturelle de convergence pour les suites de fonctions intégrables :

Définition 5.5. On dit qu'une suite (f_n) de fonctions intégrables converge L^1 vers f si $\|f_n - f\|_1 \to 0$.

Il existe trois autres notions de convergences intéressantes lorsque μ est une mesure de probabilités :

- La convergence presque partout : $f_n \to f$ presque partout si il existe un ensemble A de mesure pleine, avec $f_n(x) \to f(x)$ pour tout $x \in A$. (Notons que cela fait bien sens dans $\mathrm{L}^1(X)$, autrement dit que ça ne dépend pas des représentants choisis).
- La convergence en probabilité : $f_n \to f$ en probabilité si pour tout $\epsilon > 0$,

$$\mu(\{x : |f_n(x) - f(x)| > \epsilon\}) \to 0$$

- La convergence en loi : $f_n \to f$ en loi si pour toute fonction g réelle *continue* bornée, $\int_X g \circ f_n \to \int_X g \circ f$.

Le lemme 3.14 implique qu'une suite de fonctions qui converge L^1 converge en probabilités. Quelles autres implications pouvez-vous établir entre ces différentes notions de convergence ? La réponse dans le cours de proba !

Références

[Rud98] W. Rudin. *Analyse réelle et complexe : Cours et exercices.* Dunod, 1998.