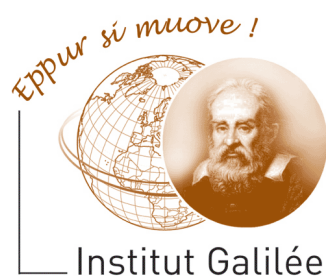

Intégration & Probabilités

Laurent Tournier



Intégration

0. PRÉLIMINAIRES SUR LES ENSEMBLES

1 Notations

Soit E un ensemble.

- Si A est une partie de E (c'est-à-dire $A \subset E$), on note $A^c = E \setminus A$ le complémentaire de A .
- On note $\mathcal{P}(E)$ l'ensemble des parties de E .
- Si la famille $(A_i)_{i \in I}$ de parties de E est disjointe (c'est-à-dire que $A_i \cap A_j = \emptyset$ pour tous $i \neq j$ dans I), on peut noter $\bigsqcup_{i \in I} A_i$ leur réunion (le “+” rappelle que les A_i sont disjoints).
- Si la suite $(A_n)_{n \in \mathbb{N}}$ de parties de E est croissante (c'est-à-dire que $A_n \subset A_{n+1}$ pour tout $n \in \mathbb{N}$), on peut noter $\bigcup_{n \in \mathbb{N}} A_n$ leur réunion (la flèche rappelle que la suite est croissante).
- Si la suite $(A_n)_{n \in \mathbb{N}}$ de parties de E est décroissante (c'est-à-dire que $A_{n+1} \subset A_n$ pour tout $n \in \mathbb{N}$), on peut noter $\bigcap_{n \in \mathbb{N}} A_n$ leur intersection (la flèche rappelle que la suite est décroissante).
- Si A est une partie de E , on note $\mathbf{1}_A$ sa fonction indicatrice :

$$\left| \begin{array}{lcl} \mathbf{1}_A : & E & \rightarrow \{0,1\} \\ & x & \mapsto \mathbf{1}_A(x) = \begin{cases} 1 & \text{si } x \in A \\ 0 & \text{si } x \notin A \end{cases} \end{array} \right.$$

- Si une suite réelle $(u_n)_{n \in \mathbb{N}}$ est croissante, on peut noter $\lim_{n \rightarrow \infty} u_n$ sa limite (dans $\overline{\mathbb{R}}$).
- Si une suite réelle $(u_n)_{n \in \mathbb{N}}$ est décroissante, on peut noter $\lim_{n \rightarrow \infty} u_n$ sa limite (dans $\overline{\mathbb{R}}$).

2 Dénombrabilité

Définition

Un ensemble E est **dénombrable** si $E = \emptyset$ ou s'il existe une application $\varphi : \mathbb{N} \rightarrow E$ surjective, c'est-à-dire que

$$E = \varphi(\mathbb{N}) = \{\varphi(n) \mid n \in \mathbb{N}\} = \{\varphi(0), \varphi(1), \varphi(2), \dots\}.$$

Autrement dit, E est dénombrable si on peut énumérer ses éléments, en faire une liste (éventuellement infinie).

Exemples.

- les ensembles finis sont dénombrables (si $E = \{x_1, \dots, x_n\}$, on a $E = \varphi(\mathbb{N})$ pour $\varphi : \mathbb{N} \rightarrow E$ définie par $\varphi(i) = x_i$ si $1 \leq i \leq n$ et $\varphi(i) = x_n$ pour tout $i > n$)
- $\mathbb{N}$ est évidemment dénombrable : $\mathbb{N} = \{0, 1, 2, 3, \dots\}$
- $\mathbb{Z}$ est dénombrable : par exemple,

$$\mathbb{Z} = \{0, 1, -1, 2, -2, 3, -3, \dots\},$$

où on alterne entiers positifs et négatifs par ordre croissant de valeur absolue.

- $\mathbb{N} \times \mathbb{N} = \{(m, n) \mid m \in \mathbb{N}, n \in \mathbb{N}\}$ est dénombrable : par exemple,

$$\mathbb{N} \times \mathbb{N} = \{(0,0), (1,0), (0,1), (2,0), (1,1), (0,2), \dots\},$$

où on énumère les couples dont la somme vaut 0, puis 1, puis 2, etc. (à chaque fois, il y en a un nombre fini).

- $\mathbb{Q}$ est dénombrable : par exemple, (quitte à répéter certains rationnels plusieurs fois)

$$\mathbb{Q} = \left\{0, \frac{1}{1}, -\frac{1}{1}, \frac{1}{2}, -\frac{1}{2}, \frac{2}{1}, -\frac{2}{1}, \frac{3}{1}, -\frac{3}{1}, \frac{1}{3}, -\frac{1}{3}, \frac{4}{1}, -\frac{4}{1}, \frac{3}{2}, -\frac{3}{2}, \frac{2}{3}, -\frac{2}{3}, \frac{1}{4}, -\frac{1}{4}, \dots\right\}$$

où on énumère les fractions dont la somme des valeurs absolues du numérateur et du dénominateur vaut 0, puis 1, puis 2, etc. (il y en a un nombre fini pour une somme donnée), en alternant positifs et négatifs.

- $\mathbb{R}$ n'est pas dénombrable : on peut le démontrer à l'aide de *l'argument de la diagonale* de Cantor. Il suffit de montrer que $[0,1]$ n'est pas dénombrable (ceci est justifié par la propriété a) suivante). Supposons, par l'absurde, que ce soit le cas : on aurait $[0,1] = \varphi(\mathbb{N}) = \{\varphi(0), \varphi(1), \dots\}$ pour une fonction $\varphi : \mathbb{N} \rightarrow [0,1]$. Mais alors on peut facilement donner un réel $x \in [0,1]$ différent de $\varphi(n)$ pour tout $n \in \mathbb{N}$: il suffit que pour tout n , la n -ième décimale de x soit choisie différente de la n -ième décimale de $\varphi(n)$, et différente de 0 et 9 (cette précision évite que x soit un nombre décimal et n'ait pas une écriture unique : $0,4999\dots = 0,5000\dots$). Ainsi pour tout n , x est différent de $\varphi(n)$ (comme leurs n -ièmes décimales diffèrent, $x \neq \varphi(n)$ si le développement de $\varphi(n)$ est unique, et $x \neq \varphi(n)$ aussi sinon car alors $\varphi(n)$ est décimal or x ne l'est pas). On obtient donc une contradiction avec le fait que la suite précédente énumère tous les éléments de $[0,1]$, d'où il résulte que $[0,1]$ n'est pas dénombrable.

Propriétés

- a) Si $E \subset F$ et F est dénombrable, alors E est dénombrable aussi.
- b) Si E et F sont dénombrables, alors $E \times F$ sont dénombrables.
- c) Si, pour tout n , E_n est dénombrable, alors $\bigcup_n E_n$ est dénombrable.

Démonstration : a) Si E ou F est vide, c'est vrai. Sinon, on choisit $x_0 \in E$, et φ telle que $F = \varphi(\mathbb{N})$. Alors $E = \psi(\mathbb{N})$ où $\psi(n) = \varphi(n)$ si $\varphi(n) \in E$ et $\psi(n) = x_0$ si $\varphi(n) \notin E$, ce qui montre que E est dénombrable.

b) On a vu que $\mathbb{N} \times \mathbb{N}$ est dénombrable : il existe $\varphi : \mathbb{N} \rightarrow \mathbb{N} \times \mathbb{N}$ telle que $\mathbb{N} \times \mathbb{N} = \varphi(\mathbb{N})$. Si E ou F est vide, $E \times F$ aussi, donc est dénombrable. Sinon, il existe φ_E et φ_F telles que $E = \varphi_E(\mathbb{N})$ et $F = \varphi_F(\mathbb{N})$. On a alors $E \times F = \psi(\mathbb{N})$ où

$$\psi : k \mapsto (\varphi_E(\varphi_1(k)), \varphi_F(\varphi_2(k)))$$

en notant $\varphi(k) = (\varphi_1(k), \varphi_2(k)) \in \mathbb{N} \times \mathbb{N}$. En effet, pour tous $x \in E$ et $y \in F$, il existe m, n tels que $x = \varphi_E(m)$ et $y = \varphi_F(n)$, et il existe k tel que $\varphi(k) = (m, n)$, d'où $(x, y) = (\varphi_E(m), \varphi_F(n)) = (\varphi_E(\varphi_1(k)), \varphi_F(\varphi_2(k))) = \psi(k)$.

c) On utilise à nouveau la fonction φ ci-dessus, telle que $\mathbb{N} \times \mathbb{N} = \varphi(\mathbb{N})$. On peut supposer que $E_n \neq \emptyset$ pour tout n car les ensembles vides ne modifient pas la réunion. Pour tout $n \in \mathbb{N}$, il existe alors $\psi_n : \mathbb{N} \rightarrow E_n$ telle que $E_n = \psi_n(\mathbb{N})$. Et on a $\bigcup_n E_n = \xi(\mathbb{N})$, où

$$\xi : k \mapsto \psi_{\varphi_1(k)}(\varphi_2(k)).$$

En effet, pour tout $x \in \bigcup_n E_n$, il existe n tel que $x \in E_n$, donc il existe m tel que $x = \psi_n(m)$, et il existe k tel que $(n, m) = \varphi(k)$, d'où $x = \xi(k)$. ■

1. ESPACE MESURÉS

On définit ici les éléments qui nous serviront de cadre pour la théorie de l'intégration.

1 Tribus

Définition

Soit E un ensemble. Une **tribu** (ou σ -algèbre) sur E est un ensemble $\mathcal{A}$ de parties de E telle que

- (i) $\emptyset \in \mathcal{A}$;
 - (ii) si $A \in \mathcal{A}$, alors $A^c \in \mathcal{A}$ (**stabilité par passage au complémentaire**)
 - (iii) si $(A_n)_{n \in \mathbb{N}}$ est une suite de parties dans $\mathcal{A}$, alors $\bigcup_{n \in \mathbb{N}} A_n \in \mathcal{A}$; (**stabilité par union dénombrable**)
- $(E, \mathcal{A})$ est un **espace mesurable**. Une partie $A \in \mathcal{A}$ est dite **mesurable**.

Les conséquences suivantes sont aussi importantes que la définition :

Propriétés

- a) $E \in \mathcal{A}$;
- b) si $A_1, \dots, A_n \in \mathcal{A}$, alors $A_1 \cup \dots \cup A_n \in \mathcal{A}$;
- c) si $(A_n)_{n \in \mathbb{N}}$ est une suite de parties dans $\mathcal{A}$, $\bigcap_{n \in \mathbb{N}} A_n \in \mathcal{A}$; (**stabilité par intersection dénombrable**)
- d) si $A_1, \dots, A_n \in \mathcal{A}$, alors $A_1 \cap \dots \cap A_n \in \mathcal{A}$;
- e) si $A, B \in \mathcal{A}$ et $A \subset B$, alors $B \setminus A \in \mathcal{A}$.

Démonstration : a) résulte de (i) et (ii) car $E = \emptyset^c$.

b) résulte de (i) et (iii) en prenant $A_k = \emptyset$ pour tout $k > n$.

c) résulte de (i) et (iii) car

$$\bigcup_n A_n = \left(\bigcap_n (A_n)^c \right)^c.$$

d) résulte de c) en prenant $A_k = E$ pour tout $k > n$.

e) résulte de (ii) et d) car $B \setminus A = B \cap A^c$. ■

Attention. Si $(A_i)_{i \in I}$ est une famille de parties mesurables, alors les ensembles $\bigcup_{i \in I} A_i$ et $\bigcap_{i \in I} A_i$ sont mesurables à condition que I est dénombrable (car on peut écrire $I = \{i_n | n \in \mathbb{N}\}$ et donc $\bigcup_i A_i = \bigcup_n A_{i_n}$); mais si I n'est pas dénombrable, alors ce n'est pas toujours vrai.

Exemples.

- $\mathcal{P}(E)$ est la **tribu discrète** sur E .
- $\{\emptyset, E\}$ est la **tribu grossière** sur E .

Définition-proposition

Soit $\mathcal{C}$ un ensemble de parties de E . Il existe une plus petite tribu qui contient $\mathcal{C}$. On la note $\sigma(\mathcal{C})$, et on l'appelle la **tribu engendrée par $\mathcal{C}$** .

Démonstration : Par la définition de tribu, on voit facilement que l'intersection d'une famille de tribus est une tribu (le vérifier en exercice). En particulier, l'intersection de toutes les tribus sur E contenant $\mathcal{C}$ est une tribu, et c'est la plus petite : par sa définition, elle est incluse dans toute tribu contenant $\mathcal{C}$. ■

– sur $\mathbb{R}^d$, la **tribu borélienne** est la tribu engendrée par les ouverts. On la note $\mathcal{B}(\mathbb{R}^d)$. On peut vérifier que $\mathcal{B}(\mathbb{R}^d)$ est aussi la tribu engendrée par les intervalles de $\mathbb{R}^d$. Ses éléments sont les **boréliens**. Ainsi, tout ensemble construit à partir d'intervalles à l'aide des opérations de passage au complémentaire, d'union dénombrable et d'intersection dénombrable, est un borélien. En pratique, tous les sous-ensembles de $\mathbb{R}$ que l'on manipule sont obtenus ainsi et sont donc boréliens.

2 Mesures

Soit $(E, \mathcal{A})$ un espace mesurable.

Définition

Une **mesure** sur $(E, \mathcal{A})$ est une application $\mu : \mathcal{A} \rightarrow [0, +\infty]$ telle que

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

(ii) pour toute suite $(A_n)_{n \in \mathbb{N}}$ de parties mesurables **disjointes**, $\mu\left(\bigsqcup_{n \in \mathbb{N}} A_n\right) = \sum_{n \in \mathbb{N}} \mu(A_n)$.

$(E, \mathcal{A}, \mu)$ est un **espace mesuré**. $\mu(E)$ est la **masse totale** de μ . On dit que μ est **finie** si $\mu(E) < \infty$.

Les conséquences suivantes sont aussi importantes que la définition :

Propriétés

a) Si $A_1, \dots, A_n \in \mathcal{A}$ sont **disjointes**, alors $\mu(A_1 \sqcup \dots \sqcup A_n) = \mu(A_1) + \dots + \mu(A_n)$.

b) Si $A, B \in \mathcal{A}$ et $A \subset B$, alors $\mu(A) \leq \mu(B)$ et, si $\mu(A) < \infty$, alors $\mu(B \setminus A) = \mu(B) - \mu(A)$.

c) Pour tous $A, B \in \mathcal{A}$, et $\mu(A \cap B) < \infty$, alors $\mu(A \cup B) = \mu(A) + \mu(B) - \mu(A \cap B)$.

d) Si $(A_n)_n$ est une suite croissante de parties mesurables, alors $\mu\left(\bigcup_n A_n\right) = \lim_n \mu(A_n)$.

e) Si $(A_n)_n$ est une suite décroissante de parties mesurables, et $\mu(A_0) < \infty$, alors $\mu\left(\bigcap_n A_n\right) = \lim_n \mu(A_n)$.

f) Pour toute suite $(A_n)_n$ de parties mesurables, $\mu\left(\bigcup_n A_n\right) \leq \sum_n \mu(A_n)$.

Démonstration : a) résulte de (ii) et (i) en prenant $A_k = \emptyset$ si $k > n$

b) On a la réunion disjointe $B = A \sqcup (B \setminus A)$ donc par a) $\mu(B) = \mu(A) + \mu(B \setminus A) \leq \mu(A) + \mu(B)$ car μ est à valeurs positives. Si $\mu(A) < \infty$, on peut retrancher cette quantité à chaque membre de l'égalité précédente pour obtenir la formule.

c) On a la réunion disjointe $A \cup B = A \sqcup (B \setminus (A \cap B))$ donc $\mu(A \cup B) = \mu(A) + \mu(B \setminus (A \cap B))$ et si $\mu(A \cap B) < \infty$, la formule se déduit alors de b).

d) Pour tout n , $A_n \subset A_{n+1}$ donc par b) $\mu(A_n) \leq \mu(A_{n+1})$: la suite $(\mu(A_n))_n$ est croissante. On pose $C_0 = A_0$ et, pour tout $n \geq 1$, $C_n = A_n \setminus A_{n-1} \in \mathcal{A}$ de sorte que, pour tout n , $A_n = C_0 \sqcup \dots \sqcup C_n$, et donc $\bigcup_n A_n = \bigsqcup_n C_n$. Alors

$$\mu\left(\bigcup_n A_n\right) = \mu\left(\bigsqcup_n C_n\right) = \sum_n \mu(C_n) = \lim_N \sum_{n=0}^N \mu(C_n) = \lim_N \mu\left(\bigsqcup_{n=0}^N C_n\right) = \lim_N \mu(A_N).$$

e) En notant $B_n = A_0 \setminus A_n$, alors la suite $(B_n)_n$ est croissante et $\bigcup_n B_n = A_0 \setminus \bigcap_n A_n$, ce qui permet de déduire e) de d) à l'aide de b) vu que pour tout n , $A_n \subset A_0$ donc $\mu(A_n) \leq \mu(A_0) < \infty$.

f) On pose $C_0 = A_0$ et, pour tout $n \geq 1$, $C_n = A_n \cap (A_0 \cup \dots \cup A_{n-1})^c \in \mathcal{A}$, de sorte que les ensembles C_n sont disjoints et, pour tout n , $A_0 \cup \dots \cup A_n = C_0 \sqcup \dots \sqcup C_n$ donc $\bigcup_n A_n = \bigsqcup_n C_n$. Alors

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

mais $C_n \subset A_n$ donc $\mu(C_n) \leq \mu(A_n)$, ce qui donne l'inégalité attendue. ■

Exemples.

– Soit E un ensemble. Sur $(E, \mathcal{P}(E))$, la **mesure de comptage** μ_E est définie par :

$$\text{pour tout } A \subset E, \quad \mu_E(A) = \begin{cases} \text{Card}(A) & \text{si } A \text{ est fini} \\ \infty & \text{si } A \text{ est infini.} \end{cases}$$

Ainsi, « μ_E place un poids 1 en chaque point de E »

– Soit $(E, \mathcal{A})$ un espace mesurable, et $x \in E$. La **mesure de Dirac en x** est la mesure δ_x définie par :

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

Ainsi, « δ_x place un poids 1 au point x »

- Si $(\mu_n)_{n \geq 0}$ est une suite de mesures sur $(E, \mathcal{A})$ et $(\alpha_n)_{n \geq 0}$ une suite de réels positifs, alors on peut définir la mesure $\mu = \sum_{n \geq 0} \alpha_n \mu_n$ par

$$\text{pour tout } A \in \mathcal{A}, \quad \mu(A) = \sum_{n \geq 0} \alpha_n \mu_n(A).$$

En particulier, si $(x_n)_{n \geq 0}$ est une suite de points de E , on peut considérer $\mu = \sum_{n \geq 0} \alpha_n \delta_{x_n}$ qui, pour tout n , « place un poids α_n en x_n ».

Définition-théorème

Il existe une unique mesure λ_d sur $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ telle que, pour tout pavé fermé $[a_1, b_1] \times \cdots \times [a_d, b_d]$,

$$\lambda_d([a_1, b_1] \times \cdots \times [a_d, b_d]) = |b_1 - a_1| \cdots |b_d - a_d|.$$

On l'appelle **mesure de Lebesgue sur $\mathbb{R}^d$** .

Démonstration : Un peu longue : voir document sur mon site internet. ■

- sur $\mathbb{R}$, la mesure $\lambda = \lambda_1$ vérifie $\lambda([a, b]) = b - a$ pour tout segment $[a, b]$ avec $a \leq b$. Cette mesure correspond donc à la *longueur* sur $\mathbb{R}$. Le théorème signifie que l'on peut définir la longueur de n'importe quel borélien, et qu'elle vérifie la condition (ii).
- sur $\mathbb{R}^2$, la mesure λ_2 vérifie $\lambda_2([a, b] \times [c, d]) = (b - a)(d - c)$ pour tout rectangle $[a, b] \times [c, d]$ avec $a \leq b$ et $c \leq d$. Cette mesure correspond donc à l'*aire* sur $\mathbb{R}^2$.
- sur $\mathbb{R}^3$, la mesure λ_3 correspond de même au *volume*.

Propriétés

a) λ_d est **invariante par translation** : pour tout $A \in \mathcal{B}(\mathbb{R}^d)$ et $a \in \mathbb{R}^d$,

$$\lambda_d(a + A) = \lambda_d(A),$$

où $a + A = \{a + x \mid x \in A\}$.

b) λ_d est **homogène de degré d** : pour tout $A \in \mathcal{B}(\mathbb{R}^d)$ et $t \in \mathbb{R}$,

$$\lambda_d(tA) = |t|^d \lambda_d(A),$$

où $tA = \{tx \mid x \in A\}$.

Pour montrer que deux mesures sont égales, il suffit de comparer leurs valeurs sur les pavés :

Proposition

Soit μ, ν deux mesures sur $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$. Si, pour tout pavé fermé P , $\mu(P) = \nu(P) < \infty$, alors $\mu = \nu$.

Définition

Soit μ une mesure sur $(E, \mathcal{A})$.

- Si $A \in \mathcal{A}$ est tel que $\mu(A) = 0$, on dit que A est **négligeable**.

On peut préciser « μ -négligeable », ou « négligeable pour la mesure μ », si la mesure μ n'est pas claire d'après le contexte.

- Si une propriété P_x est vraie pour tout $x \in A$, où A^c est négligeable pour la mesure μ , on dit que P_x est vraie pour **presque tout** x , ou encore que P est vraie **presque partout**.

On peut préciser « μ -presque partout », ou « presque partout pour la mesure μ », si la mesure μ n'est pas claire d'après le contexte.

Sans précision, sur $\mathbb{R}^d$, « presque tout » fait référence à la mesure de Lebesgue λ_d .

3 Fonctions mesurables

Les démonstrations de cette partie se trouvent sur mon site internet.

Définition

Soit $(E, \mathcal{A})$ et $(F, \mathcal{B})$ des espaces mesurables. Une application $f : E \rightarrow F$ est **mesurable** si

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

Proposition

Les fonctions continues $f : \mathbb{R}^d \rightarrow \mathbb{R}^{d'}$ sont mesurables (pour les tribus boréliennes).

On considèrera souvent des fonctions f sur $(E, \mathcal{A})$ à valeurs dans $\overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty, +\infty\}$, c'est-à-dire que l'on pourra avoir $f(x) = +\infty$ ou $f(x) = -\infty$ pour certains $x \in E$. La tribu $\mathcal{B}(\overline{\mathbb{R}})$ est formée des boréliens de $\mathbb{R}$ et de tous les ensembles de la forme $B \cup \{+\infty\}$, $B \cup \{-\infty\}$ et $B \cup \{+\infty, -\infty\}$, pour $B \in \mathcal{B}(\mathbb{R})$.

Dire que f est mesurable signifie alors que, pour tout $B \in \mathcal{B}(\mathbb{R})$, $f^{-1}(B) \in \mathcal{A}$, et aussi que $f^{-1}(\{+\infty\}) \in \mathcal{A}$ et $f^{-1}(\{-\infty\}) \in \mathcal{A}$.

Propriétés

Si $f, g : E \rightarrow \mathbb{R}$ sont mesurables, et si $(f_n)_n$ est une suite de fonctions mesurables de E dans $\mathbb{R}$, alors

- a) la somme $f + g$ est mesurable
- b) le produit fg est mesurable
- c) les fonctions $\sup_n f_n$ et $\inf_n f_n$ (à valeurs dans $\overline{\mathbb{R}}$) sont mesurables
- d) la valeur absolue $|f|$ est mesurable
- e) les fonctions $\liminf_n f_n$ et $\limsup_n f_n$ (à valeurs dans $\overline{\mathbb{R}}$) sont mesurables
- f) si $f_n(x) \rightarrow h(x) \in \overline{\mathbb{R}}$ pour tout $x \in E$, alors h est mesurable.

Un rappel sur $\liminf$ et $\limsup$ se trouve dans le chapitre suivant, page 12.

Ainsi, toute fonction de $\mathbb{R}^d$ dans $\overline{\mathbb{R}}$ obtenue à partir de fonctions continues par ces opérations est mesurable. En pratique, toutes les fonctions que l'on manipule sont obtenues ainsi et sont donc mesurables pour les tribus boréliennes.

Il en va de même pour les fonctions de $\mathbb{R}^d$ dans $\mathbb{R}^{d'}$ par la proposition suivante :

Proposition

Une fonction $f : \mathbb{R}^d \rightarrow \mathbb{R}^{d'}$ est mesurable si, et seulement si ses composantes le sont.

2. INTÉGRATION PAR RAPPORT À UNE MESURE

Soit $(E, \mathcal{A}, \mu)$ un espace mesuré.

1 Intégrale de fonctions mesurables positives

Définition

Une **fonction étagée** sur $(E, \mathcal{A})$ est une fonction mesurable $g : (E, \mathcal{A}) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ qui ne prend qu'un nombre fini de valeurs. Autrement dit, il existe $\alpha_1, \dots, \alpha_n \in \mathbb{R}$ (les valeurs) et $A_1, \dots, A_n \in \mathcal{A}$ disjoints tels que

$$\text{pour tout } x \in \mathbb{R}, \quad g(x) = \sum_{i=1}^n \alpha_i \mathbf{1}_{A_i}(x) = \begin{cases} \alpha_1 & \text{si } x \in A_1 \\ \vdots & \\ \alpha_n & \text{si } x \in A_n \\ 0 & \text{sinon.} \end{cases}$$

NB. Les fonctions en escalier sur $\mathbb{R}$ sont étagées (c'est le cas où les A_i sont des intervalles), mais il y a beaucoup plus de fonctions étagées, par exemple $g = \mathbf{1}_{\mathbb{Q}}$. Ces fonctions forment un espace vectoriel.

Définition

Si g est étagée et positive (autrement dit, $\alpha_i \geq 0$ pour $i = 1, \dots, n$) alors, avec l'écriture de g ci-dessus, on définit

$$\int g d\mu = \sum_{i=1}^n \alpha_i \mu(A_i) \in [0, +\infty],$$

avec la convention que, si $\alpha_i = 0$ et $\mu(A_i) = \infty$, alors $\alpha_i \mu(A_i) = 0$.

On peut vérifier que cette définition ne dépend pas du choix de l'écriture de g sous la forme $g = \sum_i \alpha_i \mathbf{1}_{A_i}$, que les ensembles A_i soient disjoints ou non.

En particulier,

$$\boxed{\int \mathbf{1}_A d\mu = \mu(A)}.$$

Propriétés

Soit g, h des fonctions étagées positives.

a) Pour tous réels $a, b \geq 0$, $\int (ag + bh) d\mu = a \int g d\mu + b \int h d\mu$.

b) Si $g \leq h$, alors $\int g d\mu \leq \int h d\mu$.

Démonstration : Si $g = \sum_i \alpha_i \mathbf{1}_{A_i}$ et $h = \sum_j \beta_j \mathbf{1}_{B_j}$ avec $(A_i)_i$ disjoints, et $(B_j)_j$ disjoints, alors en introduisant les ensembles $A_{ij} = A_i \cap B_j$, la fonction $ag + bh$ prend la valeur $a\alpha_i + b\beta_j$ sur A_{ij} , ce qui permet d'obtenir a); et sur A_{ij} la fonction g vaut α_i et h vaut β_j donc, si $A_{ij} \neq \emptyset$, $\alpha_i \leq \beta_j$, ce qui permet d'obtenir b). ■

Définition

Soit $f : E \rightarrow [0, +\infty]$ mesurable. On note

$$\int f d\mu = \sup_{\substack{h \text{ étagée,} \\ 0 \leq h \leq f}} \int h d\mu \in [0, +\infty],$$

et cette quantité est appelée l'**intégrale de f par rapport à μ** .

NB. On utilise aussi les notations suivantes : $\int f d\mu = \int f(x) d\mu(x) = \int f(x) \mu(dx)$ et on peut spécifier $\int_E$. Dans la suite, de même que dans cette définition, une fonction « mesurable positive » est supposée prendre ses valeurs dans $[0, +\infty]$.

Notons que par la propriété b) ci-dessous, si $f = \infty \cdot \mathbf{1}_A$ (autrement dit, $f(x) = \infty$ si $x \in A$ et $f(x) = 0$ sinon) avec $\mu(A) = 0$ alors $\int f d\mu = “\infty \cdot 0” = 0$. Ainsi, pour $\alpha \in \mathbb{R}_+ \cup \{+\infty\}$, $\int \alpha \mathbf{1}_A d\mu = \alpha \mu(A)$ avec “ $0 \cdot \infty = \infty \cdot 0 = 0$ ”.

Propriétés

Soit f, g des fonctions mesurables positives.

a) Si $f \leq g$, alors $\int f d\mu \leq \int g d\mu$.

b) Si $f = 0$ presque partout (pour la mesure μ), alors $\int f d\mu = 0$.

Démonstration : Si $f \leq g$, alors toute fonction étagée h telle que $0 \leq h \leq f$ vérifie aussi $0 \leq h \leq g$, d'où a) (le sup qui définit l'intégrale de g porte sur un ensemble plus grand, donc est plus grand).

Si $f = 0$ presque partout, et $0 \leq h \leq f$ est étagée, on a $h = \sum_i \alpha_i \mathbf{1}_{A_i}$ (avec A_i disjoints) et, si $\alpha_i > 0$, on a, pour $x \in A_i$, $f(x) \geq h(x) = \alpha_i > 0$, donc $A_i \subset \{x \in \mathbb{R} | f(x) \neq 0\}$ d'où $\mu(A_i) \leq \mu(\{f \neq 0\}) = 0$ et donc $\mu(A_i) = 0$; on déduit que $\int h d\mu = \sum_i \alpha_i \mu(A_i) = 0$. D'où b). ■

Théorème (Théorème de convergence monotone (TCM))

Soit $(f_n)_n$ une suite **croissante** de fonctions mesurables **positives**. Alors

$$\int \lim_{\uparrow n} f_n d\mu = \lim_{\uparrow n} \int f_n d\mu.$$

Démonstration : On note $f = \lim_{\uparrow n} f_n$, qui est une fonction mesurable positive.

Pour tout n , $f_n \leq f_{n+1}$, d'où $\int f_n d\mu \leq \int f_{n+1} d\mu$. Ainsi, la suite $\left(\int f_n d\mu \right)_n$ est croissante et donc converge (dans $\overline{\mathbb{R}}$).

De même, pour tout n , $f_n \leq f$, d'où $\int f_n d\mu \leq \int f d\mu$, et en passant à la limite on a

$$\lim_{\uparrow n} \int f_n d\mu \leq \int f d\mu.$$

Il reste à voir l'inégalité inverse.

Soit h une fonction étagée telle que $0 \leq h \leq f$. Soit $\varepsilon > 0$.

Pour tout $x \in E$, si $h(x) > 0$, on a $\lim_{\uparrow n} f_n(x) = f(x) \geq h(x) > (1 - \varepsilon)h(x)$ donc il existe n (grand) tel que $f_n(x) \geq (1 - \varepsilon)h(x)$; et si $h(x) = 0$ alors $f_n(x) = 0$ pour tout n donc évidemment $f_n(x) \geq (1 - \varepsilon)h(x)$. Autrement dit,

$$E = \bigcup_n E_n, \quad \text{où} \quad E_n = \{x \in E | f_n(x) \geq (1 - \varepsilon)h(x)\} \in \mathcal{A}.$$

Par définition de E_n , on a pour tout $x \in E_n$, $f_n(x) \geq (1 - \varepsilon)h(x)$. Par conséquent, comme $f_n \geq 0$, on a $f_n \geq (1 - \varepsilon)h \mathbf{1}_{E_n}$. Comme h est étagée, elle s'écrit $h = \sum_{1 \leq i \leq r} \alpha_i \mathbf{1}_{A_i}$, si bien que $h \mathbf{1}_{E_n} = \sum_{1 \leq i \leq r} \alpha_i \mathbf{1}_{A_i \cap E_n}$ est étagée elle aussi et

$$\int f_n d\mu \geq \int (1 - \varepsilon)h \mathbf{1}_{E_n} d\mu = (1 - \varepsilon) \sum_{i=1}^r \alpha_i \mu(A_i \cap E_n).$$

Soit $1 \leq i \leq r$. La suite $(A_i \cap E_n)_n$ est croissante (car $(E_n)_n$ est croissante) et sa réunion est A_i (car la réunion des E_n est E) donc

$$\lim_{\uparrow n} \mu(A_i \cap E_n) = \mu(A_i).$$

En passant à la limite dans l'inégalité précédente, on a donc

$$\lim_{\uparrow n} \int f_n d\mu \geq (1 - \varepsilon) \sum_{i=1}^r \alpha_i \mu(A_i) = (1 - \varepsilon) \int h d\mu.$$

Pour $\varepsilon \rightarrow 0$, ceci donne

$$\lim_{\uparrow n} \int f_n d\mu \geq \int h d\mu.$$

Et ceci vaut pour toute fonction h étagée telle que $0 \leq h \leq f$, donc on a finalement

$$\lim_{\uparrow n} \int f_n d\mu \geq \int f d\mu,$$

ce qui conclut la preuve. ■

Par le TCM, pour calculer $\int f d\mu$, on peut considérer $\lim_{\uparrow n} \int f_n d\mu$ pour n'importe quelle suite croissante $(f_n)_n$ qui converge vers f . Par exemple une suite de fonctions étagées :

Lemme

Si f est mesurable positive, alors il existe une suite croissante $(f_n)_n$ de fonctions étagées positives qui converge vers f .

Démonstration : On peut prendre

$$f_n(x) = \begin{cases} \frac{k}{2^n} & \text{si } f(x) \in [\frac{k}{2^n}, \frac{k+1}{2^n}[\text{ et } f(x) < n \\ n & \text{si } f(x) \geq n. \end{cases}$$

La distinction permet de s'assurer que f_n ne prend qu'un nombre fini de valeurs même si f n'est pas bornée. Le choix de 2^n assure la croissance de f_n (la subdivision se raffine à chaque pas). ■

Propriétés

Pour f, g mesurables positives, et a, b réels positifs, $\int (af + bg) d\mu = a \int f d\mu + b \int g d\mu$.

Démonstration : Le lemme fournit des suites croissantes $(f_n)_n$ et $(g_n)_n$ de fonctions étagées positives qui convergent vers f et g . Pour tout n , par les propriétés de l'intégrale des fonctions étagées positives,

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

ce qui donne l'égalité annoncée en passant à la limite grâce au TCM. ■

Le théorème de convergence monotone admet une réécriture en termes de séries :

Corollaire (Théorème de convergence monotone pour les séries positives)

Si $(f_n)_{n \geq 0}$ est une suite de fonctions mesurables positives, alors

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

Démonstration : On applique le TCM à la suite des sommes partielles $S_n = \sum_{k=0}^n f_k \geq 0$, qui est croissante (car $f_n \geq 0$), et converge vers $S = \sum_{k=0}^{\infty} f_k$. On a ainsi

$$\int S d\mu = \lim_{\uparrow n} \int S_n d\mu.$$

Or, par la propriété précédente,

$$\int S_n d\mu = \int \left(\sum_{k=0}^n f_k \right) d\mu = \sum_{k=0}^n \int f_k d\mu \xrightarrow{n} \sum_{k=0}^{\infty} \int f_k d\mu$$

d'où la conclusion. ■

Proposition (Inégalité de Markov)

Pour toute fonction mesurable positive f , et tout réel $a > 0$,

$$\mu\left(\left\{x \in E \mid f(x) \geq a\right\}\right) \leq \frac{1}{a} \int f d\mu.$$

Démonstration : L'ensemble $A = \{x \in E \mid f(x) \geq a\}$ vérifie $f \geq a \mathbf{1}_A$ (pour $x \in A$, $f(x) \geq a = a \mathbf{1}_A(x)$, et pour $x \notin A$, $f(x) \geq 0 = a \mathbf{1}_A(x)$). Par suite, en intégrant,

$$\int f d\mu \geq \int a \mathbf{1}_A d\mu = a \mu(A),$$

ce qui donne l'inégalité. ■

Corollaire

Soit f, g des fonctions mesurables **positives**.

- a) Si $\int f d\mu < \infty$, alors $f < \infty$ presque partout.
- b) $\int f d\mu = 0$ si, et seulement si $f = 0$ presque partout.
- c) Si $f = g$ presque partout, alors $\int f d\mu = \int g d\mu$.

Démonstration : a) On suppose $\int f d\mu < \infty$. Pour tout n , on note $A_n = \{f \geq n\} (= \{x \in E \mid f(x) \geq n\})$, et

$$A_\infty = \{f = \infty\} = \bigcap_n A_n.$$

Pour tout n , on a $\mu(A_n) \leq \frac{1}{n} \int f d\mu$ par l'inégalité de Markov, donc $\mu(A_1) < \infty$, et $\mu(A_n) \rightarrow 0$. On a donc

$$\mu(A_\infty) = \mu\left(\bigcap_n A_n\right) = \lim_{n \rightarrow \infty} \mu(A_n) = 0.$$

b) On suppose $\int f d\mu = 0$. On a $\{f > 0\} = \bigcup_n \{f \geq \frac{1}{n}\}$ donc $\mu(\{f > 0\}) = \lim_{n \rightarrow \infty} \mu(\{f \geq \frac{1}{n}\}) = 0$ car l'inégalité de Markov donne $\mu(\{f \geq \frac{1}{n}\}) \leq n \int f d\mu = 0$ quel que soit n . La réciproque a déjà été vue.

c) On suppose $f = g$ presque partout. En particulier, la fonction $h = \max(f, g) - \min(f, g)$ est nulle presque partout et **positive** (ce qui n'est peut-être pas le cas de $f - g$), donc $\int h d\mu = 0$. Par la propriété, on a alors $\int \max(f, g) d\mu = \int \min(f, g) d\mu$. Or $\min(f, g) \leq f \leq \max(f, g)$ et de même pour g ; en passant aux intégrales, on voit que $\int f d\mu$ et $\int g d\mu$ sont encadrées par $\int \min(f, g) d\mu$ et $\int \max(f, g) d\mu$ qui sont égales, et donc en particulier $\int f d\mu = \int g d\mu$. ■

On rappelle que, pour une suite de réels $(x_n)_n$, on définit

$$\liminf_n x_n = \sup_n \inf_{k \geq n} x_k = \lim_{n \rightarrow \infty} \inf_{k \geq n} x_k$$

et

$$\limsup_n x_n = \inf_n \sup_{k \geq n} x_k = \lim_{n \rightarrow \infty} \sup_{k \geq n} x_k.$$

On a les propriétés suivantes : $\limsup_n (-x_n) = -\liminf_n x_n$, et la suite $(x_n)_n$ converge si, et seulement si $\liminf_n x_n = \limsup_n x_n$, et dans ce cas $\lim_n x_n = \liminf_n x_n = \limsup_n x_n$. De façon générale, $\liminf_n x_n$ est la plus petite limite d'une sous-suite de $(x_n)_n$ (et $\limsup_n x_n$ la plus grande. On dit aussi que ce sont la plus petite valeur d'adhérence de la suite et la plus grande.

Pour une suite $(f_n)_n$ de fonctions $E \rightarrow \mathbb{R}$, on peut alors définir les fonctions $\liminf_n f_n$ et $\limsup_n f_n$, par : pour $x \in E$,

$$(\liminf_n f_n)(x) = \liminf_n f_n(x) \quad \text{et} \quad (\limsup_n f_n)(x) = \limsup_n f_n(x).$$

Ces définitions ont aussi un sens si on autorise les valeurs $-\infty$ et $+\infty$ pour les suites et les fonctions.

Théorème (Lemme de Fatou)

Soit $(f_n)_{n \geq 0}$ une suite de fonction mesurables positives. On a

$$\int (\liminf_n f_n) d\mu \leq \liminf_n \int f_n d\mu.$$

Démonstration : On a, par TCM,

$$\int \liminf_n f_n d\mu = \int \lim_{n \rightarrow \infty} \inf_{k \geq n} f_k d\mu = \lim_{n \rightarrow \infty} \int \inf_{k \geq n} f_k d\mu.$$

Or, pour tout n , pour tout $m \geq n$, $\inf_{k \geq n} f_k \leq f_m$, donc $\int \inf_{k \geq n} f_k d\mu \leq \int f_m d\mu$, d'où $\int \inf_{k \geq n} f_k d\mu \leq \inf_{m \geq n} \int f_m d\mu$. En passant à la limite,

$$\lim_{n \rightarrow \infty} \int \inf_{k \geq n} f_k d\mu \leq \lim_{n \rightarrow \infty} \inf_{m \geq n} \int f_m d\mu = \liminf_n \int f_n d\mu,$$

ce qui donne le résultat vu la première égalité. ■

NB. Voici trois exemples de suites $(f_n)_n$ telles que $f_n \rightarrow 0$ et $\int f_n d\mu \not\rightarrow 0$ (on a même $\int_n f_n d\mu = 1$ pour tout n). On utilise $E = \mathbb{R}$, muni de la mesure de Lebesgue $\mu = \lambda_1$.

« **bosse voyageuse** » : $f_n = \mathbf{1}_{[n, n+1]}$

« **concentration en 0** » : $f_n = n \mathbf{1}_{[\frac{1}{n}, \frac{1}{2n}]}$

« **écrasement** » : $f_n = \frac{1}{n} \mathbf{1}_{[-\frac{n}{2}, +\frac{n}{2}]}$.

2 Fonctions intégrables

Définition

Soit $f : E \rightarrow \mathbb{R}$ une fonction mesurable. f est **intégrable par rapport à μ** si $\int |f| d\mu < \infty$.

On pose alors

$$\int f d\mu = \int f_+ d\mu - \int f_- d\mu \in \mathbb{R},$$

où $f_+ = \max(0, f)$ et $f_- = \max(0, -f)$ sont les parties positive et négative de f .

On note $\mathcal{L}^1(E, \mathcal{A}, \mu)$ l'espace des fonctions intégrables par rapport à μ .

NB. On a $|f| = f_+ + f_- \geq f_-$, ce qui justifie que $\int f_- d\mu < \infty$ et donne un sens à la soustraction ci-dessus. De même, $\int f_+ d\mu < \infty$ donc $\int f d\mu$ est bien réel.

On abrège souvent $\mathcal{L}^1(E, \mu)$, voire $\mathcal{L}^1(E)$ ou même $\mathcal{L}^1$ si le contexte précise $(E, \mathcal{A}, \mu)$.

Propriétés

- a) Pour toute $f \in \mathcal{L}^1(E, \mathcal{A}, \mu)$, $\left| \int f d\mu \right| \leq \int |f| d\mu$
- b) $\mathcal{L}^1(E, \mathcal{A}, \mu)$ est un espace vectoriel, et $f \mapsto \int f d\mu$ est une application linéaire de $\mathcal{L}^1(E, \mathcal{A}, \mu)$ dans $\mathbb{R}$.
- c) Pour $f, g \in \mathcal{L}^1(E, \mathcal{A}, \mu)$, si $f \leq g$, alors $\int f d\mu \leq \int g d\mu$.
- d) Pour $f, g \in \mathcal{L}^1(E, \mathcal{A}, \mu)$, si $f = g$ presque partout, alors $\int f d\mu = \int g d\mu$.

Démonstration : a) On a $f = f_+ - f_-$ et $|f| = f_+ + f_-$, donc par inégalité triangulaire,

$$\left| \int f d\mu \right| = \left| \int f_+ d\mu - \int f_- d\mu \right| \leq \left| \int f_+ d\mu \right| + \left| \int f_- d\mu \right| = \int f_+ d\mu + \int f_- d\mu = \int (f_+ + f_-) d\mu = \int |f| d\mu.$$

b) Décomposer en parties positives et négatives... Exercice.

c) $g = f + (g - f)$ et $g - f \geq 0$ donc $\int (g - f) d\mu \geq 0$ et $\int g d\mu = \int f d\mu + \int (g - f) d\mu \geq \int f d\mu$.

d) si $f = g$ presque partout, alors $f_+ = g_+$ presque partout et $f_- = g_-$ presque partout (là où f et g sont égales, leurs parties positives et négatives aussi), donc $\int f_+ d\mu = \int g_+ d\mu$, de même pour f_- et g_- d'où $\int f d\mu = \int g d\mu$. ■

Théorème (Théorème de convergence dominée (TCD))

Soit $(f_n)_n$ une suite de fonctions mesurables $E \rightarrow \mathbb{R}$, et f une fonction mesurable $E \rightarrow \mathbb{R}$. On suppose

- (i) $f_n(x) \rightarrow f(x)$ pour presque partout $x \in E$;
- (ii) il existe $\varphi : E \rightarrow \mathbb{R}_+$ mesurable telle que $\int \varphi d\mu < \infty$ et

$$\text{pour tout } n, \text{ pour presque tout } x \in E, \quad |f_n(x)| \leq \varphi(x). \quad (\text{hypothèse de domination})$$

Alors, pour tout n , $f_n \in \mathcal{L}^1(E, \mathcal{A}, \mu)$, $f \in \mathcal{L}^1(E, \mathcal{A}, \mu)$,

$$\int f_n d\mu \xrightarrow{n} \int f d\mu \quad \text{et} \quad \int |f_n - f| d\mu \xrightarrow{n} 0.$$

Démonstration : Le fait que f_n et f soient intégrables vient des inégalités $|f_n| \leq \varphi$ et, à la limite, $|f| \leq \varphi$. Elles sont vraie presque partout, dont $\int |f_n| d\mu \leq \int \varphi d\mu < \infty$ et de même pour f .

Pour simplifier, supposons (i) et (ii) vrais partout. Les fonctions $\varphi - f_n$ et $\varphi + f_n$ sont positives. On peut leur appliquer le lemme de Fatou. Pour $\varphi - f_n$:

$$\int \liminf_n (\varphi - f_n) d\mu \leq \liminf_n \int (\varphi - f_n) d\mu \quad (*)$$

Or $\liminf_n (\varphi - f_n) = \varphi - f$ donc

$$\int \liminf_n (\varphi - f_n) d\mu = \int \varphi d\mu - \int f d\mu.$$

Et

$$\liminf_n \int (\varphi - f_n) d\mu = \int \varphi d\mu - \limsup_n \int f_n d\mu.$$

Ainsi, (*) donne

$$\limsup_n \int f_n d\mu \leq \int f d\mu.$$

Les mêmes arguments pour $\varphi + f_n$ donnent

$$\liminf_n \int f_n d\mu \geq \inf f,$$

d'où finalement

$$\int f d\mu \leq \liminf_n \int f_n d\mu \leq \limsup_n \int f_n d\mu \leq \int f d\mu,$$

ce qui montre que tous les termes sont égaux, et en particulier $\int f_n d\mu$ converge vers $\int f d\mu$. ■

On peut alors donner une formule « concrète » de calcul de $\int f d\mu$ par approximation, qui correspond à la définition évoquée lors de la présentation du cours :

Corollaire

Soit f une fonction intégrable positive. Pour toute suite de subdivisions $0 = \ell_0^{(n)} < \ell_1^{(n)} < \dots < \ell_{N(n)}^{(n)}$ de $\mathbb{R}$ telle que

$$\max_{0 \leq i < N(n)} \ell_{i+1}^{(n)} - \ell_i^{(n)} \xrightarrow[n]{} 0 \quad \text{et} \quad \ell_{N(n)}^{(n)} \rightarrow +\infty,$$

on a

$$\int f d\mu = \lim_n \sum_{i=1}^{N(n)} \ell_i^{(n)} \mu(f^{-1}([\ell_i^{(n)}, \ell_{i+1}^{(n)}])).$$

Démonstration : Il s'agit de montrer que $\int f d\mu = \lim_n \int f_n d\mu$, où on a défini les fonctions étagées

$$f_n : x \mapsto \sum_{i=1}^{N(n)} \ell_i^{(n)} \mathbf{1}_{f^{-1}([\ell_i^{(n)}, \ell_{i+1}^{(n)}]))}(x) = \begin{cases} \ell_i^{(n)} & \text{si } \ell_i^{(n)} \leq f(x) < \ell_{i+1}^{(n)}, \text{ où } 0 \leq i < N(n) \\ 0 & \text{si } f(x) \geq \ell_{N(n)}^{(n)}. \end{cases}$$

Pour tout $x \in E$ on a, pour tout n assez grand, $f(x) < \ell_{N(n)}^{(n)}$ (par la deuxième condition sur la subdivision), et il existe alors i tel que $f_n(x) = \ell_i^{(n)} \leq f(x) < \ell_{i+1}^{(n)}$ d'où

$$0 \leq f(x) - f_n(x) < \ell_{i+1}^{(n)} - \ell_i^{(n)} \leq \max_j \ell_{j+1}^{(n)} - \ell_j^{(n)} \xrightarrow[n]{} 0$$

(par la première condition). Ceci montre que $(f_n)_n$ converge vers f . Il reste à prouver une domination.

Or, si $f(x) < \ell_{N(n)}^{(n)}$, on a aussi vu que $f_n(x) \leq f(x)$. Et si $f(x) \geq \ell_{N(n)}^{(n)}$, alors $f_n(x) = 0 \leq f(x)$. Donc on a, pour tout n , $0 \leq f_n(x) \leq f(x)$. Or on a supposé f intégrable. On peut donc appliquer le TCD à la suite $(f_n)_n$, dominée par $\varphi = f$, ce qui conclut. ■

Notation. Pour $A \in \mathcal{A}$, on note $\int_A f d\mu = \int f \mathbf{1}_A d\mu$ l'intégrale de f sur A par rapport à μ , lorsqu'elle a un sens, c'est-à-dire si $f \mathbf{1}_A$ est positive ou intégrable. Ceci a d'ailleurs un sens même si f n'est pas définie hors de A (car $\mathbf{1}_A$ vaut alors 0). On dit que f est **intégrable sur** A si $\int_A |f| d\mu < \infty$.

Remarque importante. Toute cette partie s'étend aux fonctions à valeurs dans $\mathbb{C}$ et $\mathbb{R}^d$ en intégrant composante par composante : par exemple, une fonction mesurable $f : E \rightarrow \mathbb{C}$ est intégrable si $\int |f| d\mu < \infty$ et, dans ce cas,

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

Les résultats précédents restent alors vrais avec cette définition (linéarité, TCD).

3 Exemples principaux

3.1 Intégrale par rapport à une mesure atomique

(On dit que x est un **atome** de μ si $\{x\} \in \mathcal{A}$ et $\mu(\{x\}) > 0$)

Proposition

Soit $f : E \rightarrow \mathbb{R}$ une fonction.

a) Soit $x \in E$. f est intégrable par rapport à δ_x et $\int f d\delta_x = f(x)$.

b) Soit $(x_n)_{n \geq 0}$ une suite d'éléments de E et $(\alpha_n)_{n \geq 0}$ une suite de réels ≥ 0 . On pose $\mu = \sum_{n \geq 0} \alpha_n \delta_{x_n}$. Si f est positive, on a

$$\int f d\mu = \sum_{n \geq 0} \alpha_n f(x_n) \in [0, +\infty].$$

Pour f de signe quelconque, f est intégrable par rapport à μ si, et seulement si $\sum_n \alpha_n |f(x_n)| < \infty$ et, dans ce cas,

$$\int f d\mu = \sum_{n \geq 0} \alpha_n f(x_n) \in \mathbb{R}.$$

Démonstration : a) On a $f = f(x)\mathbf{1}_{\{x\}}$ presque partout, car ces fonctions coïncident en x , et $\delta_x(\{x\}^c) = 0$. Donc leurs intégrales sont égales :

$$\int f d\delta_x = \int f(x)\mathbf{1}_{\{x\}} d\delta_x = f(x)\delta_x(\{x\}) = f(x)$$

(par la définition de l'intégrale d'une fonction étagée).

b) Soit f mesurable positive. Si $g = f\mathbf{1}_F$ où $F = \{x_n \mid n \in \mathbb{N}\}$, alors $f = g$ presque partout pour la mesure μ . En effet, $f = g$ sur F , et $\mu(F^c) = 0$. On a donc $\int f d\mu = \int g d\mu$. Or on a $g = \lim_{n \uparrow} g_n$ où $g_n = \sum_{k=0}^n f(x_k)\mathbf{1}_{\{x_k\}}$ donc, par TCM,

$$\int g d\mu = \lim_{n \uparrow} \int g_n d\mu = \lim_{n \uparrow} \sum_{k=0}^n f(x_k)\mu(\{x_k\}) = \sum_{k=0}^{\infty} f(x_k)\alpha_k,$$

d'où l'égalité annoncée. Si f est de signe quelconque et intégrable, on applique ce qui précède à f_+ et f_- pour obtenir le résultat. ■

Ainsi, si μ_E est la mesure de comptage sur E et $f : E \rightarrow \mathbb{R}_+$,

$$\int f d\mu_E = \sum_{x \in E} f(x).$$

3.2 Intégrale par rapport à la mesure de Lebesgue (lien avec l'intégrale de Riemann)

On note $\lambda = \lambda_1$ (mesure de Lebesgue sur $\mathbb{R}$). Soit $a < b$.

Rappel. Une fonction $f : [a, b] \rightarrow \mathbb{R}$ est **intégrable au sens de Riemann** si, pour tout $\varepsilon > 0$, il existe des fonctions en escalier φ et ψ sur $[a, b]$ telles que

$$\varphi \leq f \leq \psi \quad \text{et} \quad \int_a^b (\psi - \varphi) < \varepsilon,$$

où l'intégrale de la fonction en escalier $\psi - \varphi$ est définie élémentairement (comme pour les fonctions étagées : si $\varphi = \sum_i \alpha_i \mathbf{1}_{[x_i, x_{i+1}[}$ alors $\int_a^b \varphi = \sum_i \alpha_i (x_{i+1} - x_i)$). Et dans ce cas on note

$$\int_a^b f = \sup_{\substack{\varphi \text{ en escalier,} \\ \varphi \leq f}} \int_a^b \varphi = \inf_{\substack{\psi \text{ en escalier,} \\ f \leq \psi}} \int_a^b \psi$$

Théorème

Si f est intégrable au sens de Riemann sur $[a, b]$, alors f est intégrable par rapport à λ sur $[a, b]$, et

$$\int_{[a, b]} f d\lambda = \int_a^b f.$$

Démonstration : Comme $\lambda([x_i, x_{i+1}[) = x_{i+1} - x_i$ (et $\lambda(\{x_i\}) = 0$), la formule est vraie pour les fonctions en escalier (en comparant à la formule pour l'intégrale des fonctions étagées).

Soit f une fonction intégrable au sens de Riemann. f est limite de fonctions en escalier, donc est mesurable. Soit $\varepsilon > 0$. Il existe φ, ψ en escalier telles que $\varphi \leq f \leq \psi$ et $\int_a^b (\psi - \varphi) < \varepsilon$ (où l'intégrale est au sens de Riemann ou de Lebesgue, par ce qui précède). En particulier f est bornée donc intégrable sur $[a, b]$: on a $|f| \leq M$ (où $M = \|\varphi\|_\infty + \|\psi\|_\infty$) donc $\int_{[a,b]} |f| d\mu \leq \int_{[a,b]} M d\mu = (b-a)M < \infty$. On a (par les propriétés de l'intégrale de Riemann)

$$\int_a^b \varphi \leq \int_a^b f \leq \int_a^b \psi$$

et aussi (par les propriétés de l'intégrale de Lebesgue)

$$\int_{[a,b]} \varphi d\lambda \leq \int_{[a,b]} f d\lambda \leq \int_{[a,b]} \psi d\lambda.$$

Or $\int_a^b \varphi = \int_{[a,b]} \varphi d\lambda$ car φ est en escalier, et de même pour ψ . Ainsi les deux intégrales de f appartiennent au même intervalle de largeur $< \varepsilon$, donc

$$\left| \int_a^b f - \int_{[a,b]} f d\lambda \right| < \varepsilon.$$

Ceci vaut pour tout $\varepsilon > 0$, d'où la conclusion. ■

Par suite, si I est un intervalle, pour $f : I \rightarrow \mathbb{R}$ mesurable positive, ou intégrable par rapport à λ , on pourra noter

$$\int_I f = \int_I f(x) dx = \int_I f d\lambda,$$

même si f n'est pas intégrable au sens de Riemann, sans confusion possible.

On pourra donc, pour des intégrales au sens de Riemann, appliquer les théorèmes précédents (convergence monotone, dominée, etc.) ; et pour des intégrales de Lebesgue de fonctions intégrables au sens de Riemann, utiliser les propriétés bien connues (intégration par parties, lien entre intégrale et primitive, etc.).

3.3 Intégrale par rapport à une mesure à densité

Soit $(E, \mathcal{A})$ un espace mesurable.

On rappelle que, si $f : E \rightarrow \mathbb{R}$ est mesurable, et $A \in \mathcal{A}$, on note $\int_A f d\mu = \int f \mathbf{1}_A d\mu$ l'intégrale de f sur A , lorsque $f \mathbf{1}_A$ est à valeurs positives, ou lorsque $f \mathbf{1}_A$ est intégrable.

On vérifie facilement avec le TCM que, si f est positive, $A \mapsto \int_A f d\mu$ est une mesure, d'où la définition :

Définition

Si f est une fonction mesurable $E \rightarrow [0, +\infty]$, et μ une mesure sur E , la **mesure de densité f par rapport à μ** est la mesure $f \cdot \mu$ (aussi notée $f(x)d\mu(x)$) définie par :

$$\text{pour tout } A \in \mathcal{A}, \quad (f \cdot \mu)(A) = \int_A f d\mu = \int f \mathbf{1}_A d\mu.$$

NB. A est négligeable pour $f \cdot \mu$ dès que A est négligeable pour μ ou que f est nulle sur A .

Proposition

Soit f une fonction mesurable $E \rightarrow [0, +\infty]$, et μ une mesure sur E .

a) Pour toute fonction mesurable $g : E \rightarrow [0, +\infty]$, on a

$$\int g d(f \cdot \mu) = \int g f d\mu = \int g(x) f(x) d\mu(x),$$

b) Une fonction $g : E \rightarrow \mathbb{R}$ est intégrable par rapport à $f \cdot \mu$ si, et seulement si fg est intégrable par rapport à μ et, dans ce cas,

$$\int g d(f \cdot \mu) = \int g f d\mu = \int g(x) f(x) d\mu(x).$$

Ceci justifie la notation $f \cdot \mu = f(x)d\mu(x)$. Pour la mesure de Lebesgue, vu le lien avec l'intégrale de Riemann, on notera aussi $f(x)dx$ pour $f \cdot \lambda$. Par extension, vu que $1 \cdot \mu = \mu$, on pourra parfois noter $d\mu(x)$ pour désigner la mesure μ , et donc dx pour désigner la mesure de Lebesgue λ (ou λ_d).

4 Intégrales dépendant d'un paramètre

Dans cette partie, on considère une famille $(f(t, \cdot))_{t \in I}$ de fonctions, indexée par un intervalle réel I : plutôt qu'une suite, il s'agit d'un « continuum » de fonctions, que l'on peut aussi voir comme une fonction $f : I \times E \rightarrow \mathbb{R}$. Lorsque les fonctions $f(t, \cdot)$ sont intégrables, on peut considérer la fonction

$$F : t \mapsto F(t) = \int f(t, \cdot) d\mu = \int f(t, x) d\mu(x),$$

que l'on appelle une **intégrale à paramètre** car elle dépend du réel t , appelé un paramètre. Le théorème de convergence dominée permet alors d'obtenir facilement des résultats de régularité sur F .

Théorème (Théorème de continuité sous l'intégrale)

Soit $f : (t, x) \mapsto f(t, x)$ une fonction mesurable de $I \times E$ dans $\mathbb{R}^d$ ou $\mathbb{C}$ (où I est un intervalle de $\mathbb{R}$).

On suppose que :

- **(continuité par rapport au paramètre)** pour μ -presque tout $x \in E$, $t \mapsto f(t, x)$ est continue sur I ;
- **(domination)** il existe une fonction $\varphi : E \rightarrow \mathbb{R}_+$ mesurable telle que $\int \varphi d\mu < \infty$ et, pour tout $t \in I$, pour μ -presque tout $x \in E$,

$$|f(t, x)| \leq \varphi(x).$$

Alors la fonction

$$F : t \mapsto F(t) = \int f(t, x) d\mu(x)$$

est bien définie pour tout $t \in I$, et est continue sur I .

Démonstration : Soit $t \in I$. On montre que F est continue en t . On utilise la caractérisation des limites par les suites : pour montrer que $F(u) \rightarrow F(t)$ quand $u \rightarrow t$, il suffit de montrer que, pour toute suite $(t_n)_n$ dans I qui converge vers t , $F(t_n) \rightarrow F(t)$.

Soit $(u_n)_n$ une suite dans I qui converge vers t . Appliquons le TCD à la suite des fonctions $f_n : x \mapsto f(t_n, x)$. Par la première hypothèse, on a, pour presque tout $x \in E$, $f_n(x) \rightarrow f(t, x)$. Par la seconde hypothèse, on a, pour tout n , pour presque tout $x \in E$,

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

et $\int \varphi d\mu < \infty$. Le TCD donne alors exactement $F(t_n) \xrightarrow{n} F(t)$, ce qui conclut. ■

Théorème (Théorème de dérivation sous l'intégrale)

Soit $f : (t, x) \mapsto f(t, x)$ une fonction de $I \times E$ dans $\mathbb{R}^d$ ou $\mathbb{C}$. On suppose que :

- **(existence de F)** pour tout $t \in I$, $x \mapsto f(t, x)$ est intégrable ;
- **(dérivabilité par rapport au paramètre)** pour μ -presque tout $x \in E$, $t \mapsto f(t, x)$ est dérivable sur I , de dérivée notée $\frac{\partial f}{\partial t}$;
- **(domination de la dérivée)** il existe une fonction $\varphi : E \rightarrow \mathbb{R}_+$ mesurable telle que $\int \varphi d\mu < \infty$ et, pour tout $t \in I$, pour μ -presque tout $x \in E$,

$$\left| \frac{\partial f}{\partial t}(t, x) \right| \leq \varphi(x).$$

Alors la fonction

$$F : t \mapsto F(t) = \int f(t, x) d\mu(x)$$

est dérivable sur I et, pour tout $t \in I$,

$$F'(t) = \int \frac{\partial f}{\partial t}(t, x) d\mu(x).$$

Démonstration : Soit $t \in I$. Soit une suite $(t_n)_n$ qui converge vers t dans $I \setminus \{t\}$. Il faut montrer que

$$\frac{F(t_n) - F(t)}{t_n - t} \xrightarrow{n} \int \frac{\partial f}{\partial t}(t, x) d\mu(x).$$

Or par linéarité

$$\frac{F(t_n) - F(t)}{t_n - t} = \int \frac{f(t_n, x) - f(t, x)}{t_n - t} d\mu(x),$$

donc on va appliquer le TCD à la suite $(g_n)_n$ définie par

$$g_n : x \mapsto \frac{f(t_n, x) - f(t, x)}{t_n - t}.$$

Pour presque tout $x \in E$, la deuxième hypothèse donne $g_n(x) \rightarrow \frac{\partial f}{\partial t}(t, x)$. Et l'inégalité des accroissements finis et la troisième hypothèse donnent : pour tout n , pour presque tout $x \in E$,

$$|g_n(x)| \leq \sup_{u \in [t_n, t]} \left| \frac{\partial f}{\partial t}(u, x) \right| \leq \varphi(x),$$

et $\int \varphi d\mu < \infty$. Le TCD donne alors $\int g_n d\mu \rightarrow \int \frac{\partial f}{\partial t}(t, \cdot) d\mu$, ce qui conclut. ■

Remarque : Si de plus la fonction $t \mapsto \frac{\partial f}{\partial t}(t, x)$ est continue pour presque tout x , alors le théorème de continuité montre que F' est continue, et donc que F est de classe $\mathcal{C}^1$ sur I .

Deuxième remarque : Pour montrer que F est de classe $\mathcal{C}^2, \mathcal{C}^3, \dots$, on peut appliquer le théorème plusieurs fois. Pour montrer que F est de classe $\mathcal{C}^\infty$, on peut montrer par récurrence que F est de classe $\mathcal{C}^n$ pour tout n .

3. INTÉGRATION SUR UN ESPACE PRODUIT, CHANGEMENT DE VARIABLES

Soit $(E, \mathcal{A}, \mu)$ et $(F, \mathcal{B}, \nu)$ deux espaces mesurés. On suppose dans ce chapitre que μ et ν sont **σ -finies** : il existe une suite croissante $(E_n)_n$ de parties de E telle que

$$E = \bigcup_n E_n \quad \text{et} \quad \text{pour tout } n, \mu(E_n) < \infty,$$

et de même pour ν .

Pour une fonction $f : E \times F \rightarrow \mathbb{R}$, on peut considérer $\int_E \left(\int_F f(x, y) d\nu(y) \right) d\mu(x)$. On va montrer que l'on peut voir cette double intégration comme une simple intégration sur l'espace $E \times F$. Ce point de vue aura l'avantage de mettre en évidence la symétrie des rôles de E et F , et donc de montrer que l'on peut intervertir l'ordre entre les intégrales précédentes. Il fournit de plus une façon de construire une mesure sur l'espace $E \times F$, qui permettrait de construire l'intégrale de Lebesgue sur $\mathbb{R}^2$ à partir de l'intégrale de Lebesgue sur $\mathbb{R}$, et qui jouera aussi un rôle important dans la notion d'indépendance en probabilités.

1 Produit d'espaces mesurés

On souhaite faire de $E \times F (= \{(x, y) \mid x \in E, y \in F\})$ un espace mesuré, c'est-à-dire le munir d'une tribu et d'une mesure, déduites de celles de E et F .

Définition

La **tribu produit** de $\mathcal{A}$ et $\mathcal{B}$ est la tribu $\mathcal{A} \otimes \mathcal{B}$ engendrée par les pavés $A \times B$ où $A \in \mathcal{A}$ et $B \in \mathcal{B}$:

$$\mathcal{A} \otimes \mathcal{B} = \sigma\left(\{A \times B \mid A \in \mathcal{A}, B \in \mathcal{B}\}\right).$$

Dans le cas des boréliens de $\mathbb{R}^d$, cette opération redonne les tribus déjà connues :

Proposition

Pour tous d, d' , $\mathcal{B}(\mathbb{R}^d) \otimes \mathcal{B}(\mathbb{R}^{d'}) = \mathcal{B}(\mathbb{R}^{d+d'})$.

Par suite, $\mathcal{B}(\mathbb{R}^d) = \mathcal{B}(\mathbb{R}) \otimes \cdots \otimes \mathcal{B}(\mathbb{R}) = \mathcal{B}(\mathbb{R})^{\otimes d}$ en définissant de même une mesure produit de d mesures.

Pour tout ensemble $C \subset E \times F$, on définit ses « tranches » : pour tout $x \in E$,

$$C_x = \{y \in F \mid (x, y) \in C\} \subset F$$

pour tout $y \in F$,

$$C^y = \{x \in E \mid (x, y) \in C\} \subset E.$$

On pourrait vérifier que, si $C \in \mathcal{A} \otimes \mathcal{B}$, alors les tranches sont mesurables.

Théorème

Il existe une unique mesure m sur $(E \times F, \mathcal{A} \otimes \mathcal{B})$ telle que

$$\text{pour tous } A \in \mathcal{A} \text{ et } B \in \mathcal{B}, \quad m(A \times B) = \mu(A)\nu(B),$$

(avec ici $\infty \cdot 0 = 0 \cdot \infty = 0$). On la note $m = \mu \otimes \nu$. De plus, pour tout $C \in \mathcal{A} \otimes \mathcal{B}$,

$$\mu \otimes \nu(C) = \int_E \nu(C_x) d\mu(x) = \int_F \mu(C^y) d\nu(y).$$

Cette mesure s'appelle la **mesure produit** de μ par ν . La dernière formule décrit une intégration « par tranches » : la mesure de C est l'intégrale des mesures de tranches, horizontales ou verticales.

Dans le cas de la mesure de Lebesgue sur $\mathbb{R}^d$, le produit redonne les mesures déjà connues :

Proposition

Pour tous d, d' , $\lambda_d \otimes \lambda_{d'} = \lambda_{d+d'}$.

Par suite, en notant $\lambda = \lambda_1$, on a $\lambda_d = \lambda \otimes \cdots \otimes \lambda = \lambda^{\otimes d}$, en définissant par récurrence le produit de d mesures.

2 Théorèmes de Fubini

Les propriétés de la mesure produit se transfèrent aisément aux propriétés de l'intégrale et permettent de résoudre la question initiale.

Théorème (Théorème de Fubini-Tonelli)

Pour toute fonction mesurable $f : E \times F \rightarrow [0, +\infty]$,

$$\int_{E \times F} f d(\mu \otimes \nu) = \int_E \left(\int_F f(x, y) d\nu(y) \right) d\mu(x) = \int_F \left(\int_E f(x, y) d\mu(x) \right) d\nu(y).$$

Démonstration : La formule est vraie si $f = \mathbf{1}_C$ où $C \in \mathcal{A} \otimes \mathcal{B}$: c'est le précédent théorème. Par suite, la linéarité de l'intégrale montre qu'elle est vraie pour toute fonction étagée positive. Finalement, si f est mesurable positive, alors $f = \lim_{\uparrow n} f_n$ pour une suite croissante $(f_n)_n$ de fonctions étagées positives (sur $E \times F$), ce qui donne en particulier pour tout $x \in E$, $f_x = f(x, \cdot) = \lim_{\uparrow n} f_n(x, \cdot)$ et donc, en appliquant deux fois le TCM,

$$\lim_{\uparrow n} \int_E \left(\int_F f_n(x, \cdot) d\nu \right) d\mu(x) = \int_E \left(\lim_{\uparrow n} \int_F f_n(x, \cdot) d\nu \right) d\mu(x) = \int_E \left(\int_F \lim_{\uparrow n} f_n(x, \cdot) d\nu \right) d\mu(x) = \int_E \left(\int_F f(x, \cdot) d\nu \right) d\mu(x)$$

et, en appliquant une fois le TCM,

$$\lim_{\uparrow n} \int_{E \times F} f_n d(\mu \otimes \nu) = \int_{E \times F} f d(\mu \otimes \nu).$$

L'égalité pour f_n donne alors l'égalité pour f en passant à la limite. L'autre égalité s'obtient de même en échangeant l'ordre d'intégration. ■

En conséquence de ce théorème, on pourra noter $\int_E \int_F f(x, y) d\nu(y) d\mu(x)$ sans parenthèses lorsque f est mesurable positive.

On déduit la mesure produit de deux mesures à densité en appliquant le théorème à $(x, y) \mapsto f(x)g(y)$:

Proposition

Soit $f : E \rightarrow \mathbb{R}_+$ et $g : F \rightarrow \mathbb{R}_+$ mesurables. On a : $(f \cdot \mu) \otimes (g \cdot \nu) = (f \otimes g) \cdot (\mu \otimes \nu)$, où $(f \otimes g)(x, y) = f(x)g(y)$. Autrement dit,

$$(f(x) d\mu(x)) \otimes (g(y) d\nu(y)) = f(x)g(y) d(\mu \otimes \nu)(x, y).$$

Par ailleurs, le théorème précédent justifie la notation $d\mu(x) d\nu(y)$ pour la mesure $d(\mu \otimes \nu)(x, y) = \mu \otimes \nu$ et la proposition prend alors la forme naturelle suivante :

$$(f(x) d\mu(x)) (g(y) d\nu(y)) = f(x)g(y) d\mu(x) d\nu(y).$$

Dans le cas de $\mathbb{R}^2$ avec la mesure de Lebesgue, la mesure produit des mesures $f(x)dx$ et $g(y)dy$ est donc naturellement $f(x)g(y)dx dy$.

En décomposant une fonction f de signe quelconque en $f = f_+ - f_-$, on obtient facilement :

Théorème (Théorème de Fubini-Lebesgue)

Pour toute fonction mesurable f sur $E \times F$ à valeurs dans $\mathbb{R}^d$ ou $\mathbb{C}$, telle que

$$\int_{E \times F} |f(x, y)| d(\mu \otimes \nu)(x, y) < \infty,$$

on a

$$\int_{E \times F} f(x, y) d(\mu \otimes \nu)(x, y) = \int_E \left(\int_F f(x, y) d\nu(y) \right) d\mu(x) = \int_F \left(\int_E f(x, y) d\mu(x) \right) d\nu(y).$$

NB. Par le théorème de Fubini-Tonelli, la condition équivaut à

$$\int_E \left(\int_F |f(x, y)| d\nu(y) \right) d\mu(x) < \infty$$

ou

$$\int_F \left(\int_E |f(x, y)| d\mu(x) \right) d\nu(y) < \infty.$$

3 Changements de variables

3.1 Mesure image

On définit ici une façon de « transporter » une mesure d'un espace à un autre, par une fonction. Cette opération sera notamment importante en probabilités pour définir la loi d'une variable aléatoire.

Soit $\varphi : (E, \mathcal{A}) \rightarrow (F, \mathcal{B})$ une application mesurable. On rappelle que μ est une mesure sur $(E, \mathcal{A})$.

Définition

La **mesure image de μ par φ** est la mesure $\varphi_*\mu$ sur F donnée par :

$$\text{pour tout } B \in \mathcal{B}, \quad \varphi_*\mu(B) = \mu(\varphi^{-1}(B)).$$

C'est bien une mesure : $\varphi_*\mu(\emptyset) = \mu(\varphi^{-1}(\emptyset)) = \mu(\emptyset) = 0$ et, si les B_n sont des parties mesurables disjointes de F , alors les images réciproques $\varphi^{-1}(B_n)$ sont disjointes et leur réunion est $\varphi^{-1}(\biguplus_n B_n)$, donc

$$\varphi_*\mu\left(\biguplus_n B_n\right) = \mu\left(\varphi^{-1}\left(\biguplus_n B_n\right)\right) = \mu\left(\biguplus_n \varphi^{-1}(B_n)\right) = \sum_n \mu(\varphi^{-1}(B_n)) = \sum_n \varphi_*\mu(B_n).$$

Théorème (*Théorème de transfert*)

a) Pour toute fonction mesurable $f : F \rightarrow [0, +\infty]$,

$$\int_F f(y) d(\varphi_*\mu)(y) = \int_E f(\varphi(x)) d\mu(x).$$

b) Pour toute fonction mesurable f sur F à valeurs dans $\mathbb{R}^d$ ou $\mathbb{C}$, f est intégrable par rapport à $\varphi_*\mu$ si, et seulement si $f \circ \varphi$ est intégrable par rapport à μ et, dans ce cas,

$$\int_F f(y) d(\varphi_*\mu)(y) = \int_E f(\varphi(x)) d\mu(x).$$

Démonstration : a) La formule est vraie si $f = \mathbf{1}_B$ où $B \in \mathcal{B}$: c'est exactement la définition de $\varphi_*\mu$. Par suite, la linéarité de l'intégrale montre qu'elle est vraie pour toute fonction étagée positive. Finalement, si f est mesurable positive, alors $f = \lim_{\uparrow n} f_n$ pour une suite croissante $(f_n)_n$ de fonctions étagées positives (sur F), ce qui donne $f \circ \varphi = \lim_{\uparrow n} f_n \circ \varphi$ et donc, en appliquant le TCM, la formule pour la fonction f_n fournit à la limite la formule pour f .

b) Pour f à valeurs réelles de signe quelconque, appliquer le a) à $|f|$ donne

$$\int |f| d(\varphi_*\mu) = \int |f \circ \varphi| d\mu,$$

ce qui montre que f est intégrable par rapport à $\varphi_*\mu$ si, et seulement si $f \circ \varphi$ est intégrable par rapport à μ . Supposons que c'est le cas. On a par définition de l'intégrale puis par a)

$$\int f d(\varphi_*\mu) = \int f_+ d(\varphi_*\mu) - \int f_- d(\varphi_*\mu) \stackrel{a)}{=} \int f_+ \circ \varphi d\mu - \int f_- \circ \varphi d\mu = \int (f_+ - f_-) \circ \varphi d\mu = \int f \circ \varphi d\mu.$$

Pour f à valeurs dans $\mathbb{R}^d$ ou $\mathbb{C}$, on applique ce qui précède à chaque composante. ■

3.2 Changements de variables dans $\mathbb{R}^d$

• Cas linéaire

Proposition

Soit M une application linéaire $\mathbb{R}^d \rightarrow \mathbb{R}^d$. On a, pour $B \in \mathcal{B}(\mathbb{R}^d)$,

$$\lambda_d(M(B)) = |\det M| \lambda_d(B).$$

Autrement dit, si M est inversible,

$$M_*\lambda_d = \frac{1}{|\det M|} \lambda_d.$$

En particulier, pour $B = [0,1]^d$, on a l'interprétation suivante du déterminant de M : c'est le volume du parallélotope engendré par les vecteurs colonnes de M .

Et si $f : \mathbb{R}^d \rightarrow \mathbb{R}$ est intégrable par rapport à λ_d , le théorème de transfert donne

$$\int_{\mathbb{R}^d} f(M(x)) dx = \int_{\mathbb{R}^d} f(y) d(M_* \lambda_d)(y) = \frac{1}{|\det M|} \int_{\mathbb{R}^d} f(y) dy.$$

• Cas des $\mathcal{C}^1$ -difféomorphismes

On a vu que l'image de λ_d par une application linéaire bijective M est $\frac{1}{|\det M|} \lambda_d$.

Dans le cas de l'image par une application *non linéaire* φ , on peut néanmoins vouloir appliquer *localement* le cas linéaire. En effet, une application $\varphi : U \rightarrow D$ (où U et D sont des ouverts de $\mathbb{R}^d$) est **différentiable** si, près de chaque point $x \in U$, elle peut s'approcher par une application linéaire $d\varphi_x$ appelée sa différentielle :

$$\varphi(x+h) = \varphi(x) + d\varphi_x(h) + o_{h \rightarrow 0}(|h|).$$

L'application linéaire $d\varphi_x$ a pour matrice la **matrice jacobienne** de $\varphi = \begin{pmatrix} \varphi_1 \\ \vdots \\ \varphi_d \end{pmatrix}$ en x :

$$\text{Jac}_\varphi(x) = \left(\frac{\partial \varphi_i}{\partial x_j}(x) \right)_{1 \leq i, j \leq d},$$

et son déterminant est le **jacobien** de φ en x :

$$J_\varphi(x) = \det(\text{Jac}_\varphi(x)).$$

Par le cas linéaire, on s'attend à ce que l'image de λ_d par φ soit, au voisinage de $y = \varphi(x)$, proche de $\frac{1}{|J_\varphi(x)|} \lambda_d = |J_{\varphi^{-1}}(y)| \lambda_d$ (à condition que $J_\varphi(x) \neq 0$), ce qui suggère, puisque le facteur $|J_{\varphi^{-1}}(y)|$ dépend du point y près duquel on considère la mesure, que la mesure image est (partout) la mesure à densité $|J_{\varphi^{-1}}(z)| d\lambda_d(z)$ et donc, en intégrant, que $\int f(\varphi(x)) dx = \int f(y) |J_{\varphi^{-1}}(y)| dy$.

Ceci est justifié si $\varphi : U \rightarrow D$ est un $\mathcal{C}^1$ -**difféomorphisme**, c'est-à-dire que φ est de classe $\mathcal{C}^1$, bijective, et que φ^{-1} est aussi de classe $\mathcal{C}^1$ (la dernière condition revient aussi à dire que $J_\varphi(x) \neq 0$ pour tout $x \in U$) :

Théorème (Théorème de changement de variable dans $\mathbb{R}^d$)

Soit U, D des ouverts de $\mathbb{R}^d$. Soit $f : D \rightarrow \mathbb{R}$ mesurable, et $\varphi : U \rightarrow D$ un $\mathcal{C}^1$ -difféomorphisme.

a) Si f est positive, alors

$$\int_D f(y) dy = \int_U f(\varphi(x)) |J_\varphi(x)| dx$$

et

$$\int_U f(\varphi(x)) dx = \int_D f(y) |J_{\varphi^{-1}}(y)| dy.$$

b) Si f est intégrable sur D , la première égalité précédente a un sens (autrement dit, $u \mapsto f(\varphi(u)) |J_\varphi(u)|$ est intégrable sur U) et est vraie. Si $f \circ \varphi$ est intégrable sur U , alors il en est de même de la deuxième.

c) En particulier, la mesure image de λ_d par φ est la mesure de densité $|J_{\varphi^{-1}}|$ par rapport à λ_d :

$$\varphi_*(\lambda_d)|_U = |J_{\varphi^{-1}}| \cdot (\lambda_d)|_D.$$

(Ici, $(\lambda_d)|_D$ est la restriction de λ_d à D , puisque φ n'est définie que sur D)

4. ESPACES L^1 ET L^2

Soit $(E, \mathcal{A}, \mu)$ un espace mesuré. On rappelle la définition

$$\mathcal{L}^1(E, \mathcal{A}, \mu) = \left\{ f : E \rightarrow \mathbb{R} \text{ mesurable} \mid \int |f| d\mu < \infty \right\},$$

et on peut définir de même $\mathcal{L}_{\mathbb{C}}^1(E, \mathcal{A}, \mu)$ pour les fonctions à valeurs dans $\mathbb{C}$.
Plus généralement, si $p > 0$, on note

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

et

$$\mathcal{L}^\infty(E, \mathcal{A}, \mu) = \left\{ f : E \rightarrow \mathbb{R} \text{ mesurable} \mid \text{il existe } C > 0 \text{ tel que } |f(x)| \leq C \text{ pour } \mu\text{-presque tout } x \in E \right\}$$

(on dit que $\mathcal{L}^\infty(E, \mathcal{A}, \mu)$ est l'espace des fonctions « bornées presque partout » sur E)

Comme pour $\mathcal{L}^1$, on abrège souvent $\mathcal{L}^p(E)$ et $\mathcal{L}^\infty(E)$ si la mesure est donnée par le contexte. En particulier, systématiquement, $\mathcal{L}^p(\mathbb{R}) = \mathcal{L}^p(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d), \lambda_d)$.

1 Espace L^1

Pour $f, g : E \rightarrow \mathbb{R}$ mesurables, on rappelle que $f = g$ presque partout (ou μ -presque partout) si

$$\mu(\{x \in E \mid f(x) \neq g(x)\}) = 0$$

et que, dans ce cas, f est intégrable si, et seulement si g est intégrable, et alors $\int f d\mu = \int g d\mu$.

On note $\|f\|_1 = \int |f| d\mu$ pour $f \in \mathcal{L}^1(E)$. On déduit des propriétés de l'intégrale que :

- pour tout $\lambda \in \mathbb{R}$, pour tout $f \in \mathcal{L}^1(E)$, $\|\lambda f\|_1 = |\lambda| \|f\|_1$
- pour tous $f, g \in \mathcal{L}^1(E)$, $\|f + g\|_1 \leq \|f\|_1 + \|g\|_1$.
- pour tout $f \in \mathcal{L}^1(E)$, $\|f\|_1 = 0$ si, et seulement si $f = 0$ presque partout.

À cause du dernier point, $\|\cdot\|_1$ n'est pas une norme : en général, il y a des fonctions non nulles f telles que $\|f\|_1 = 0$. Ceci mène aux définitions suivantes :

Définition

On note $f \sim g$ si $f = g$ presque partout, et on définit $L^1(E, \mathcal{A}, \mu) = \mathcal{L}^1(E, \mathcal{A}, \mu) / \sim$, c'est-à-dire que l'on obtient $L^1(E, \mathcal{A}, \mu)$ en identifiant entre elles les fonctions égales presque partout : les éléments de $L^1(E, \mathcal{A}, \mu)$ sont les classes de fonctions égales entre elles presque partout.

Par exemple, l'élément $0 \in L^1(E, \mathcal{A}, \mu)$ est l'ensemble des fonctions nulles presque partout.

Puisque deux fonctions égales presque partout ont la même intégrale, on pourra sans ambiguïté noter $\int f d\mu$ pour $f \in L^1(E, \mathcal{A}, \mu)$ pour désigner l'intégrale de *n'importe quelle* fonction dans la classe f , et ainsi procéder comme si les éléments de $L^1(E, \mathcal{A}, \mu)$ étaient des fonctions, c'est-à-dire que l'on pourra *confondre* un élément de L^1 (une classe de fonctions) et l'un quelconque de ses représentants (l'une de ces fonctions), à ceci près que si $f, g \in L^1(E, \mathcal{A}, \mu)$ sont égales presque partout, alors $f = g$ par définition.

En particulier, si on définit $\|f\|_1 = \int |f| d\mu$ pour $f \in L^1(E, \mathcal{A}, \mu)$, alors on a les mêmes propriétés qu'avant, **et** $\|f\|_1 = 0$ implique $f = 0$. Ainsi, $\|\cdot\|_1$ est une **norme** sur $L^1(E, \mathcal{A}, \mu)$: par l'identification entre elles des fonctions égales presque partout, on s'est artificiellement ramené à ce qu'une seule « fonction » soit de norme nulle.

Si une suite $(f_n)_n$ dans $L^1(E, \mathcal{A}, \mu)$ converge vers $f \in L^1(E, \mathcal{A}, \mu)$ pour cette norme, on notera $f_n \xrightarrow[n]{L^1(E, \mathcal{A}, \mu)} f$, ou simplement $f_n \xrightarrow[n]{L^1} f$.

Théorème

L^1 l'application $f \mapsto \|f\|_1$ est une norme sur $L^1(E, \mathcal{A}, \mu)$ et, muni de cette norme, l'espace vectoriel $L^1(E, \mathcal{A}, \mu)$ est complet. C'est donc un espace de Banach.

On rappelle qu'un espace est **complet** si toute suite de Cauchy dans cet espace converge : si $(f_n)_n$ est une suite de fonctions dans L^1 telle que $\sup_{p \geq 0} \|f_{n+p} - f_n\|_1 \xrightarrow{n} 0$, alors il existe $f \in L^1$ telle que $\|f_n - f\|_1 \xrightarrow{n} 0$.

On donne quelques résultats utiles de densité :

Proposition

- a) Les fonctions étagées intégrables sont denses dans $L^1(E, \mathcal{A}, \mu)$.
- b) Les fonctions en escalier sont denses dans $L^1(\mathbb{R})$.
- c) Les fonctions de classe $\mathcal{C}^\infty$ à support compact sont denses dans $L^1(\mathbb{R}^d)$.

2 Espace L^2

On note, pour $f \in \mathcal{L}^2(E, \mathcal{A}, \mu)$,

$$\|f\|_2 = \left(\int |f|^2 d\mu \right)^{1/2}$$

et, plus généralement, si $f, g \in \mathcal{L}^2(E, \mathcal{A}, \mu)$, la fonction fg est intégrable puisque $|fg| \leq \frac{|f|^2 + |g|^2}{2}$ (cette inégalité s'obtient en développant $(|f| - |g|)^2 \geq 0$, ce qui est évident), et on note

$$\langle f, g \rangle = \int fg d\mu.$$

$\langle \cdot, \cdot \rangle$ est bilinéaire symétrique positive : pour tous $f, g, h \in \mathcal{L}^2(E, \mathcal{A}, \mu)$, et tout $\lambda \in \mathbb{R}$,

$$\langle \lambda f + g, h \rangle = \lambda \langle f, h \rangle + \langle g, h \rangle, \quad \langle f, g \rangle = \langle g, f \rangle, \quad \text{et} \quad \langle f, f \rangle = \|f\|_2^2 \geq 0.$$

En revanche, $\langle \cdot, \cdot \rangle$ n'est pas définie positive : $\|f\|_2 = 0$ implique $f = 0$ presque partout. Donc $\|\cdot\|_2$ n'est pas une norme, et $\langle \cdot, \cdot \rangle$ n'est pas un produit scalaire sur $\mathcal{L}^2(E)$.

Comme précédemment, on définit $L^2(E, \mathcal{A}, \mu) = \mathcal{L}^2(E, \mathcal{A}, \mu) / \sim$, où $f \sim g$ lorsque $f = g$ presque partout, c'est-à-dire que l'on identifie entre elles les fonctions de $\mathcal{L}^2(E)$ qui sont égales presque partout. Alors :

Théorème

Sur l'espace $L^2(E, \mathcal{A}, \mu)$, l'application $(f, g) \mapsto \langle f, g \rangle$ est un produit scalaire associé à la norme $f \mapsto \|f\|_2$ et, muni de ce produit scalaire, l'espace vectoriel $L^2(E, \mathcal{A}, \mu)$ est complet. C'est donc un espace de Hilbert.

Signalons l'expression que prend l'inégalité de Cauchy-Schwarz : pour toutes fonctions $f, g \in L^2(E)$,

$$\left(\int fg d\mu \right)^2 \leq \int f^2 d\mu \int g^2 d\mu.$$

Les résultats de densité donnés pour L^1 sont aussi valables pour L^2 :

Proposition

- a) Les fonctions étagées intégrables sont denses dans $L^2(E, \mathcal{A}, \mu)$.
- b) Les fonctions en escalier sont denses dans $L^2(\mathbb{R})$.
- c) Les fonctions de classe $\mathcal{C}^\infty$ à support compact sont denses dans $L^2(\mathbb{R}^d)$.

La remarque suivante sera très utile en probabilités :

Proposition

Si $\mu(E) < \infty$, alors $L^2(E) \subset L^1(E)$.

Démonstration : Soit $f \in L^2(E)$. Il faut montrer que $\int |f| d\mu < \infty$. On peut utiliser l'inégalité de Cauchy-Schwarz avec $g = 1$:

$$\left(\int |f| d\mu \right)^2 \leq \int |f|^2 d\mu \int 1 d\mu = \|f\|_2^2 \mu(E)$$

et le dernier terme est fini car $f \in L^2(E)$ et $\mu(E) < \infty$ par l'hypothèse. ■

Probabilités

1. FONDEMENTS DES PROBABILITÉS

L'objectif de ce chapitre est de constater que la théorie de l'intégration développée dans la première partie du cours fournit un cadre rigoureux pour les probabilités. La théorie sera donc la même, mais l'interprétation en est différente : on cherche à fournir un modèle mathématique pour une « expérience aléatoire ».

Une première partie va donc consister à relire les résultats de théorie de l'intégration en ayant en tête cette intuition. Ceci va de pair avec un nouveau vocabulaire, que l'on va commencer par introduire.

1 Définitions

1.1 Espace de probabilités

Définition

Un **espace de probabilité** est un espace mesuré $(\Omega, \mathcal{A}, P)$ où la mesure P a pour masse totale 1 :

$$P(\Omega) = 1.$$

On appelle P une **probabilité**, ou une **mesure de probabilité**. Ω est parfois appelé l'**univers**, ou l'**espace des éventualités**. Les parties mesurables $A \in \mathcal{A}$ sont appelés des **événements**. Un événement est **presque sûr** si $P(A) = 1$; on dit aussi que A est réalisé **presque sûrement** (en abrégé, *p.s.*).

Une interprétation en est la suivante :

- Ω représente l'ensemble de toutes les éventualités possibles, toutes les réalisations possibles du hasard dans l'expérience aléatoire considérée.
- $\mathcal{A}$ est l'ensemble des « événements », c'est-à-dire des ensembles d'éventualités dont on peut évaluer la probabilité.
- Pour $A \in \mathcal{A}$, $P(A)$ représente la probabilité d'occurrence de l'événement A . On peut s'en faire diverses intuitions, qui pourront être justifiées par la théorie qui va suivre :
 - un point de vue *a priori*, où des considérations de symétrie, par exemple, ou un calcul lié aux propriétés physiques mises en jeu par l'expérience, permettent de justifier la répartition des probabilités (par exemple, pour un dé équilibré, l'occurrence de chaque face devrait avoir même probabilité, donc $1/6$),
 - un point de vue *a posteriori*, où $P(A)$ est vu comme la fréquence asymptotique de réalisation de l'événement A si on répète l'expérience un grand nombre de fois (par exemple, si on tire le même dé un grand nombre de fois, on observe que chaque face apparaît en moyenne approximativement lors de $1/6$ des tirages, et cette approximation a tendance à s'améliorer avec le nombre de tirages).

NB. Malgré l'importance théorique de l'espace de probabilité $(\Omega, \mathcal{A}, P)$, on verra dans la suite qu'une particularité fondamentale de la théorie des probabilités est qu'il ne sera souvent pas nécessaire de spécifier l'espace de probabilités car on ne le verra qu'à travers les « variables aléatoires ».

1.2 Variables aléatoires

Soit $(\Omega, \mathcal{A}, P)$ un espace de probabilité.

Définition

Une **variable aléatoire** (en abrégé, *v.a.*) est une application mesurable $X : \Omega \rightarrow E$, où $(E, \mathcal{E})$ est un espace mesurable.

On parle de **variable aléatoire réelle** si l'espace d'arrivée est $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$.

La définition suivante est fondamentale.

Définition

La **loi** d'une variable aléatoire $X : \Omega \rightarrow E$ est la mesure image de P par X . C'est donc la mesure de probabilité P_X sur $(E, \mathcal{E})$ donnée par

$$P_X(B) = P(X^{-1}(B)) = P(\{\omega \in \Omega \mid X(\omega) \in B\}) \quad \text{pour } B \in \mathcal{E}.$$

On dit que X **suit la loi** μ si la loi de X est μ .

Notation fonctionnelle. On utilisera en général la notation $\{X \in B\} = X^{-1}(B)$, de sorte que la définition s'écrit

$$P_X(B) = P(X \in B).$$

De même, on écrira par exemple, pour une variable aléatoire réelle X , $\{\sin(X) \geq 0\} = \{\omega \in \Omega \mid \sin(X(\omega)) \geq 0\}$.

Variables discrètes et continues

Deux familles de lois méritent une attention particulière : les lois dites discrètes et continues. Attention, ce ne sont que des cas particuliers, et de nombreuses lois ne sont ni discrètes ni continues.

• **Variables aléatoires discrètes.** Dans le cas où X prend ses valeurs dans un espace E dénombrable, on dit que X est une variable aléatoire **discrète**, et dans ce cas la loi de X est donnée par les valeurs $p_x = P(X = x)$ pour $x \in E$. En effet, pour tout $B \subset E$,

$$P_X(B) = P(X \in B) = P\left(\bigcup_{x \in B} \{X = x\}\right) = \sum_{x \in B} P(X = x) = \sum_{x \in B} p_x.$$

Autrement dit,

$$P_X = \sum_{x \in E} p_x \delta_x.$$

Donner la loi de X revient donc à calculer les valeurs p_x pour $x \in E$.

• **Variables aléatoires continues (ou à densité).** Dans le cas où X est à valeurs dans $\mathbb{R}^d$ et la loi de X admet une densité f par rapport à la mesure de Lebesgue, on dit que X est une variable aléatoire **continue**, ou **à densité**, de densité f .

Autrement dit, X a pour densité f si f est une fonction mesurable positive qui vérifie, pour tout $A \in \mathcal{B}(\mathbb{R}^d)$,

$$P_X(A) = P(X \in A) = \int_A f(x) dx = \int \mathbf{1}_A(x) f(x) dx.$$

Remarquons que $1 = P(X \in \mathbb{R}^d) = \int f(x) dx$.

Une fonction mesurable $f : \mathbb{R}^d \rightarrow \mathbb{R}$ est une **densité** si $f(x) \geq 0$ pour tout $x \in \mathbb{R}^d$ et $\int f(x) dx = 1$. Toute densité f définit une loi de probabilité.

Propriétés

Si X est une variable aléatoire de densité f , alors

- a) pour tout $x \in \mathbb{R}^d$, $P(X = x) = 0$. Autrement dit, pour tout $x \in \mathbb{R}^d$, presque sûrement, $X \neq x$
- b) presque sûrement, $X \in \{x \in \mathbb{R}^d \mid f(x) > 0\}$.

Démonstration : a) On a $\lambda_d(\{x\}) = 0$ et donc

$$P(X = x) = \int_{\{x\}} f(t) dt = \lambda(\{x\}) f(x) = 0.$$

b) Si f est nulle sur $B \in \mathcal{B}(\mathbb{R}^d)$ (autrement dit, $f(x) = 0$ pour tout $x \in B$), alors $P(X \in B) = \int_B f(t) dt = 0$ donc p.s., $X \in B^c$. Le résultat correspond au cas où $B = \{x \in \mathbb{R}^d \mid f(x) = 0\}$, car f est évidemment nulle sur B . ■

Tribu engendrée

Définition

Soit X une variable aléatoire à valeurs dans $(E, \mathcal{E})$. La **tribu engendrée** par X est

$$\sigma(X) = \{X^{-1}(B) \mid B \in \mathcal{E}\} \subset \mathcal{A}.$$

C'est la tribu sur Ω qui contient tous les événements qui ne dépendent que de X . La proposition suivante montre que, de même, les fonctions mesurables par rapport à $\sigma(X)$ sont celles qui ne dépendent que de X :

Proposition

Soit X, Y des variables aléatoires à valeurs dans $\mathbb{R}^m$ et $\mathbb{R}^n$. Y est mesurable par rapport à $\sigma(X)$ si et seulement s'il existe une fonction mesurable $f : \mathbb{R}^m \rightarrow \mathbb{R}^n$ telle que $Y = f(X)$ p.s.

1.3 Espérance

Définition

Soit X une variable aléatoire réelle. Son **espérance** est

$$E[X] = \int_{\Omega} X(\omega) dP(\omega),$$

ce qui, en tant qu'intégrale d'une fonction mesurable, est bien défini dans les deux cas suivants :

- si $X \geq 0$ (et dans ce cas $E[X] \in [0, \infty]$)
- si X est intégrable, c'est-à-dire $E[|X|] = \int |X| dP < \infty$.

On interprète $E[X]$ comme la moyenne de la variable aléatoire X .

On a en particulier $E[\mathbf{1}_B] = P(B)$ et, pour toute constante $c \in \mathbb{R}$, $E[c] = cP(\Omega) = c$.

Le théorème de transfert (du chapitre sur les changements de variables) s'écrit comme suit :

Proposition (Théorème de transfert)

Soit X une variable aléatoire à valeurs dans $(G, \mathcal{G})$ et φ une fonction mesurable $G \rightarrow \mathbb{R}$ telle que $E[\varphi(X)]$ est bien définie. Alors

$$E[\varphi(X)] = \int_G \varphi(x) dP_X(x).$$

Ceci montre que l'espérance de toute fonction d'une variable aléatoire X ne dépend que de la loi P_X de X , et non de la façon exacte donc X est définie comme fonction sur Ω .

• Si X est discrète à valeurs dans G , on a, par le théorème de transfert, pour toute fonction $\varphi : G \rightarrow \mathbb{R}$ positive (ou telle que $\varphi(X)$ est intégrable),

$$E[\varphi(X)] = \sum_{x \in G} \varphi(x) P(X = x).$$

• Si X est continue sur $\mathbb{R}^d$, de densité f , le théorème de transfert donne, pour toute fonction $\varphi : \mathbb{R}^d \rightarrow \mathbb{R}$ positive (ou telle que $\varphi(X)$ est intégrable),

$$E[\varphi(X)] = \int_{\mathbb{R}^d} \varphi(x) f(x) dx.$$

Tous les résultats vus pour les intégrales sont toujours valables. Écrivons-en quelques-uns à titre d'exemple :

Proposition

Soit X une variable aléatoire réelle **positive**.

- (Inégalité de Markov) Pour tout $a > 0$,

$$P(X \geq a) \leq \frac{E[X]}{a}.$$

- Si $E[X] < \infty$, alors $X < \infty$ presque sûrement.
- Si $E[X] = 0$, alors $X = 0$ presque sûrement.

Proposition

Soit $(X_n)_{n \geq 0}$ une suite de variables aléatoires **positives**.

- (TCM) Si la suite $(X_n)_n$ est croissante et converge vers X , alors $\lim_n \uparrow E[X_n] = E[X]$.

- (TCM pour les séries) On a $E\left[\sum_{n \geq 0} X_n\right] = \sum_{n \geq 0} E[X_n]$.

Proposition (Théorème de convergence dominée)

Soit $(X_n)_{n \geq 0}$ une suite de variables aléatoires réelles, et X une variable aléatoire réelle.

Si $X_n \xrightarrow[n]{p.s.} X$ p.s. et s'il existe Z intégrable telle que $|X_n| \leq Z$ p.s., alors

$$E[X_n] \xrightarrow[n]{} E[X].$$

On pourra utiliser les théorèmes de Fubini ou encore les théorèmes pour les intégrales (espérances) à paramètre. Enfin, on définit aussi les espaces $L^1(\Omega, \mathcal{A}, P)$ et $L^2(\Omega, \mathcal{A}, P)$, abrégés en L^1 et L^2 , et on rappelle dans ce cas l'inclusion $L^2 \subset L^1$ (en effet, pour tout v.a. réelle X , $E[|X|] = E[|1 \cdot X|] \leq E[X^2]^{1/2} E[1]^{1/2} = E[X^2]^{1/2}$). Autrement dit, les v.a. de carré intégrable sont intégrables.

Définition

Soit X une variable aléatoire de carré intégrable. La **variance** de X est le réel positif

$$\text{Var}(X) = E\left[\left(X - E[X]\right)^2\right].$$

L'**écart-type** de X est le réel positif

$$\sigma(X) = \sqrt{\text{Var}(X)}.$$

La variance de X est la moyenne du carré de l'écart entre X et sa moyenne. C'est donc une mesure de la « dispersion » de la loi de X autour de son espérance $E[X]$, de même que l'écart-type. L'écart-type de X a l'intérêt d'être homogène à X , c'est-à-dire que $\sigma(aX) = a\sigma(X)$ pour $a \geq 0$ (tandis que $\text{Var}(aX) = a^2 \text{Var}(X)$), de sorte qu'il pourra être pertinent de comparer les valeurs de $\sigma(X)$ à celles de $X - E[X]$; on verra cela plus loin.

En développant le carré et en utilisant la linéarité de l'espérance, on obtient

$$\text{Var}(X) = E[X^2 - 2XE[X] + E[X]^2] = E[X^2] - 2E[X]E[X] + E[X]^2$$

d'où l'expression importante

$$\text{Var}(X) = E[X^2] - E[X]^2.$$

Proposition (Inégalité de Tchebychev)

Si X est une variable aléatoire de carré intégrable, et $a > 0$, alors

$$P(|X - E[X]| \geq a) \leq \frac{\text{Var}(X)}{a^2}.$$

Démonstration : Pour tout $a > 0$, on a, pour tout $\omega \in \Omega$, $|X(\omega) - E[X]| \geq a$ si, et seulement si $|X(\omega) - E[X]|^2 \geq a^2$, donc

$$\{|X - E[X]| \geq a\} = \{|X - E[X]|^2 \geq a^2\}$$

et par conséquent

$$P(|X - E[X]| \geq a) = P(|X - E[X]|^2 \geq a^2).$$

Or

$$P(|X - E[X]|^2 \geq a^2) \leq \frac{E[|X - E[X]|^2]}{a^2}$$

en appliquant l'inégalité de Markov à la variable aléatoire positive $(X - E[X])^2$. La conclusion vient de la définition de la variance. ■

1.4 Interprétations probabilistes

Le tableau suivant résume les correspondances entre la théorie de l'intégration et les probabilités, pour ce qui est du vocabulaire et de la notation :

Intégration	Probabilités
espace mesuré $(E, \mathcal{A}, \mu)$	espace de probabilités $(\Omega, \mathcal{A}, P)$
point $x \in E$	éventualité, réalisation $\omega \in \Omega$
espace E	univers, ensemble des éventualités Ω
partie mesurable $A \in \mathcal{A}$	événement $A \in \mathcal{A}$
fonction mesurable $f : E \rightarrow F$	variable aléatoire $X : \Omega \rightarrow F$
$A \cup B$	A ou B
$A \cap B$	A et B
A et B disjoints ($A \cap B = \emptyset$)	A et B incompatibles
complémentaire $A^c = E \setminus A$	négation $A^c = \Omega \setminus A$ (aussi noté $\overline{A}$)
$A \subset B$	A implique B
mesure μ	probabilité P
A^c est négligeable ($\mu(A^c) = 0$)	A est presque sûr ($P(A) = 1$)
presque partout, ... (en abrégé, p.p.)	presque sûrement, ... (en abrégé, p.s.)
mesure image de μ par f	loi de X (mesure image de P par X)
intégrale $\int_E f d\mu$	espérance (moyenne) $E[X] = \int_{\Omega} X dP$

2 Caractériser une loi

On sera souvent amené à déterminer la loi d'une variable aléatoire. Une première approche, qui découle de la définition, consistera à utiliser l'une des équivalences données par la proposition suivante :

Proposition

Soit X et Y des v.a. à valeurs dans $\mathbb{R}^d$. Les propriétés suivantes sont équivalentes :

- (i) X et Y ont même loi
- (ii) pour toute partie $A \in \mathcal{B}(\mathbb{R}^d)$, $P(X \in A) = P(Y \in A)$
- (iii) pour toute fonction mesurable $f : \mathbb{R}^d \rightarrow \mathbb{R}_+$, $E[f(X)] = E[f(Y)]$
- (iv) pour toute fonction C^∞ à support compact f , $E[f(X)] = E[f(Y)]$.

Dans certains cas, on utilisera plutôt le calcul d'une fonction associée à X et qui caractérise sa loi : fonction de répartition, fonction génératrice ou fonction caractéristique.

2.1 Fonction de répartition (variables aléatoires réelles)

Définition

Soit X une variable aléatoire réelle. Sa **fonction de répartition** est la fonction

$$\begin{aligned} F_X : \mathbb{R} &\rightarrow [0,1] \\ t &\mapsto F_X(t) = P(X \leq t). \end{aligned}$$

On constate que F_X ne dépend que de la loi de X : $F_X(t) = P_X([-\infty, t])$, donc on pourra dire aussi que F_X est la fonction de répartition de la loi de X .

Proposition

- a) La loi de X est déterminée par sa fonction de répartition : si $F_X(t) = F_Y(t)$ pour tout $t \in \mathbb{R}$, alors X et Y ont même loi.
- b) La fonction F_X est croissante, continue à droite, et admet pour limites 0 en $-\infty$ et 1 en $+\infty$.
- c) Pour tout $t \in \mathbb{R}$, la limite de F_X en t à gauche est

$$F_X(t^-) = P(X < t)$$

donc F_X est continue en t si et seulement si $P(X = t) = 0$.

- d) Toute fonction qui vérifie les propriétés de b) est la fonction de répartition d'une loi de probabilité sur $\mathbb{R}$.

Démonstration : On prouve seulement b) et c). Pour tous $s < t$, on a $\{X \leq s\} \subset \{X \leq t\}$ donc

$$F_X(s) = P(X \leq s) \leq P(X \leq t) = F_X(t),$$

ce qui montre que F_X est croissante. Soit $t \in \mathbb{R}$. Comme F_X est croissante, elle admet une limite à droite en t , que l'on peut obtenir comme limite le long de la suite $(t + \frac{1}{n})_n$:

$$\lim_{s \rightarrow t^+} F_X(s) = \lim_n F_X\left(t + \frac{1}{n}\right).$$

Or $F_X(t + \frac{1}{n}) = P(X \leq t + \frac{1}{n})$ et les événements $\{X \leq t + \frac{1}{n}\}$ forment une suite décroissante dont l'intersection est l'événement $\{X \leq t\}$, donc

$$\lim_n F_X\left(t + \frac{1}{n}\right) = \lim_n P\left(X \leq t + \frac{1}{n}\right) = P(X \leq t) = F_X(t).$$

Ceci montre la continuité de F_X à droite en t .

De même, la limite en $-\infty$ existe par croissance de F_X , et on a par exemple

$$\lim_{t \rightarrow -\infty} F_X(t) = \lim_n F_X(-n)$$

or $F_X(-n) = P(X \leq -n)$ et les événements $\{X \leq -n\}$ forment une suite décroissante dont l'intersection est $\{X = -\infty\} = \emptyset$, d'où

$$\lim_n F_X(-n) = \lim_n P(X \leq -n) = P(\emptyset) = 0.$$

ceci montre que la limite de F_X en $-\infty$ est 0. ■

Les cas des variables discrètes et continues sont particulièrement importants :

Proposition

a) Si X est une variable aléatoire discrète, F_X est une fonction constante entre les éléments de

$$E = \{x \in \mathbb{R} \mid P(X = x) > 0\},$$

et telle que le saut en $x \in E$ a pour hauteur $P(X = x)$.

b) Si X est une variable aléatoire de densité f_X , on a

$$\text{pour tout } x \in \mathbb{R}, \quad F_X(x) = \int_{-\infty}^x f_X(t) dt,$$

donc, si f_X est continue par morceaux, F_X est la primitive de f_X nulle en $-\infty$; et on a la dérivée $(F_X)'(x) = f_X(x)$ pour tout x où f_X est continue.

Inversement, si F_X est continue sur $\mathbb{R}$, et de classe $\mathcal{C}^1$ par morceaux, alors X admet pour densité $f_X = F_X'$ (avec une valeur quelconque aux points où F_X n'est pas dérivable).

2.2 Fonction génératrice (variables aléatoires entières)

Définition

Soit X une variable aléatoire à valeurs dans $\mathbb{N}$. La **fonction génératrice** de X est la fonction

$$G_X : s \mapsto G_X(s) = E[s^X] = \sum_{n=0}^{\infty} s^n P(X = n).$$

Notons tout de suite que G_X est au moins définie sur $[-1, 1]$, en effet c'est une série entière et comme $\sum_{n=0}^{\infty} P(X = n) = 1$ converge, $G_X(1)$ et $G_X(-1)$ convergent absolument. Ainsi le rayon de convergence est supérieur ou égal à 1.

Proposition

G_X caractérise la loi de X : si pour deux variables X et Y à valeurs dans $\mathbb{N}$ on a $G_X(s) = G_Y(s)$ pour tout $s \in]-1, 1[$, alors X et Y ont même loi.

Démonstration : Comme le rayon de convergence de G_X et G_Y est ≥ 1 , on peut dériver terme à terme sur $]-1, 1[$ (on pourrait aussi appliquer le théorème de dérivation pour le justifier) : pour tout $s \in]-1, 1[$, pour tout $k \geq 0$,

$$G_X^{(k)}(s) = \sum_{n=k}^{\infty} n(n-1) \cdots (n-k+1) s^{n-k} P(X = n) = k! P(X = k) + \frac{k!}{1!} s P(X = 1) + \frac{k!}{2!} s^2 P(X = 2) + \cdots$$

d'où $G_X^{(k)}(0) = k! P(X = k)$ et de même pour Y . Comme G_X et G_Y sont égales sur $]-1, 1[$, leurs dérivées successives en 0 sont les mêmes d'où, pour tout k , $P(X = k) = P(Y = k)$, ce qui implique que X et Y ont même loi (puisqu'elles sont à valeurs dans $\mathbb{N}$). ■

2.3 Fonction caractéristique (variables aléatoires à valeurs dans $\mathbb{R}^d$)

Définition

Soit X une variable aléatoire à valeurs dans $\mathbb{R}^d$. La **fonction caractéristique** de X est la fonction

$$\left| \begin{array}{ll} \Phi_X : \mathbb{R}^d & \rightarrow \mathbb{C} \\ t & \mapsto \Phi_X(t) = E[e^{it \cdot X}], \end{array} \right.$$

où $\cdot$ désigne le produit scalaire (ou le produit usuel si $d = 1$).

Comme $|e^{it \cdot X}| = 1$ et $E[1] = 1 < \infty$, la fonction $t \mapsto e^{it \cdot X}$ est intégrable par rapport à P donc Φ_X est bien définie sur $\mathbb{R}^d$, et vérifie d'ailleurs $|\Phi_X(t)| \leq 1$ par inégalité triangulaire.

Si X a pour densité f , alors pour tout $t \in \mathbb{R}$,

$$\Phi_X(t) = \int e^{itx} f(x) dx$$

donc Φ_X est la (conjuguée de la) transformée de Fourier de f .

Proposition

Φ_X caractérise la loi de X : si pour deux variables X et Y à valeurs dans $\mathbb{R}^d$ on a $\Phi_X(t) = \Phi_Y(t)$ pour tout $t \in \mathbb{R}^d$, alors X et Y ont même loi.

On peut donner quelques propriétés de Φ_X , qui s'obtiennent via les théorèmes de continuité et dérivation sous l'intégrale.

Proposition

Soit X une variable aléatoire réelle.

- a) La fonction Φ_X est continue sur $\mathbb{R}$ et $\Phi_X(0) = 1$.
- b) Si X est intégrable, alors Φ_X est de classe $\mathcal{C}^1$ sur $\mathbb{R}$ et $\Phi'_X(0) = iE[X]$.
- c) De manière générale, si $E[|X|^m] < \infty$, alors Φ_X est de classe $\mathcal{C}^m$ sur $\mathbb{R}$ et $\Phi_X^{(m)}(0) = i^m E[X^m]$, d'où le développement limité

$$\Phi_X(t) = 1 + \sum_{n=1}^m \frac{(it)^n}{n!} E[X^n] + o_{t \rightarrow 0}(t^m).$$

3 Indépendance

3.1 Probabilité conditionnelle

Définition

Si $A, B \in \mathcal{A}$ sont deux événements, avec $P(B) > 0$, la **probabilité de A sachant B** est

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

C'est la probabilité que A se réalise si on a l'information que B est réalisé.

On note que $A \mapsto P(A|B)$ est une probabilité sur $(\Omega, \mathcal{A})$: c'est la mesure de densité $\frac{1_B}{P(B)}$ par rapport à P . On peut donc considérer l'espérance par rapport à cette probabilité et elle est donnée, pour toute variable aléatoire X positive (ou intégrable), par

$$E[X|B] = \int_{\Omega} X dP(\cdot|B) = \frac{1}{P(B)} \int X 1_B dP = \frac{E[X 1_B]}{P(B)}.$$

Proposition

Soit $(B_n)_{1 \leq n \leq N}$ une suite (avec $N \in \mathbb{N}$ ou $N = \infty$) d'événements qui partitionne Ω , c'est-à-dire que

$$\Omega = \bigsqcup_{1 \leq n \leq N} B_n.$$

Pour tout événement A , et toute variable aléatoire X positive ou intégrable,

a) (**Formule des probabilités totales**) $P(A) = \sum_{n=1}^N P(A \cap B_n) = \sum_{n=1}^N P(A|B_n)P(B_n)$

et $E[X] = \sum_{n=1}^N E[X 1_{B_n}] = \sum_{n=1}^N E[X|B_n]P(B_n)$

b) (**Formule de Bayes**) pour tout $1 \leq n \leq N$, $P(B_n|A) = \frac{P(A|B_n)P(B_n)}{\sum_{k=1}^N P(A|B_k)P(B_k)}.$

Démonstration : a) La première formule se déduit du fait que $A = \bigsqcup_n (A \cap B_n)$, et la deuxième vient de $X = \sum_n X 1_{B_n}$.

La formule de Bayes se déduit directement de $P(B_n|A) = \frac{P(B_n \cap A)}{P(A)} = \frac{P(A|B_n)P(B_n)}{P(A)}$ et de a). ■

3.2 Événements indépendants

Définition

Deux événements A, B sont **indépendants** si $P(A \cap B) = P(A)P(B)$.

Si $P(B) > 0$, ceci revient donc à $P(A|B) = P(A)$: savoir que B est réalisé n'affecte pas la probabilité que A soit réalisé ou non. Cela correspond bien au sens courant d'« indépendance ».

On généralise la définition :

Définition

Les événements $A_1, \dots, A_n$ sont **indépendants** si, pour tous $i_1 < \dots < i_k$,

$$P(A_{i_1} \cap \dots \cap A_{i_k}) = P(A_{i_1}) \cdots P(A_{i_k}).$$

Autrement dit, par exemple, A, B, C sont indépendants si

$$P(A \cap B \cap C) = P(A)P(B)P(C), \quad P(A \cap B) = P(A)P(B), \quad P(A \cap C) = P(A)P(C) \quad \text{et} \quad P(B \cap C) = P(B)P(C).$$

Attention, il ne suffit pas de vérifier la première égalité, ou les trois suivantes.

3.3 Variables aléatoires indépendantes

Définition

Les variables aléatoires $X_1, \dots, X_n$, à valeurs dans $(E_1, \mathcal{E}_1), \dots, (E_n, \mathcal{E}_n)$ sont **indépendantes** si pour tous $B_1 \in \mathcal{E}_1, \dots, B_n \in \mathcal{E}_n$,

$$P(\{X_1 \in B_1\} \cap \dots \cap \{X_n \in B_n\}) = P(X_1 \in B_1) \cdots P(X_n \in B_n).$$

Intuitivement, $X_1, \dots, X_n$ sont indépendantes si la connaissance de certaines d'entre elles n'apporte aucune information sur les autres : cela correspond encore à la notion intuitive d'« indépendance ».

La définition d'indépendance rappelle la mesure produit : on peut la réécrire sous la forme

$$\begin{aligned} P_{(X_1, \dots, X_n)}(B_1 \times \dots \times B_n) &= P_{X_1}(B_1) \cdots P_{X_n}(B_n) \\ &= P_{X_1} \otimes \dots \otimes P_{X_n}(B_1 \times \dots \times B_n). \end{aligned}$$

On a ici noté $P_{(X_1, \dots, X_n)}$ la loi du vecteur $(X_1, \dots, X_n)$. Ceci nous donne le résultat suivant :

Théorème

Les variables aléatoires $X_1, \dots, X_n$ sont indépendantes si, et seulement si la loi du vecteur $(X_1, \dots, X_n)$ est le produit des lois de $X_1, \dots, X_n$:

$$P_{(X_1, \dots, X_n)} = P_{X_1} \otimes \dots \otimes P_{X_n}.$$

On a alors, pour toutes fonctions $f_1, \dots, f_n$ mesurables positives, ou intégrables,

$$E[f_1(X_1) \cdots f_n(X_n)] = E[f_1(X_1)] \cdots E[f_n(X_n)].$$

Le second point est la traduction du théorème de Fubini.

En particulier, si X et Y sont indépendantes et positives (ou intégrables), $E[XY] = E[X]E[Y]$.

Proposition

Soit $X_1, \dots, X_n$ des variables aléatoires réelles.

a) Si $X_1, \dots, X_n$ sont de carré intégrables et **indépendantes**, alors

$$\text{Var}(X_1 + \dots + X_n) = \text{Var}(X_1) + \dots + \text{Var}(X_n).$$

b) Si $X_1, \dots, X_n$ sont à valeurs dans $\mathbb{N}$ et **indépendantes**, alors

$$\text{pour tout } s \in [-1, 1], \quad G_{X_1 + \dots + X_n}(s) = G_{X_1}(s) \cdots G_{X_n}(s).$$

c) Si $X_1, \dots, X_n$ sont **indépendantes**, alors

$$\text{pour tout } t \in \mathbb{R}, \quad \Phi_{X_1 + \dots + X_n}(t) = \Phi_{X_1}(t) \cdots \Phi_{X_n}(t).$$

Démonstration : On raisonne pour deux variables indépendantes X et Y .

a) On a $E[X + Y] = E[X] + E[Y]$ donc

$$\begin{aligned}\text{Var}(X + Y) &= E\left[\left((X - E[X]) + (Y - E[Y])\right)^2\right] \\ &= E[(X - E[X])^2] + 2E[(X - E[X])(Y - E[Y])] + E[(Y - E[Y])^2] \\ &= \text{Var}(X) + E[X - E[X]]E[Y - E[Y]] + \text{Var}(Y)\end{aligned}$$

où la dernière égalité utilise l'indépendance. Et $E[X - E[X]] = E[X] - E[X] = 0$, d'où le résultat.

b) On a, pour tout s ,

$$G_{X+Y}(s) = E[s^{X+Y}] = E[s^X s^Y] = E[s^X]E[s^Y] = G_X(s)G_Y(s).$$

c) On a, pour tout $t \in \mathbb{R}$,

$$\Phi_{X+Y}(t) = E[e^{it(X+Y)}] = E[e^{itX} e^{itY}] = E[e^{itX}]E[e^{itY}] = \Phi_X(t)\Phi_Y(t).$$

Ces propriétés permettent de calculer la fonction génératrice (ou caractéristique) de la somme $X + Y$ de deux v.a. indépendantes à partir de celles de X et de Y , et donc d'en déduire la loi de $X + Y$.

Remarque. La définition d'indépendance de variables aléatoires peut aussi s'introduire en définissant l'indépendance d'une famille de tribus, qui sera utile en *Probabilités 2* :

Définition

Les tribus $\mathcal{A}_1, \dots, \mathcal{A}_n \subset \mathcal{A}$ sont **indépendantes** si, pour tous $A_1 \in \mathcal{A}_1, \dots, A_n \in \mathcal{A}_n$, ces événements sont indépendants.

Les variables aléatoires $X_1, \dots, X_n$ sont **indépendantes** si les tribus engendrées $\sigma(X_1), \dots, \sigma(X_n)$ le sont.

On donne deux façons générales d'obtenir de nouvelles variables aléatoires indépendantes à partir d'une première famille de variables indépendantes :

Propriétés

Soit $X_1, \dots, X_n$ des variables aléatoires indépendantes.

a) Pour toutes fonctions mesurables $f_1, \dots, f_n$, les variables aléatoires $f_1(X_1), \dots, f_n(X_n)$ sont indépendantes.

b) (« **Indépendance par paquets** ») Pour tous $1 \leq i_1 < \dots < i_k \leq n$, les variables aléatoires

$$(X_1, \dots, X_{i_1}), (X_{i_1+1}, \dots, X_{i_2}), \dots, (X_{i_{k-1}+1}, \dots, X_{i_k})$$

sont indépendantes.

En combinant a) et b), ceci montre que si on regroupe $X_1, \dots, X_n$ en paquets disjoints, ces paquets sont indépendants, et donc toutes les familles de v.a. définies comme fonctions qui dépendent de paquets disjoints sont indépendantes : si X, Y, Z, T sont indépendantes et à valeurs réelles, alors $X + Z$ et YT sont indépendantes, de même que $\frac{X}{T}$, Y^2 et $\sqrt{|Z|}$, par exemple.

Indépendance d'une famille infinie. On dit qu'une famille infinie d'événements, de variables aléatoires ou de tribus, est indépendante si toute sous-famille finie est indépendante.

4 Lois classiques

Lois discrètes

- Si E est un ensemble, et $x \in E$, la **loi de Dirac en x** est la loi d'une variable aléatoire X à valeurs dans E telle que

$$P(X = x) = 1.$$

On dit que X est constante (p.s.). *Ce n'est donc pas une variable « aléatoire » mais plutôt « déterministe ».*

- Si E est un ensemble fini, de cardinal n , la **loi uniforme sur E** est la loi d'une variable aléatoire X à valeurs dans E telle que

$$\text{pour tout } x \in E, \quad P(X = x) = \frac{1}{n}.$$

- La **loi de Bernoulli de paramètre $p \in [0,1]$** est la loi (notée $\mathcal{B}(p)$) d'une variable aléatoire X à valeurs dans $\{0,1\}$ telle que

$$P(X = 1) = p, \quad P(X = 0) = 1 - p.$$

On interprète X comme le résultat du lancer d'une pièce biaisée ayant probabilité p de tomber sur pile.

- La **loi binomiale de paramètres $n \in \mathbb{N}$ et $p \in [0,1]$** est la loi (notée $\mathcal{B}(n,p)$) d'une variable aléatoire X à valeurs dans $\{0,1,\dots,n\}$ telle que

$$\text{pour } k = 0, \dots, n, \quad P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}.$$

On interprète X comme le nombre de piles obtenus en n lancers indépendants de la pièce précédente. Ceci résulte de la proposition suivante.

Proposition

Soit $n \in \mathbb{N}$ et $p \in [0,1]$. Si $X_1, \dots, X_n$ sont des variables aléatoires indépendantes, de loi $\mathcal{B}(p)$, alors leur somme

$$S = X_1 + \dots + X_n$$

suit la loi binomiale de paramètres n et p .

- La **loi géométrique de paramètre $p \in]0,1[$** est la loi (notée $\mathcal{G}(p)$) d'une variable aléatoire X à valeurs dans $\mathbb{N}^*$ telle que

$$\text{pour tout } k \in \mathbb{N}^*, \quad P(X = k) = (1-p)^{k-1} p.$$

On interprète X comme l'instant où l'on obtient pile pour la première fois dans une suite de lancers indépendants de la pièce précédente. Ceci résulte de la proposition suivante.

Proposition

Soit $p \in]0,1[$. Si $X_1, X_2, \dots$ est une suite de variables aléatoires indépendantes, de loi $\mathcal{B}(p)$, alors la variable aléatoire

$$N = \min\{n \geq 1 \mid X_n = 1\}$$

suit la loi géométrique de paramètre p .

- La **loi de Poisson de paramètre $\lambda \in]0, +\infty[$** est la loi (notée $\mathcal{P}(\lambda)$) d'une variable aléatoire X à valeurs dans $\mathbb{N}$ telle que

$$\text{pour tout } k \in \mathbb{N}, \quad P(X = k) = e^{-\lambda} \frac{\lambda^k}{k!}.$$

On interprète X comme un nombre d'événements « rares » qui se produisent parmi une « longue » suite de tirages indépendants. Un énoncé plus précis est le suivant.

Proposition

Soit $(p_n)_{n \geq 1}$ une suite de réels dans $[0,1]$ telle que $np_n \xrightarrow{n} \lambda > 0$. Si, pour tout n , S_n suit la loi binomiale de paramètres n et p_n , alors

$$P(S_n = k) \xrightarrow{n} e^{-\lambda} \frac{\lambda^k}{k!}.$$

Lois continues

- La **loi uniforme** sur $[a, b]$ (où $a < b$) est la loi (notée $\mathcal{U}([a, b])$) de densité

$$f(x) = \frac{1}{b-a} \mathbf{1}_{[a,b]}(x).$$

(Si X suit la loi uniforme sur $[a, b]$, alors $X \in [a, b]$ p.s.)

- La **loi exponentielle de paramètre** sur $\lambda \in]0, +\infty[$ est la loi (notée $\mathcal{E}(\lambda)$) de densité

$$f(x) = \lambda e^{-\lambda x} \mathbf{1}_{\mathbb{R}_+}(x).$$

(Si X suit la loi exponentielle de paramètre $\lambda > 0$, alors $X > 0$ p.s.)

- La **loi normale (ou gaussienne) de moyenne** $m \in \mathbb{R}$ **et de variance** $\sigma^2 \in]0, +\infty[$ est la loi (notée $\mathcal{N}(m, \sigma^2)$) de densité

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-m)^2}{2\sigma^2}}.$$

L'appellation est justifiée en vérifiant que m est la moyenne et σ^2 la variance de cette loi. La loi $\mathcal{N}(0, 1)$, appelée loi normale standard, a donc pour densité $\frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$.

5 Compléments sur les vecteurs aléatoires

Une variable aléatoire X à valeurs dans $\mathbb{R}^d$ est parfois appelée vecteur aléatoire. Si on note $X = (X_1, \dots, X_d)$ ses composantes, alors $X_1, \dots, X_d$ sont des variables aléatoires réelles appelées les **marginales** de X . Leurs lois sont les **lois marginales** de la loi de X . Inversement, la loi de X est la **loi jointe** de $X_1, \dots, X_d$.

Attention, il ne suffit pas de connaître les lois marginales pour connaître la loi de X .

On s'intéresse au cas des variables à densité.

Proposition

Soit $X = (X_1, \dots, X_d)$ une variable aléatoires à valeurs dans $\mathbb{R}^d$ de densité la fonction f . Alors $X_1, \dots, X_d$ ont aussi des densités $f_1, \dots, f_d$ données, pour $i = 1, \dots, d$, par

$$f_i : x_i \mapsto f_i(x_i) = \int_{\mathbb{R}^{d-1}} f(x_1, \dots, x_d) dx_1 \cdots dx_{i-1} dx_{i+1} \cdots dx_d.$$

Par exemple, si (X, Y) dans $\mathbb{R}^2$ a pour densité $f_{(X,Y)}$, alors X et Y ont pour densités

$$f_X : x \mapsto f_X(x) = \int_{\mathbb{R}} f_{(X,Y)}(x, y) dy \quad \text{et} \quad f_Y : y \mapsto f_Y(y) = \int_{\mathbb{R}} f_{(X,Y)}(x, y) dx.$$

L'indépendance entre les marginales se voit par le fait que la densité est un produit de fonctions ne dépendant que d'une variable : il résulte directement de la proposition sur la mesure produit de mesures à densité (p. 20) que

Proposition

Soit $X_1, \dots, X_d$ des variables aléatoires réelles.

- a) Si, pour $i = 1, \dots, d$, X_i a pour densité f_i , et $X_1, \dots, X_d$ sont indépendantes, alors la loi de $(X_1, \dots, X_d)$ a pour densité la fonction

$$f : (x_1, \dots, x_d) \mapsto f_1(x_1) \cdots f_d(x_d).$$

- b) Inversement si la loi de $(X_1, \dots, X_d)$ a une densité qui s'écrit sous la forme

$$f(x_1, \dots, x_n) = g_1(x_1) \cdots g_d(x_d),$$

alors $X_1, \dots, X_d$ sont indépendantes, et pour $i = 1, \dots, d$, la densité de X_i est proportionnelle à g_i , et est donc

$$f_i : x \mapsto f_i(x) = \frac{1}{\int_{\mathbb{R}} g_i(y) dy} g_i(x).$$

Image par un $\mathcal{C}^1$ -difféomorphisme. Si $X = (X_1, \dots, X_d)$ a pour densité f , et si presque sûrement $X \in U$, où U est un ouvert de $\mathbb{R}^d$, alors pour tout $\mathcal{C}^1$ -difféomorphisme $\varphi : U \rightarrow V$ (où V est un ouvert de $\mathbb{R}^d$) la variable aléatoire $(Y_1, \dots, Y_d) = \varphi(X_1, \dots, X_d)$ admet une densité. C'est une conséquence du théorème de changement de variable. On calcule la densité en suivant le schéma suivant : pour toute fonction mesurable positive $g : \mathbb{R}^d \rightarrow \mathbb{R}$,

$$\begin{aligned} E[g(Y_1, \dots, Y_d)] &= E[g(\varphi(X_1, \dots, X_d))] = \int g(\varphi(x_1, \dots, x_d)) dP_{(X_1, \dots, X_d)}(x_1, \dots, x_d) \\ &= \int_U g(\varphi(x_1, \dots, x_d)) f(x_1, \dots, x_d) dx_1 \cdots dx_d \\ &= \int_V g(y_1, \dots, y_d) f(\varphi^{-1}(y_1, \dots, y_d)) |J_{\varphi^{-1}}(y_1, \dots, y_d)| dy_1 \cdots dy_d \end{aligned}$$

à l'aide du théorème de changement de variable. La dernière expression est aussi $E[g(Z_1, \dots, Z_d)]$ où $(Z_1, \dots, Z_d)$ a pour densité

$$(y_1, \dots, y_d) \mapsto f(\varphi^{-1}(y_1, \dots, y_d)) |J_{\varphi^{-1}}(y_1, \dots, y_d)|$$

(c'est bien une densité : cette fonction est positive, et on voit que son intégrale vaut 1 en prenant $g = 1$ ci-dessus). Comme l'égalité $E[g(Y_1, \dots, Y_d)] = E[g(Z_1, \dots, Z_d)]$ vaut pour toute fonction mesurable positive g , on en déduit que $(Y_1, \dots, Y_d)$ et $(Z_1, \dots, Z_d)$ ont même loi, donc la fonction ci-dessus est aussi la densité de $(Y_1, \dots, Y_d)$. La formule n'est pas à retenir, il vaut mieux faire le calcul à chaque fois (ce qui réduit le risque d'erreur).

2. SUITES DE VARIABLES ALÉATOIRES

Au terme du chapitre précédent, on dispose de toutes les notions fondamentales pour définir et étudier des questions de probabilités. L'objectif de ce chapitre est maintenant d'en faire usage pour aboutir aux deux principaux théorèmes de la théorie des probabilités : la loi des grands nombres et le théorème central limite.

Ces deux résultats portent sur le comportement asymptotique de suites de variables aléatoires. On commence donc par introduire divers outils et définitions, notamment des notions de convergence pour les suites de variables aléatoires, dont l'intérêt va bien au-delà des théorèmes qui sont le but principal de ce chapitre.

Dans tout ce chapitre, $(\Omega, \mathcal{A}, P)$ est un espace de probabilités.

1 Suites d'événements : deux outils

1.1 Intersection dénombrable d'événements presque sûrs

Lemme

Si $(A_i)_{i \in I}$ est une famille d'événements presque sûrs, c'est-à-dire que $P(A_i) = 1$ pour tout $i \in I$, et si I est dénombrable, alors $\bigcap_{i \in I} A_i$ est presque sûr.

Démonstration : En effet le complémentaire de $\bigcap_{i \in I} A_i$ est $\bigcup_{i \in I} (A_i)^c$ et, comme I est dénombrable,

$$P\left(\bigcup_{i \in I} (A_i)^c\right) \leq \sum_{i \in I} P(A_i^c) = 0.$$

Ainsi, une intersection dénombrable d'événements presque sûrs est presque sûre. Ceci peut se voir comme un échange de quantificateurs : si, pour tout $i \in I$, p.s. A_i a lieu, et que I est dénombrable, alors p.s., pour tout $i \in I$, A_i a lieu.

1.2 Lemme de Borel-Cantelli

Considérons une suite $(A_n)_{n \geq 0}$ d'événements. On définit à partir de ceux-ci deux nouveaux événements : la **limite supérieure** de la suite $(A_n)_{n \geq 0}$

$$\limsup_n A_n = \bigcap_{n \geq 0} \bigcup_{k \geq n} A_k = \{\text{il existe une infinité d'indices } n \text{ tels que } A_n \text{ se réalise}\}$$

et la **limite inférieure** de la suite $(A_n)_{n \geq 0}$

$$\liminf_n A_n = \bigcup_{n \geq 0} \bigcap_{k \geq n} A_k = \{\text{à partir d'un certain rang, } A_n \text{ se réalise}\}.$$

On peut noter que $(\limsup_n A_n)^c = \liminf_n (A_n)^c$, ou encore (pour faire le lien avec les notions de limites supérieure et inférieure pour les suites de réels) que $\mathbf{1}_{\limsup_n A_n} = \limsup_n \mathbf{1}_{A_n}$ et $\mathbf{1}_{\liminf_n A_n} = \liminf_n \mathbf{1}_{A_n}$.

Théorème (Lemme de Borel-Cantelli)

Soit $(A_n)_{n \geq 0}$ une suite d'événements.

a) On suppose que $\sum_{n \geq 0} P(A_n) < \infty$. Alors

$$P(\limsup_n A_n) = 0.$$

b) On suppose que **les événements** $(A_n)_n$ **sont indépendants** et que $\sum_{n \geq 0} P(A_n) = \infty$. Alors

$$P(\limsup_n A_n) = 1.$$

Démonstration : a) On note N le nombre d'événements de la suite $(A_n)_n$ qui se réalisent : c'est la variable aléatoire

$$N = \sum_{n \geq 0} \mathbf{1}_{A_n}.$$

Par le théorème de convergence monotone pour les séries positives,

$$E[N] = \sum_{n \geq 0} E[\mathbf{1}_{A_n}] = \sum_{n \geq 0} P(A_n),$$

donc si $\sum_n P(A_n) < \infty$, alors $E[N] < \infty$, d'où $N < \infty$ p.s., comme annoncé car $P(\limsup_n A) = P(N = \infty)$.

b) Montrons d'abord une propriété utile :

Lemme

Pour toute suite $(A_n)_{n \geq 0}$ d'événements indépendants, $P\left(\bigcap_{n \geq 0} A_n\right) = \prod_{n \geq 0} P(A_n)$.

Démonstration : Pour tout N , on note $B_N = A_0 \cap \dots \cap A_N$. Alors la suite $(B_N)_N$ est décroissante, donc

$$\lim_{N \rightarrow \infty} P(B_N) = P\left(\bigcap_N B_N\right) = P\left(\bigcap_n A_n\right).$$

Or, vu l'indépendance, $P(B_N) = P(A_0) \cdots P(A_N)$, qui converge vers $\prod_{n \geq 0} P(A_n)$, d'où le lemme. ■

On suppose les $(A_n)_n$ indépendants, et $\sum_n P(A_n) = \infty$. On a

$$P(\limsup_n A_n) = P\left(\bigcap_n \bigcup_{k \geq n} A_k\right) = \lim_{n \rightarrow \infty} P\left(\bigcup_{k \geq n} A_k\right).$$

Or, pour tout n , en utilisant le lemme (pour la suite $(A_k^c)_n$),

$$P\left(\bigcup_{k \geq n} A_k\right) = 1 - P\left(\bigcap_{k \geq n} (A_k)^c\right) = 1 - \prod_{k \geq n} P((A_k)^c) = 1 - \prod_{k \geq n} (1 - P(A_k))$$

et, par l'inégalité $1 - x \leq e^{-x}$, on a

$$0 \leq \prod_{k \geq n} (1 - P(A_k)) \leq \prod_{k \geq n} e^{-P(A_k)} = e^{-\sum_{k \geq n} P(A_k)} = 0$$

car la série diverge, d'où $P(\bigcup_{k \geq n} A_k) = 1$. Et donc $P(\limsup_n A_n) = 1$, comme annoncé. ■

2 Loi d'une suite de variables aléatoires

Si $X_1, X_2, \dots$ sont des variables aléatoires réelles alors, on l'a vu, pour tout $d \geq 1$, $(X_1, X_2, \dots, X_d)$ est une variable aléatoire à valeurs dans $\mathbb{R}^d$ (dont la loi décrit notamment les corrélations possibles entre $X_1, \dots, X_d$). Pour voir $(X_n)_{n \geq 0}$ comme une « suite aléatoire », il faut munir l'espace $E = \mathbb{R}^{\mathbb{N}}$ des suites réelles d'une tribu adaptée.

Définition

La **tribu produit infini**, aussi appelée **tribu cylindrique**, sur $E = \mathbb{R}^{\mathbb{N}}$ est la tribu

$$\mathcal{B} = \mathcal{B}(\mathbb{R})^{\otimes \mathbb{N}} = \sigma(\mathcal{C})$$

où $\mathcal{C}$ est l'ensemble des « cylindres », c'est-à-dire des événements qui ne dépendent que d'un nombre fini de coordonnées :

$$\mathcal{C} = \bigcup_{n \geq 0} \{B_0 \times \dots \times B_n \times \mathbb{R}^{\mathbb{N}} \mid B_0, \dots, B_n \in \mathcal{B}(\mathbb{R})\}.$$

On peut vérifier alors que $(X_n)_{n \geq 0}$ est une variable aléatoire à valeurs dans $(E, \mathcal{B})$ si et seulement si, pour tout $n \geq 0$, X_n est une variable aléatoire réelle.

En conséquence de la définition de $\mathcal{B}$, pour décrire la loi de la suite $(X_n)_{n \geq 0}$, il suffit de donner la loi de $(X_0, \dots, X_n)$ pour tout $n \geq 0$. En particulier, si la suite $(X_n)_{n \geq 0}$ est indépendante, alors (par définition) les variables aléatoires $X_0, \dots, X_n$ sont indépendantes pour tout n , donc la loi du vecteur $(X_0, \dots, X_n)$ est la loi produit des lois de $X_0, \dots, X_n$, de sorte que la loi de la suite $(X_n)_{n \geq 0}$ ne dépend que des lois de X_1 , de X_2 , etc. ; c'est la loi « produit infini » de ces lois.

Notons que l'on n'a pas justifié rigoureusement l'existence de cette loi produit infini, et donc de suites de variables aléatoires indépendantes ayant des lois prescrites. Faisons-le dans le cas particulier important d'une suite de tirages à pile ou face équilibrés (ce cas permettrait d'ailleurs d'obtenir un cas plus général).

Proposition

Soit U une variable aléatoire de loi uniforme sur $[0,1]$. On note $X_1, X_2, \dots$ les chiffres (bits) de son développement en base 2 : $X_i \in \{0,1\}$ et

$$U = \overline{0.X_1X_2\cdots} = \sum_{n=1}^{\infty} \frac{X_n}{2^n} \quad (*)$$

(Autrement dit, $X_n = \lfloor 2^n U \rfloor - 2 \lfloor 2^{n-1} U \rfloor$). Alors les variables aléatoires X_n , $n \geq 1$, sont indépendantes et suivent toutes la loi $\mathcal{B}(1/2)$.

Inversement, pour une telle suite $(X_n)_{n \geq 1}$, la variable aléatoire U définie par $(*)$ suit la loi uniforme sur $[0,1]$.

Démonstration : Soit $N \in \mathbb{N}^*$. Pour tous $\varepsilon_1, \dots, \varepsilon_N \in \{0,1\}$, on a

$$P(X_1 = \varepsilon_1, \dots, X_N = \varepsilon_N) = P(U \in [x, x + 2^{-N}]),$$

où $x = \sum_{n=1}^N \varepsilon_n 2^{-n}$. Cette probabilité vaut donc $\int_{[x, x+2^{-N}]} \mathbf{1}_{[0,1]}(t) dt = 2^{-N}$. En sommant sur les 2^{N-1} choix possibles de $\varepsilon_1, \dots, \varepsilon_{N-1}$, on obtient, pour $\varepsilon_N \in \{0,1\}$,

$$P(X_N = \varepsilon_N) = 2^{N-1} 2^{-N} = \frac{1}{2},$$

ce qui montre que X_N suit la loi de Bernoulli de paramètre $1/2$. Par le même argument, il en va de même de $X_1, \dots, X_{N-1}$. On a donc, pour tous $\varepsilon_1, \dots, \varepsilon_N \in \{0,1\}$,

$$P(X_1 = \varepsilon_1, \dots, X_N = \varepsilon_N) = 2^{-N} = P(X_1 = \varepsilon_1) \cdots P(X_N = \varepsilon_N),$$

ce qui montre que $X_1, \dots, X_N$ sont indépendantes. Ceci est vrai pour tout N , d'où la conclusion.

La seconde partie résulte de la première vu que U s'exprime en fonction de la suite $(X_n)_n$ dont on a décrit la loi. ■

Signalons l'abréviation courante suivante : « $(X_n)_{n \geq 0}$ est une suite de v.a. i.i.d. » signifie que $(X_n)_{n \geq 0}$ est une suite de variables aléatoires indépendantes et identiquement distribuées (c'est-à-dire qu'elles suivent toutes la même loi).

3 Convergence d'une suite de variables aléatoires

Soit $(X_n)_{n \geq 0}$ une suite de variables aléatoires réelles, et X une variable aléatoire réelle.

Les définitions suivantes s'étendent aisément au cas de vecteurs aléatoires.

3.1 Convergence en probabilité

Définition

La suite $(X_n)_{n \geq 0}$ **converge vers X en probabilité**, ce que l'on note aussi $X_n \xrightarrow[n]{(p)} X$, si

$$\text{pour tout } \delta > 0, \quad P(|X_n - X| > \delta) \xrightarrow[n]{} 0.$$

Notons que la limite est unique (à égalité presque sûre près) :

Propriété

Si $X_n \xrightarrow[n]{(p)} X$ et $X_n \xrightarrow[n]{(p)} X'$, alors $X = X'$ p.s..

Donnons un exemple fondamental, qui illustre également l'utilisation de l'inégalité de Tchébychev.

Proposition (Loi faible des grands nombres L^2)

Soit $(X_n)_{n \geq 1}$ une suite de variables aléatoires indépendantes et de même loi, de carré intégrable (c'est-à-dire que $E[(X_1)^2] < \infty$). On a

$$\frac{X_1 + \cdots + X_n}{n} \xrightarrow[n]{(p)} E[X_1].$$

NB. Comme les variables aléatoires $X_1, X_2, \dots$, ont même loi, elles ont même espérance : $E[X_1] = E[X_2] = \cdots$, et de même $E[(X_1)^2] = E[(X_2)^2] = \cdots$.

Démonstration : Notons $\bar{X}_n = \frac{1}{n}(X_1 + \dots + X_n)$. Cette variable aléatoire vérifie

$$E[\bar{X}_n] = \frac{1}{n}(E[X_1] + \dots + E[X_n]) = \frac{1}{n}(E[X_1] + \dots + E[X_1]) = E[X_1]$$

et

$$\text{Var}(\bar{X}_n) = \frac{1}{n^2} \text{Var}(X_1 + \dots + X_n) = \frac{1}{n^2} (\text{Var}(X_1) + \dots + \text{Var}(X_n))$$

car $X_1, \dots, X_n$ sont indépendantes,

$$= \frac{1}{n^2} (\text{Var}(X_1) + \dots + \text{Var}(X_1))$$

car $X_1, \dots, X_n$ ont même loi,

$$= \frac{\text{Var}(X_1)}{n}.$$

Soit $\delta > 0$. L'inégalité de Tchébychev fournit donc ici

$$P(|\bar{X}_n - E[X_1]| > \delta) = P(|\bar{X}_n - E[\bar{X}_n]| > \delta) \leq \frac{\text{Var}(\bar{X}_n)}{\delta^2} = \frac{\text{Var}(X_1)}{n\delta^2} \xrightarrow[n]{} 0.$$

C'est le résultat annoncé. ■

Supposons que $(X_n)_{n \geq 0}$ est une suite de v.a. i.i.d. de loi $\mathcal{B}(1/2)$ (c'est-à-dire une suite de tirages à pile ou face indépendants, équilibrés). Alors X_1 est bornée, donc de carré intégrable, et la proposition donne

$$\frac{X_1 + \dots + X_n}{n} \xrightarrow[n]{(p)} E[X_1] = \frac{1}{2}.$$

Or les X_i valent 0 ou 1, donc $X_1 + \dots + X_n$ est aussi le nombre d'indices entre 1 et n pour lesquels $X_i = 1$, de sorte que $\frac{X_1 + \dots + X_n}{n}$ est la proportion de tirages parmi $X_1, \dots, X_n$ qui valent 1 (ou encore le nombre de « piles »). On a donc obtenu que quand le nombre de tirages tend vers $+\infty$, la probabilité que la proportion de piles s'écarte de $1/2$ de plus que δ converge vers 0. C'est une première forme de la loi des grands nombres, dite faible car la convergence a lieu en probabilité. Elle n'assure cependant pas que, pour chaque suite de tirages, on observe une convergence des proportions vers $1/2$; ce serait une « convergence presque sûre ».

3.2 Convergence presque sûre

Ce qui suit n'est pas une définition à proprement parler mais plutôt une redite ; on l'écrit ainsi pour la mettre en parallèle avec les autres modes de convergence.

Définition

La suite $(X_n)_{n \geq 0}$ **converge vers X presque sûrement**, noté $X_n \xrightarrow[n]{} X$ p.s., si, presque sûrement, $(X_n)_{n \geq 0}$ converge vers X , c'est-à-dire si

$$P(X_n \xrightarrow[n]{} X) = 1.$$

Autrement dit $(X_n)_{n \geq 0}$ converge vers X p.s. s'il existe un événement $\Omega' \subset \Omega$ tel que $P(\Omega') = 1$ et, pour toute réalisation $\omega \in \Omega'$, la suite réelle $(X_n(\omega))_{n \geq 0}$ converge vers $X(\omega)$.

Proposition

a) Si $X_n \xrightarrow[n]{} X$ p.s., alors $X_n \xrightarrow[n]{(p)} X$.

b) $X_n \xrightarrow[n]{} X$ p.s. si, et seulement si, pour tout $\delta > 0$, $P(\limsup_n \{|X_n - X| > \delta\}) = 0$. En particulier, si

$$\sum_{n \geq 0} P(|X_n - X| > \delta) < \infty$$

pour tout $\delta > 0$, alors $X_n \xrightarrow[n]{} X$ p.s.

c) Si $X_n \xrightarrow[n]{(p)} X$, alors il existe une sous-suite $(X_{\varphi(k)})_k$ qui converge vers X p.s..

Démonstration : a) On suppose $X_n \xrightarrow[n]{p.s.} X$. Soit $\delta > 0$. Vu la convergence, on a presque sûrement $|X_n - X| \leq \delta$ pour n grand, et donc l'événement contraire, « pour tout n , il existe $k \geq n$ tel que $|X_k - X| > \delta$ », est négligeable :

$$0 = P(\forall n, \exists k \geq n, |X_k - X| > \delta) = \lim_n \downarrow P(\exists k \geq n, |X_k - X| > \delta)$$

Or $P(\exists k \geq n, |X_k - X| > \delta) \geq P(|X_n - X| > \delta)$, donc cette dernière probabilité converge vers 0.

b) On a $X_n \not\xrightarrow[n]{p.s.} X$ si, et seulement si, il existe $N \in \mathbb{N}$ tel que, pour une infinité de n , $|X_n - X| > \frac{1}{N}$. Autrement dit,

$$1 - P(X_n \rightarrow X) = P\left(\bigcup_N \limsup_n \{|X_n - X| > \frac{1}{N}\}\right) = \lim_N \uparrow P(\limsup_n \{|X_n - X| > \frac{1}{N}\}).$$

L'équivalence b) en résulte rapidement. Le cas particulier est une conséquence des lemmes de la section 1.

c) Si $X_n \rightarrow X$ en probabilité, on peut construire la suite $\varphi(k)$ par récurrence de telle sorte que, pour tout k , $\varphi(k+1) > \varphi(k)$ et

$$P\left(|X_{\varphi(k+1)} - X| > \frac{1}{k+1}\right) \leq 2^{-(k+1)}.$$

Alors, pour tout $\delta > 0$,

$$\sum_k P(|X_{\varphi(k)} - X| > \delta) \leq \delta^{-1} + \sum_{k \geq \delta^{-1}} P(|X_{\varphi(k)} - X| > \delta) \leq \delta^{-1} + \sum_k P(|X_{\varphi(k)} - X| > \frac{1}{k}) \leq \delta^{-1} + \sum_k 2^{-k} < \infty,$$

donc le b) s'applique à la suite $(X_{\varphi(k)})_k$, qui converge donc p.s. vers X . ■

Théorème (Loi forte des grands nombres)

Soit $(X_n)_{n \geq 1}$ une suite de variables aléatoires indépendantes et de même loi, intégrables (c'est-à-dire que $E[|X_1|] < \infty$). On a

$$\frac{X_1 + \dots + X_n}{n} \xrightarrow[n]{p.s.} E[X_1]$$

Démonstration : Dans le **cas particulier** où les variables sont dans L^4 (c'est-à-dire $E[(X_1)^4] < \infty$), qui couvre le cas borné (tirages à pile-ou-face, sondages...), on peut donner une preuve très courte qui s'apparente à celle de la loi faible des grands nombres L^2 . Notons $S_n = X_1 + \dots + X_n$. On peut supposer $E[X_1] = 0$ (car ensuite, on peut écrire $X_n = (X_n - E[X_1]) + E[X_1]$ pour s'y ramener). En développant puis en regroupant les termes, et en notant $X = X_1$,

$$\begin{aligned} E[(S_n)^4] &= E\left[\sum_{1 \leq i,j,k,l \leq n} X_i X_j X_k X_l\right] = \sum_{1 \leq i,j,k,l \leq n} E[X_i X_j X_k X_l] \\ &= nE[X^4] + 4n(n-1)E[X]E[X^3] + 6n(n-1)(n-2)E[X]^2E[X^2] + 3n(n-1)E[X^2]^2 + n(n-1)(n-2)(n-3)E[X]^4 \\ &= nE[X^4] + 3n(n-1)E[X^2]^2 \underset{n \rightarrow \infty}{\sim} 3n^2E[X^2]^2 \end{aligned}$$

d'où (par le TCM pour les séries positives), $E\left[\sum_{n \geq 1} \left(\frac{S_n}{n}\right)^4\right] = \sum_{n \geq 1} \frac{E[(S_n)^4]}{n^4} < \infty$, et par conséquent $\sum_n \left(\frac{S_n}{n}\right)^4 < \infty$ p.s., donc le terme général $\frac{(S_n)^4}{n^4}$ tend vers 0 p.s., et en particulier $\frac{S_n}{n} \rightarrow 0$ p.s., ce que l'on voulait démontrer. ■

Pour des tirages à pile ou face, ce résultat traduit l'observation : pour (presque) toute suite de tirages, la proportion de piles converge vers 1/2. Plus généralement, si la probabilité de pile est p , la proportion de piles converge vers p . Ceci justifie l'interprétation de la probabilité $P(A)$ *a posteriori* comme fréquence d'occurrence de l'événement A si on répète l'expérience **de façon indépendante** un grand nombre de fois.

3.3 Convergences L^1 et L^2

En tant que fonctions mesurables, on peut définir pour les variables aléatoires les normes L^1 et L^2 et donc les convergences associées :

Définition

La suite $(X_n)_{n \geq 0}$ de variables aléatoires L^1 **converge vers X dans L^1** si $\|X_n - X\|_1 \xrightarrow[n]{} 0$, c'est-à-dire si

$$E[|X_n - X|] \xrightarrow[n]{} 0.$$

La suite $(X_n)_{n \geq 0}$ de variables aléatoires L^2 **converge vers X dans L^2** si $\|X_n - X\|_2 \xrightarrow[n]{} 0$, c'est-à-dire si

$$E[|X_n - X|^2] \xrightarrow[n]{} 0.$$

L'inégalité de Markov donne la proposition suivante.

Proposition

Si $X_n \xrightarrow[n]{L^p} X$ (où $p = 1$ ou 2), alors $X_n \xrightarrow[n]{(p)} X$.

Démonstration : En effet, pour tout $\delta > 0$,

$$P(|X_n - X| > \delta) = P(|X_n - X|^p > \delta^p) \leq \frac{E[|X_n - X|^p]}{\delta^p} \xrightarrow[n]{} 0. \quad \blacksquare$$

En revanche, aucune implication n'existe en général entre convergence presque sûre et convergence L^1 ou convergence L^2 . On a par contre la propriété suivante qui résulte directement du théorème de convergence dominée :

Proposition

- a) Si $X_n \xrightarrow[n]{} X$ p.s., et si $|X_n| \leq Z$ pour tout n , où $E[Z] < \infty$, alors $X_n \xrightarrow[n]{L^1} X$.
 b) Si $X_n \xrightarrow[n]{} X$ p.s., et si $|X_n| \leq Z$ pour tout n , où $E[Z^2] < \infty$, alors $X_n \xrightarrow[n]{L^2} X$.

Par ailleurs, la convergence L^2 implique la convergence L^1 :

Proposition

Si $X_n \xrightarrow[n]{L^2} X$, alors $X_n \xrightarrow[n]{L^1} X$.

Démonstration : Par l'inégalité de Cauchy-Schwarz pour les variables $|X_n - X|$ et 1 ,

$$\|X_n - X\|_1 = E[|X_n - X|] \leq E[|X_n - X|^2]^{1/2} E[1]^{1/2} = \|X_n - X\|_2,$$

d'où le résultat. \blacksquare

4 Convergence en loi

4.1 Définition et propriétés

Définition

La suite $(X_n)_{n \geq 0}$ **converge vers X en loi**, noté $X_n \xrightarrow[n]{(loi)} X$, si, pour toute fonction continue bornée $\varphi : \mathbb{R} \rightarrow \mathbb{R}$, $E[\varphi(X_n)] \xrightarrow[n]{} E[\varphi(X)]$.

Une remarque essentielle est que la variable aléatoire X pourrait être remplacée par n'importe quelle variable aléatoire ayant la même loi : la variable aléatoire limite n'est pas unique, mais sa loi est unique. C'est pourquoi on peut aussi dire que X_n **converge en loi vers la loi μ** , noté $X_n \xrightarrow[n]{(loi)} \mu$ si $X_n \xrightarrow[n]{(loi)} X$ où X suit la loi μ .

De plus, on note que cette convergence ne dépend que de la loi de X_n pour chaque n , et non pas de la loi jointe de $(X_n)_{n \geq 0}$: il s'agit en fait de la convergence de la loi de X_n vers la loi de X . En cela, cette convergence est relativement différente des précédentes.

On dispose de nombreuses manières souvent plus pratiques de démontrer une convergence en loi.

Théorème

Les assertions suivantes sont équivalentes :

- (i) $(X_n)_n$ converge vers X en loi.
- (ii) pour toute fonction $\varphi : \mathbb{R} \rightarrow \mathbb{R}$ continue bornée, $E[\varphi(X_n)] \xrightarrow[n]{} E[\varphi(X)]$.
- (iii) pour toute fonction $\varphi : \mathbb{R} \rightarrow \mathbb{R}$ continue à support compact, $E[\varphi(X_n)] \xrightarrow[n]{} E[\varphi(X)]$.
- (iv) en tout point $t \in \mathbb{R}$ où la fonction de répartition F_X est continue (c'est-à-dire que $P(X = t) = 0$),

$$F_{X_n}(t) \xrightarrow[n]{} F_X(t).$$

- (v) pour tout réel t , $\Phi_{X_n}(t) \xrightarrow[n]{} \Phi_X(t)$. (L'équivalence de (v) avec (i) est appelée le **théorème de Lévy**)

Le point (iv) s'avère notamment utile quand X_n est définie comme min ou max de v.a. indépendantes.
 Le point (v) s'avère notamment utile quand X_n est définie comme somme de v.a. indépendantes.

La convergence en loi est plus faible que toutes les autres, en effet :

Proposition

- a) Si $X_n \xrightarrow[n]{(p)} X$, alors $X_n \xrightarrow[n]{(loi)} X$.
 b) Dans le cas où X est constante égale à $c \in \mathbb{R}$, $X_n \xrightarrow[n]{(p)} c$ équivaut à $X_n \xrightarrow[n]{(loi)} c$.

On peut signaler des cas particuliers importants :

Proposition

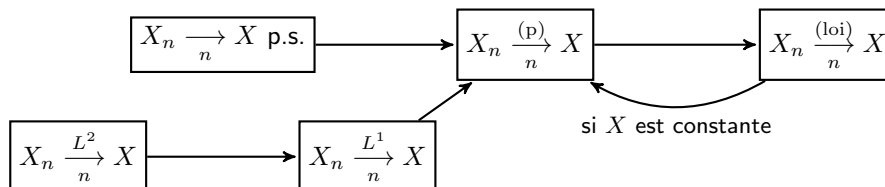
- a) Si les v.a. X_n sont à valeurs dans $\mathbb{Z}$, alors $X_n \xrightarrow[n]{(loi)} X$ si, et seulement si, pour tout $x \in \mathbb{Z}$,
- $$P(X_n = x) \xrightarrow[n]{} P(X = x).$$
- b) Si, pour tout n , la v.a. X_n a une densité f_{X_n} , et si $f_{X_n}(x) \xrightarrow[n]{} f_X(x)$ pour presque tout $x \in \mathbb{R}$, où f_X est la densité d'une variable aléatoire X , alors

$$X_n \xrightarrow[n]{(loi)} X.$$

Attention, convergence en loi n'équivaut pas à convergence des densités !

NB. Par le point a), la propriété citée lors de la définition de la loi de Poisson peut se voir comme une convergence en loi : si X_n suit la loi $\mathcal{B}(n, p_n)$ où $np_n \xrightarrow[n]{} \lambda > 0$, alors $X_n \xrightarrow[n]{(loi)} \mathcal{P}(\lambda)$.

Le schéma suivant résume les implications entre les modes de convergence :



4.2 Théorème central limite

On rappelle que la loi normale $\mathcal{N}(m, \sigma^2)$, de moyenne m et de variance σ^2 , a pour fonction caractéristique

$$\Phi : t \mapsto e^{itm - t^2 \sigma^2 / 2}.$$

En particulier, la loi $\mathcal{N}(0, 1)$ a pour fonction caractéristique $t \mapsto e^{-t^2/2}$.

Théorème (Théorème central limite)

Soit $(X_n)_{n \geq 1}$ une suite de variables aléatoires indépendantes et de même loi, de carré intégrable. On note $m = E[X_1]$ et $\sigma^2 = \text{Var}(X_1)$. On a

$$\frac{X_1 + \cdots + X_n - nm}{\sqrt{n}} \xrightarrow[n]{(\text{loi})} \mathcal{N}(0, \sigma^2).$$

NB. Le cas particulier $\sigma^2 = 0$ correspond au cas où presque sûrement $X_n = m$ (variables “aléatoires” constantes) et est donc sans intérêt (on pose $\mathcal{N}(0, 0) = \delta_0$). On suppose dorénavant $\sigma > 0$. Signalons une autre écriture de cet énoncé :

$$\sqrt{n} \left(\frac{X_1 + \cdots + X_n}{n} - m \right) \xrightarrow[n]{(\text{loi})} \mathcal{N}(0, \sigma^2).$$

De plus, si Z suit la loi $\mathcal{N}(0, 1)$ alors σZ suit la loi $\mathcal{N}(0, \sigma^2)$, de sorte que l’on peut aussi écrire ces énoncés sous la forme

$$\frac{X_1 + \cdots + X_n - nm}{\sqrt{n\sigma^2}} \xrightarrow[n]{(\text{loi})} \mathcal{N}(0, 1) \quad \text{et} \quad \frac{\sqrt{n}}{\sigma} \left(\frac{X_1 + \cdots + X_n}{n} - m \right) \xrightarrow[n]{(\text{loi})} \mathcal{N}(0, 1).$$

Ceci montre que l’« erreur » dans la loi des grands nombres est de l’ordre de $\frac{\sigma}{\sqrt{n}}$ et qu’elle est approximativement distribuée selon une loi normale.

On peut aussi en donner une expression plus élémentaire, à l’aide de l’équivalence (iv) précédente :

$$\text{pour tous } a < b, \quad P \left(a \leq \frac{X_1 + \cdots + X_n - nm}{\sqrt{n\sigma^2}} \leq b \right) \xrightarrow[n]{} \int_a^b e^{-t^2/2} \frac{dt}{\sqrt{2\pi}}$$

(NB : on peut appliquer (iv) en tout $t \in \mathbb{R}$ car la loi limite est continue).

Démonstration : On utilise l’équivalence (v) du théorème. Remarquons d’abord que, quitte à remplacer X_n par $X_n - E[X_n]$, on peut supposer que $m = E[X_n] = 0$. On a, pour tout n ,

$$\begin{aligned} \Phi_{\frac{X_1 + \cdots + X_n}{\sqrt{n}}}(t) &= E \left[e^{it \frac{X_1 + \cdots + X_n}{\sqrt{n}}} \right] = E \left[e^{it \frac{X_1}{\sqrt{n}}} \cdots e^{it \frac{X_n}{\sqrt{n}}} \right] \\ &= E \left[e^{it \frac{X_1}{\sqrt{n}}} \right] \cdots E \left[e^{it \frac{X_n}{\sqrt{n}}} \right] = E \left[e^{it \frac{X_1}{\sqrt{n}}} \right]^n \end{aligned}$$

car $X_1, \dots, X_n$ sont indépendantes et de même loi, d’où

$$\Phi_{\frac{X_1 + \cdots + X_n}{\sqrt{n}}}(t) = \Phi_{X_1} \left(\frac{t}{\sqrt{n}} \right)^n.$$

Rappelons-nous maintenant (voir section 2.3) que, si $E[(X_1)^2] < \infty$, alors Φ_{X_1} a un développement limité à l’ordre 2 en 0 donné par

$$\Phi_{X_1}(t) = 1 + itE[X_1] - \frac{t^2}{2} E[(X_1)^2] + o(t^2) = 1 - \frac{t^2 \sigma^2}{2} + o(t^2)$$

(vu que $E[X_1] = 0$). Comme $\frac{t}{\sqrt{n}} \rightarrow_n 0$, on a donc, pour tout réel t ,

$$\Phi_{X_1} \left(\frac{t}{\sqrt{n}} \right) = 1 - \frac{t^2 \sigma^2}{2n} + o_n \left(\frac{1}{n} \right),$$

et par suite

$$\Phi_{\frac{X_1 + \cdots + X_n}{\sqrt{n}}}(t) = \left(1 - \frac{t^2 \sigma^2}{2n} + o_n \left(\frac{1}{n} \right) \right)^n = \exp \left(n \ln \left(1 - \frac{t^2 \sigma^2}{2n} + o_n \left(\frac{1}{n} \right) \right) \right) = \exp \left(-\frac{t^2 \sigma^2}{2} + o_n(1) \right),$$

autrement dit

$$\Phi_{\frac{X_1 + \cdots + X_n}{\sqrt{n}}}(t) \xrightarrow[n]{} \exp \left(-\frac{t^2 \sigma^2}{2} \right) = \Phi_Z(t),$$

où Z suit la loi $\mathcal{N}(0, \sigma^2)$. Vu l’équivalence (v), il en résulte que $\frac{X_1 + \cdots + X_n}{\sqrt{n}}$ converge en loi vers Z , ce que l’on voulait démontrer. ■