
Intégration & Probabilités

Cours : Quentin Berger, Laurent Tournier
TD : Yueyun Hu

Ce document, ainsi que d'autres documents liés au cours, peut être trouvé sur la page Moodle du cours.



Table des matières

PARTIE I — Intégration —	5
CHAPITRE 1 — Préliminaires	7
1.1 Notations	7
1.1.1 Ensembles	7
1.1.2 Droite réelle achevée $\overline{\mathbb{R}}$	7
1.2 Dénombrabilité	8
1.3 Sommabilité (de réels positifs)	9
1.4 Rappels sur l'intégrabilité de fonctions positives usuelles	10
1.5 Limite inférieure, limite supérieure	11
CHAPITRE 2 — Espace mesurés	13
2.1 Tribus et ensembles mesurables	13
2.2 Mesures	14
2.3 Fonctions mesurables	17
CHAPITRE 3 — Intégration par rapport à une mesure	19
3.1 Intégrale de fonctions mesurables positives	19
3.2 Intégrale de fonctions à valeurs dans $\mathbb{R}$	23
3.3 Intégrale de fonctions à valeurs dans $\mathbb{R}^d$ ou $\mathbb{C}$	25
3.4 Exemples principaux	25
3.4.1 Intégrale par rapport à une mesure discrète	25
3.4.2 Intégrale par rapport à la mesure de Lebesgue (et intégrale de Riemann)	27
3.5 Mesures à densité	29
3.6 Intégrales dépendant d'un paramètre	30
3.7 Espaces de fonctions intégrables	32
3.7.1 Espace $L^1(E, \mathcal{A}, \mu)$	33
3.7.2 Espace $L^2(E, \mathcal{A}, \mu)$	33
CHAPITRE 4 — Intégration sur un espace produit, changement de variables	35
4.1 Produit d'espaces mesurés et mesure produit	35
4.2 Théorèmes de Fubini	36
4.3 Changements de variables	37
4.3.1 Mesure image et formule de transfert	37
4.3.2 Changements de variables dans $\mathbb{R}^d$	38

CHAPITRE 5 — Fondements des probabilités	43
5.1 Espaces de probabilités	43
5.2 Variables aléatoires	44
5.2.1 Variables aléatoires discrètes	44
5.2.2 Variables aléatoires continues (ou à densité).	45
5.3 Espérance	46
5.4 Variance	48
5.5 Vecteurs aléatoires de $\mathbb{R}^d$	49
CHAPITRE 6 — Indépendance	51
6.1 Probabilité conditionnelle et indépendance d'événements	51
6.1.1 Probabilité conditionnelle	51
6.1.2 Événements indépendants	52
6.2 Indépendance de variables aléatoires	52
6.2.1 Indépendance et loi produit	53
6.2.2 Covariance	54
6.3 Propriétés de l'indépendance	55
Récapitulatif de lois importantes	56
CHAPITRE 7 — Caractériser une loi	57
7.1 Fonction de répartition (variables aléatoires réelles)	57
7.2 Fonction génératrice (variables aléatoires entières)	58
7.3 Méthode de la fonction test (variables à densité)	59
7.4 Fonction caractéristique (variables aléatoires à valeurs dans $\mathbb{R}^d$)	60
CHAPITRE 8 — Suites de variables aléatoires	61
8.1 Suites d'événements : deux outils	61
8.1.1 Intersection dénombrable d'événements presque sûrs	61
8.1.2 Lemme de Borel-Cantelli	61
8.2 À propos des suites infinies de variables aléatoires	62
8.3 Convergence d'une suite de variables aléatoires	64
8.3.1 Convergence en probabilité	64
8.3.2 Convergence presque sûre	64
8.3.3 Convergences L^1 et L^2	66
8.4 Convergence en loi	67
8.4.1 Définition et propriétés	67
8.4.2 Théorème central limite	68

En complément, quelques références

- [1] M. Briane, G. Pagès, *Analyse — Théorie de l'intégration*, Vuibert.
- [2] W. Appel, *Probabilités pour les non-probabilistes*, H&K.
- [3] O. Garet, A. Kurtzmann, *De l'intégration aux probabilités*, Ellipses.
- [4] P. Billingsley, *Probability and Measure*, Wiley.

Première partie

— Intégration —

Chapitre 1

Préliminaires

On introduit ici quelques notations pratiques et notamment celle de $\overline{\mathbb{R}}$, et on revoit deux notions qui joueront un rôle central dans la suite : la **dénombrabilité** et la **sommabilité**. On rappelle ensuite quelques intégrales ou sommes usuelles. On donne enfin la définition des limites inférieure et supérieure.

1.1 Notations

1.1.1 Ensembles

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 (la forme "carrée" 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 $\bigcup_{n \in \mathbb{N}} \uparrow A_n$ 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 $\bigcap_{n \in \mathbb{N}} \downarrow A_n$ leur intersection (la flèche rappelle que la suite est décroissante).
- Si A est une partie de E , on note $\mathbf{1}_A : E \rightarrow \{0,1\}$ sa fonction indicatrice :

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

On utilisera parfois une extension de cette notation : $\mathbf{1}_P$ vaudra 1 si la propriété P est vraie, et 0 sinon. Par exemple, pour tous réels x et y , $\mathbf{1}_{(x>0)} = \mathbf{1}_{]0,\infty[}(x)$, ou encore $\mathbf{1}_{(x<y)} = 1$ si $x < y$ et $= 0$ sinon.

1.1.2 Droite réelle achevée $\overline{\mathbb{R}}$

On note $\overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty, +\infty\}$, où $+\infty$ et $-\infty$ sont deux points que l'on ajoute à $\mathbb{R}$.

Ordre. On étend l'ordre de $\mathbb{R}$ à $\overline{\mathbb{R}}$ par $-\infty \leq x \leq +\infty$ pour tout $x \in \mathbb{R}$. Ainsi, $+\infty$ est un majorant de tous les réels, et $-\infty$ est un minorant de tous les réels. De l'existence des **bornes inférieures et supérieures** sur $\mathbb{R}$ on déduit leur existence sur $\overline{\mathbb{R}}$: pour $A \subset \overline{\mathbb{R}}$ non vide, $\sup A$ est le plus petit majorant de A dans $\overline{\mathbb{R}}$, et $\inf A$ est son plus grand minorant. On ajoutera la convention $\sup \emptyset = -\infty$ et $\inf \emptyset = +\infty$.

Disposant d'un ordre sur $\overline{\mathbb{R}}$, on peut considérer, pour $x \in \mathbb{R}$, les **intervalles**

$$[x, +\infty] =]x, +\infty[\cup \{+\infty\} = \{y \in \overline{\mathbb{R}} \mid y \geq x\},$$

$$[-\infty, x] =]-\infty, x[\cup \{-\infty\} = \{y \in \overline{\mathbb{R}} \mid y \leq x\},$$

de même pour $]x, +\infty[$ et $] -\infty, x[$. On a aussi $[-\infty, +\infty] = \overline{\mathbb{R}}$, et on constate que cela est cohérent avec les notations usuelles comme

$$[x, +\infty[= \{y \in \overline{\mathbb{R}} \mid x \leq y < \infty\}, \quad \text{ou} \quad]-\infty, +\infty[= \{y \in \overline{\mathbb{R}} \mid -\infty < y < \infty\} = \mathbb{R}.$$

Toutefois, dans $\overline{\mathbb{R}}$, on appelle **intervalles ouverts** les intervalles de la forme $\{x \in \overline{\mathbb{R}} \mid a < x < b\}$, $\{x \in \overline{\mathbb{R}} \mid x > a\}$ et $\{x \in \overline{\mathbb{R}} \mid x < a\}$, où $a, b \in \overline{\mathbb{R}}$, c'est-à-dire l'intervalle $\mathbb{R}$, les intervalles de la forme $]a, b[$ avec $a, b \in \mathbb{R}$, et les intervalles de la forme $]x, +\infty]$ ou de la forme $[-\infty, x[$.

Limites. Avec cette notion généralisée d'intervalle, on peut généraliser la définition de limite de suite : pour toute suite $(x_n)_{n \geq 0}$ dans $\overline{\mathbb{R}}$, et pour $\ell \in \overline{\mathbb{R}}$, on dit que $(x_n)_n$ converge vers ℓ dans $\overline{\mathbb{R}}$ si : pour tout intervalle ouvert I de $\overline{\mathbb{R}}$ contenant ℓ , il existe un rang n_0 tel que, pour tout $n \geq n_0$, $x_n \in I$.

Si la suite $(x_n)_n$ est à valeurs dans $\mathbb{R}$, cette définition redonne les notions usuelles de limite :

- $x_n \rightarrow \ell \in \mathbb{R}$ si et seulement si, pour tous a, b avec $a < \ell < b$, il existe n_0 tel que, pour $n \geq n_0$, $a < x_n < b$;
- $x_n \rightarrow +\infty$ si et seulement si, pour tout a , il existe n_0 tel que, pour $n \geq n_0$, $x_n > a$;
- $x_n \rightarrow -\infty$ si et seulement si, pour tout a , il existe n_0 tel que, pour $n \geq n_0$, $x_n < a$.
- Si une suite réelle $(x_n)_{n \in \mathbb{N}}$ est croissante, on peut noter $\lim_{\uparrow n} x_n$ sa limite (dans $\overline{\mathbb{R}} = \mathbb{R} \cup \{+\infty, -\infty\}$).
- Si une suite réelle $(x_n)_{n \in \mathbb{N}}$ est décroissante, on peut noter $\lim_{\downarrow n} x_n$ sa limite (dans $\overline{\mathbb{R}} = \mathbb{R} \cup \{+\infty, -\infty\}$).

On peut noter que cette notion de convergence est donnée aussi par la distance

$$d(x, y) = |f(y) - f(x)|,$$

où la fonction $f : \overline{\mathbb{R}} \rightarrow [a, b]$ est une bijection croissante (exemple : $f = \arctan$ étendue à $\overline{\mathbb{R}}$ par $\arctan(-\infty) = -\frac{\pi}{2}$ et $\arctan(+\infty) = \frac{\pi}{2}$).

On peut vérifier que l'espace $\overline{\mathbb{R}}$ est compact : toute suite dans $\overline{\mathbb{R}}$ admet une sous-suite qui converge ! On se rappelle en effet que toute suite réelle bornée (c.-à-d. contenue dans un segment $[-a, a]$) admet une sous-suite convergente (théorème de Bolzano-Weierstrass) et on remarque que toute suite non bornée admet une sous-suite qui diverge vers $+\infty$ ou $-\infty$, donc converge aussi dans $\overline{\mathbb{R}}$.

Opérations. On étend aussi les opérations d'addition et de multiplication à $\overline{\mathbb{R}}$ de façon naturelle, à l'exception de $0 \cdot \pm\infty$ et $(+\infty) + (-\infty)$ qui ne sont pas définies.

1.2 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, c'est-à-dire en faire une liste (éventuellement vide, finie ou 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 $i > n$). On notera $|E| = \text{Card}(E)$ le nombre d'éléments de E .
- $\mathbb{N}$ est évidemment dénombrable : $\mathbb{N} = \{0, 1, 2, 3, \dots\}$.

Ex. 1.1. Montrer que si E est un ensemble dénombrable, alors E est soit fini (ou vide), soit en bijection avec $\mathbb{N}$.

Autres exemples.

- $\mathbb{Z}$ 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 -ème décimale de x soit choisie différente de la n -ème décimale de $\varphi(n)$, et différente de 0 et 9 (cette précision évite que x soit un nombre décimal et ait une écriture non unique : $0,4999\dots = 0,5000\dots$). 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 $n \in \mathbb{N}$, E_n est dénombrable, alors $\bigcup_{n \in \mathbb{N}} E_n$ 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 $x = \xi(k)$. $\square$

Ex. 1.2. Les ensembles suivants sont-ils dénombrables ?

1. $\mathbb{Q}[X]$, ensemble des polynômes à coefficients rationnels

2. $\mathbb{R} \setminus \mathbb{Q}$, ensemble des nombres irrationnels

3. $\{0,1\}^{\mathbb{N}}$, ensemble des suites infinies de 0 et de 1 (*Indication : si on pouvait énumérer toutes ces suites, on pourrait énumérer tous les développements en base 2 des réels dans $[0,1[...$*)

4. $\mathcal{P}(\mathbb{N})$, ensemble des parties de $\mathbb{N}$ (*Indication : associer à chaque partie $A \subset \mathbb{N}$ la suite $(\mathbf{1}_{\{n \in A\}})_{n \in \mathbb{N}} \in \{0,1\}^{\mathbb{N}}$*)

5. $\mathcal{P}_{\text{finie}}(\mathbb{N})$, ensemble des parties finies de $\mathbb{N}$.

1.3 Sommabilité (de réels positifs)

On peut définir la somme de n'importe quelle famille de réels **positifs**, quitte à donner la valeur $+\infty$:

Définition. Pour tout ensemble I (dénombrable ou non), et toute famille $(\alpha_i)_{i \in I}$ de réels positifs, la **somme** de la famille $(\alpha_i)_{i \in I}$ est donnée par

$$\sum_{i \in I} \alpha_i = \sup_{J \subset I, \text{ fini}} \sum_{i \in J} \alpha_i \in [0, +\infty].$$

Lorsque $\sum_{i \in I} \alpha_i < \infty$, on dit que $(\alpha_i)_{i \in I}$ est **sommable**.

Si $I = \mathbb{N}$, et plus généralement si I est dénombrable, on peut vérifier que

$$\sum_{i \in I} \alpha_i = \lim_{n \rightarrow \infty} \sum_{i \in I_n} \alpha_i,$$

où $(I_n)_{n \geq 1}$ est n'importe quelle suite croissante de parties finies de I telle que $\bigcup_n I_n = I$.

L'hypothèse de dénombrabilité de I est en fait essentiellement vérifiée pour toutes les familles sommables :

Proposition 1.1. Soit I un ensemble. S'il existe une famille $(\alpha_i)_{i \in I}$ de réels > 0 telle que $\sum_{i \in I} \alpha_i < \infty$, alors I est dénombrable.

Démonstration. Supposons $S = \sum_{i \in I} \alpha_i < \infty$. Pour tout $n > 0$, l'ensemble $A_n = \{i \in I \mid \alpha_i \geq \frac{S}{n}\}$ contient au plus n éléments (car $S \geq \sum_{i \in A_n} \alpha_i \geq \text{Card}(A_n) \cdot S/n$), donc en particulier est fini, donc dénombrable, or $I = \bigcup_{n \in \mathbb{N}^*} A_n$ si bien que la conclusion vient de la propriété c). $\square$

1.4 Rappels sur l'intégrabilité de fonctions positives usuelles

On utilise ici la définition vue en L1-L2 ou CPGE de l'intégrale de fonctions continues (par morceaux) sur un segment, pour revoir certaines intégrales usuelles et des critères d'intégrabilité. Comme dans le cas des sommes, on suppose les fonctions **positives**, ce qui permet de définir leur intégrale sur un intervalle ouvert même si elle diverge :

Définition. Soit $a \in \mathbb{R}$ et $b \in [a, +\infty]$. Soit $f : [a, b[\rightarrow \mathbb{R}_+$ une fonction continue par morceaux, positive. Alors la fonction $x \mapsto \int_a^x f(t)dt$ est croissante sur $[a, b[$, en raison de la positivité de f , et en particulier, $\lim_{x \rightarrow b^-} \int_a^x f(t)dt$ existe dans $[0, \infty]$: on l'appelle l'**intégrale** de f sur $[a, b[$. Ainsi,

$$\int_{[a, b[} f(t)dt = \lim_{x \uparrow b^-} \int_a^x f(t)dt = \sup_{x \in [a, b[} \int_{[a, x[} f(t)dt \in [0, +\infty].$$

Lorsque $\int_{[a, b[} f(t)dt < \infty$, on dit que f est **intégrable sur** $[a, b[$.

Symétriquement, pour $f :]a, b] \rightarrow \mathbb{R}_+$ continue par morceaux, on définit $\int_{[a, b]} f(t)dt = \lim_{x \downarrow a} \int_x^b f(t)dt$.

Enfin, pour $f :]a, b[\rightarrow \mathbb{R}_+$ continue par morceaux, on définit $\int_{[a, b]} f(t)dt = \int_{[a, c]} f(t)dt + \int_{[c, b]} f(t)dt$ où $c \in]a, b[$ est arbitraire (la somme ne dépend pas du choix de c , par la relation de Chasles). Notons que cette intégrale vaut $+\infty$ dès qu'au moins un des deux termes vaut $+\infty$.

Ex. 1.3. Pour $\alpha \in \mathbb{R}$, la fonction $t \mapsto \frac{1}{t^\alpha}$ est-elle intégrable sur $[1, +\infty[$? Sur $]0, 1]$? Sur $]0, +\infty[$?

Ex. 1.4. Pour $\alpha \in \mathbb{R}$, la famille $(\frac{1}{n^\alpha})_{n \geq 1}$ est-elle sommable? Utiliser une comparaison avec $\int_1^\infty \frac{dt}{t^\alpha}$.

Ex. 1.5. Pour $\lambda \in \mathbb{R}$, la fonction $t \mapsto e^{\lambda t}$ est-elle intégrable sur $[0, +\infty[$?

Ex. 1.6. La fonction $x \mapsto -\ln(x)$ est-elle intégrable sur $]0, 1]$?

Proposition 1.2. Soit f, g deux fonctions continues par morceaux, positives sur $[a, b[$ (où $b \in \mathbb{R} \cup \{+\infty\}$).

a) Si $0 \leq f \leq g$ et g est intégrable sur $[a, b[$, alors f aussi, et $\int_{[a, b[} f(t)dt \leq \int_{[a, b[} g(t)dt$.

b) Si $f(t) \underset{t \rightarrow b^-}{\sim} g(t)$, alors f est intégrable sur $[a, b[$ si et seulement si g est intégrable sur $[a, b[$. De plus,

$$\begin{aligned} \text{si elles ne sont pas intégrables,} \quad & \text{alors} \quad \int_a^x f(t)dt \underset{x \rightarrow b^-}{\sim} \int_a^x g(t)dt \quad (\text{et tend vers } \infty), \\ \text{et si elles sont intégrables,} \quad & \text{alors} \quad \int_{[x, b[} f(t)dt \underset{x \rightarrow b^-}{\sim} \int_{[x, b[} g(t)dt \quad (\text{et tend vers } 0). \end{aligned}$$

Notons que b) est conséquence de a) : si f et g sont équivalentes au voisinage de b à gauche, alors on a un encadrement $cf \leq g \leq Cf$ pour des constantes $c, C > 0$, au voisinage de b à gauche, ce qui permet d'utiliser a). De plus c et C peuvent être choisies aussi proches de 1 que souhaité, avec $0 < c < 1 < C$, d'où les équivalents.

Ex. 1.7. Écrire formellement la démonstration des deux derniers équivalents.

Ex. 1.8. Déterminer si les intégrales suivantes (de fonctions continues, positives) sont finies ou infinies :

$$\int_{]0, \infty[} \frac{dt}{e^t - 1}, \quad \int_{[0, \infty[} \frac{te^{-\sqrt{t}}}{1 + t^2} dt, \quad \int_{]0, 1]} \cos^2 \frac{1}{t} dt, \quad \int_{[0, 1]} \frac{dt}{1 - \sqrt{t}}.$$

1.5 Limite inférieure, limite supérieure

Les limites inférieure et supérieure d'une suite de réels sont des outils pratiques pour étudier le comportement asymptotique d'une suite, convergente ou non.

Définition. Soit $(x_n)_{n \in \mathbb{N}}$ une suite de nombres réels. La **limite inférieure** de la suite $(x_n)_n$ est

$$\liminf_n x_n = \sup_n \left(\inf_{k \geq n} x_k \right) = \lim_{n \rightarrow \infty} \left(\inf_{k \geq n} x_k \right) \in \overline{\mathbb{R}}$$

et la **limite supérieure** de la suite $(x_n)_n$ est

$$\limsup_n x_n = \inf_n \left(\sup_{k \geq n} x_k \right) = \lim_{n \rightarrow \infty} \left(\sup_{k \geq n} x_k \right) \in \overline{\mathbb{R}}.$$

On rappelle qu'une **valeur d'adhérence** d'une suite réelle $(x_n)_n$ est la limite dans $\overline{\mathbb{R}}$ d'une sous-suite convergente $(x_{\varphi(n)})_n$.

Proposition 1.3. Pour toutes suites réelles $(x_n)_n$ et $(y_n)_n$, on a les propriétés suivantes :

a) $\liminf_n x_n$ est la plus petite valeur d'adhérence de la suite $(x_n)_n$: il existe une sous-suite $(x_{\varphi(n)})_n$ qui converge vers $\liminf_n x_n$, et si une sous-suite $(x_{\psi(n)})_n$ admet une limite $\ell \in \overline{\mathbb{R}}$, alors $\liminf_n x_n \leq \ell$.
 $\limsup_n x_n$ est la plus grande valeur d'adhérence de la suite $(x_n)_n$: il existe une sous-suite $(x_{\varphi(n)})_n$ qui converge vers $\limsup_n x_n$, et si une sous-suite $(x_{\psi(n)})_n$ admet une limite $\ell \in \overline{\mathbb{R}}$, alors $\ell \leq \limsup_n x_n$.

b) $\liminf_n x_n \leq \limsup_n x_n$

c) $\limsup_n (-x_n) = -\liminf_n x_n$

d) $(x_n)_n$ converge si et seulement si $\liminf_n x_n = \limsup_n x_n$, et dans ce cas

$$\lim_n x_n = \liminf_n x_n = \limsup_n x_n.$$

e) S'il existe un rang à partir duquel on a $x_n \leq y_n$, alors on peut "passer à la $\liminf$ et $\limsup$ " :

$$\limsup_n x_n \leq \limsup_n y_n \quad \text{et} \quad \liminf_n x_n \leq \liminf_n y_n.$$

f) S'il existe une infinité d'entiers n tels que $x_n \leq y_n$, alors

$$\liminf_n x_n \leq \limsup_n y_n.$$

Démonstration. On montre juste a). On raisonne pour $\liminf$, le cas de $\limsup$ se traitant symétriquement. Si $(x_{\varphi(n)})_n$ est une sous-suite admettant une limite $\ell \in \overline{\mathbb{R}}$ alors pour tout n , on a clairement $x_{\varphi(n)} \geq \inf_{k \geq \varphi(n)} x_k$ d'où, en passant à la limite, $\ell \geq \liminf_n x_n$. Construisons maintenant une sous-suite de $(x_n)_n$ qui converge vers $\liminf_n x_n$.

- Si la suite $(x_n)_n$ n'est pas minorée, alors $\inf_{k \geq n} x_k = -\infty$ pour tout n , donc $\liminf_n x_n = -\infty$; et il existe bien une sous-suite de $(x_n)_n$ qui converge vers $-\infty$: définir $x_{\varphi(0)} = x_0$ puis, pour tout n , on peut par exemple choisir $\varphi(n+1) > \varphi(n)$ tel que $x_{\varphi(n+1)} < -n$.

- Supposons maintenant $(x_n)_n$ minorée, d'où $\inf_{k \geq n} x_k \in \mathbb{R}$ pour tout n . On choisit $\varphi(0) = 0$ puis, pour tout n , on choisit $\varphi(n+1) > \varphi(n)$ tel que $x_{\varphi(n+1)} < \inf_{k \geq \varphi(n)} x_k + 2^{-n}$ (c'est possible par définition de la borne inférieure). Alors, pour tout n , $x_{\varphi(n+1)}$ est encadré entre $\inf_{k \geq \varphi(n)} x_k$ et $\inf_{k \geq \varphi(n)} x_k + 2^{-n}$, et ces deux suites convergent vers $\liminf_n x_n \in]-\infty, +\infty]$, donc $x_{\varphi(n+1)} \rightarrow \liminf_n x_n$. $\square$

Ex. 1.9. Soit $(x_n)_n$ une suite réelle. Montrer que, pour tout réel a , on a $\liminf_n x_n < a$ si et seulement si il existe $\delta > 0$ tel que, pour une infinité d'entiers $n \in \mathbb{N}$, $x_n < a - \delta$; et que $\liminf_n x_n > a$ si et seulement si il existe $\delta > 0$ tel que, pour tout n à partir d'un certain rang, $x_n > a + \delta$.

Chapitre 2

Espace mesurés

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

2.1 Tribus et ensembles mesurables

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

- (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**)

Le couple $(E, \mathcal{A})$ est appelé **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 $A_1, \dots, A_n \in \mathcal{A}$, alors $A_1 \cap \dots \cap A_n \in \mathcal{A}$;

d) 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**)

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 $k > n$. d) résulte de (i) et (iii) car $\bigcup_n A_n = \left(\bigcap_n (A_n)^c \right)^c$. c) résulte de d) en prenant $A_k = E$ pour tout $k > n$. e) résulte de (ii) et d) car $B \setminus A = B \cap A^c$. □

paragraphe*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 soit dénombrable (car on peut écrire $I = \{i_n \mid n \in \mathbb{N}\}$ 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. On vérifie que l'intersection d'une famille de tribus est une tribu (exercice). Notamment, l'intersection de toutes les tribus sur E contenant $\mathcal{C}$ est une tribu, et c'est la plus petite : elle est incluse dans toute tribu contenant $\mathcal{C}$. □

Ex. 2.1. Si $A \subset E$, montrer que la tribu engendrée par $\{A\}$ est $\{\emptyset, A, A^c, E\}$.

Définition (Tribu borélienne). Sur $\mathbb{R}^d$, la **tribu borélienne** est la tribu engendrée par les sous-ensembles ouverts. On la note $\mathcal{B}(\mathbb{R}^d)$. Ses éléments sont les ensembles **boréliens**.

Ex. 2.2. Soit O un ouvert de $\mathbb{R}^d$. Montrer que, pour tout $x \in O$, il existe un pavé fermé $P = [a_1, b_1] \times \cdots \times [a_d, b_d]$ avec $a_1, b_1, \dots, a_d, b_d \in \mathbb{Q}$, inclus dans O , et contenant x . En déduire que O est l'union d'une famille dénombrable de pavés fermés. Conclure que $\mathcal{B}(\mathbb{R}^d)$ est engendré par les pavés fermés de $\mathbb{R}^d$.

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 de $\mathbb{R}$. En pratique, tous les sous-ensembles de $\mathbb{R}$ que l'on manipule sont obtenus ainsi et sont donc boréliens. (Mais attention, il existe quand même des ensembles non boréliens !)

2.2 Mesures

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

Définition (Mesure). 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**, $\mu\left(\bigsqcup_{n \in \mathbb{N}} A_n\right) = \sum_{n \in \mathbb{N}} \mu(A_n)$. (**σ -additivité**)

Le triplet $(E, \mathcal{A}, \mu)$ est appelé **espace mesuré**, et $\mu(E)$ est la **masse totale** de μ .

On dit que μ est **finie** si $\mu(E) < \infty$, et que μ est une **mesure de probabilité** si $\mu(E) = 1$.

Les conséquences suivantes sont aussi importantes que la définition :

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

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) **Continuité croissante ou par l'intérieur :**

Si $(A_n)_n$ est une suite croissante de parties mesurables, alors $\mu\left(\bigcup_n A_n\right) = \lim_{n \rightarrow \infty} \mu(A_n)$.

e) **Continuité décroissante ou par l'extérieur :**

Si $(A_n)_n$ est une suite décroissante de parties mesurables, et $\mu(A_0) < \infty$, alors $\mu\left(\bigcap_n A_n\right) = \lim_{n \rightarrow \infty} \mu(A_n)$.

f) **Sous-additivité :** Pour toute suite $(A_n)_n$ de parties mesurables, $\mu\left(\bigcup_{n \in \mathbb{N}} A_n\right) \leq \sum_{n \in \mathbb{N}} \mu(A_n)$.

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) $\mu(B) = \mu(A) + \mu(B \setminus A) \geq \mu(A)$ 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 , $A_n = C_0 \sqcup \cdots \sqcup C_n$, et donc $\bigcup_n A_n = \bigsqcup_n C_n$. Alors

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

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 $n \geq 1$, $C_n = A_n \cap (A_0 \cup \cdots \cup A_{n-1})^c \in \mathcal{A}$, de sorte que les ensembles C_n sont disjoints et, pour tout n , $A_0 \cup \cdots \cup A_n = C_0 \sqcup \cdots \sqcup C_n$ 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. $\square$

On peut généraliser la propriété c) au cas de trois ensembles mesurables $A, B, C \in \mathcal{A}$. En effet, en appliquant la propriété c) plusieurs fois, on obtient :

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

On peut encore généraliser cette identité au cas de n ensembles mesurables.

Proposition 2.1 (Principe d'inclusion-exclusion). *Pour tous $A_1, \dots, A_n \in \mathcal{A}$, on a*

$$\mu\left(\bigcup_{i=1}^n A_i\right) = \sum_{k=1}^n (-1)^{k+1} \sum_{I \subset \{1, \dots, n\}, |I|=k} \mu\left(\bigcap_{i \in I} A_i\right).$$

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 $\mu = \sum_{n \geq 0} \alpha_n \mu_n$ par¹

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

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 ». Une telle mesure est appelée une **mesure discrète**.

Mesure de Lebesgue sur $\mathbb{R}^d$. Le résultat suivant introduit un exemple fondamental :

Définition-Théorème (Mesure de Lebesgue). *Il existe une unique mesure λ_d sur $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ telle que, pour tout pavé fermé $[a_1, b_1] \times \dots \times [a_d, b_d]$,*

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

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

Démonstration. Un peu longue : voir les références mentionnées en tout début de poly. □

- 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*.

Ex. 2.3. On considère la mesure $\mu := \lambda_1 + \mu_{\mathbb{Z}}$, où $\mu_{\mathbb{Z}}$ est la mesure de comptage sur $\mathbb{Z}$. Déterminer $\mu([a, b])$ pour tous réels $a < b$.

Propriétés. La mesure de Lebesgue λ_d sur $\mathbb{R}^d$ est

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

$$\lambda_d(a + A) = \lambda_d(A), \quad \text{où } a + A = \{a + x \mid x \in A\}.$$

1. Le fait que μ est bien une mesure nécessite un échange de sommations, qui sera justifié plus tard.

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

$$\lambda_d(tA) = |t|^d \lambda_d(A), \quad \text{où } tA = \{tx \mid x \in A\}.$$

Vu l'unicité dans la définition de λ_d , il suffit en effet de vérifier ces propriétés lorsque A est un pavé. Plus généralement, pour montrer que deux mesures sur $\mathbb{R}^d$ sont égales, il suffit de comparer leurs valeurs sur les pavés :

Proposition 2.2. 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 $\mu = \nu$.

Démonstration. Résultat admis. Pour plus d'informations, voir les références mentionnées en début de poly ; l'outil clé est le *lemme de classe monotone*. $\square$

Quelques définitions supplémentaires.

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 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 $\mu(A^c) = 0$, 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 », si la mesure μ n'est pas claire d'après le contexte.

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

Ex. 2.4 – Important. 1. Montrer que, si A et B sont négligeables (pour une mesure μ), alors $A \cup B$ est négligeable. Généraliser ceci à une suite $(A_n)_{n \in \mathbb{N}}$ d'ensembles négligeables.

2. En déduire que si deux propriétés P et Q sont vraies presque partout, alors la propriété « P et Q » est vraie presque partout. De même, montrer que si pour tout $n \in \mathbb{N}$ une propriété P_n est vraie presque partout, alors la propriété « pour tout n , P_n » est vraie presque partout.

Ex. 2.5. Donner un exemple d'une famille (non-dénombrable) $(A_i)_{i \in I}$ d'ensembles négligeables de $\mathbb{R}$ telle que $\bigcup_{i \in I} A_i$ ne l'est pas.

Définition. Soit μ une mesure sur $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$. Le **support** de μ est l'ensemble

$$\text{Supp}(\mu) = \{x \in \mathbb{R}^d \mid \forall \varepsilon > 0, \mu(B(x, \varepsilon)) > 0\}.$$

On peut aussi vérifier que support de μ est le plus petit fermé de $\mathbb{R}^d$ dont le complémentaire est μ -négligeable : c'est le plus petit fermé qui « porte toute la masse » de μ .

Ex. 2.6. Déterminer le support des mesures suivantes sur $\mathbb{R}$:

$$\delta_0, \quad \lambda, \quad \sum_{n \in \mathbb{N}} \delta_n, \quad \sum_{n \in \mathbb{N}^*} \delta_{1/n}.$$

Définition. Soit μ une mesure sur $(E, \mathcal{A})$, et $B \in \mathcal{A}$ un ensemble mesurable. La **restriction de la mesure μ à B** est la mesure $\mu|_B$ sur $(E, \mathcal{A})$ définie par :

$$\text{pour tout } A \in \mathcal{A}, \quad \mu|_B(A) = \mu(A \cap B).$$

La **restriction de la tribu $\mathcal{A}$ à B** est la tribu $\mathcal{A}|_B$ sur B définie par : (de façon équivalente)

$$\mathcal{A}|_B = \{A \cap B \mid A \in \mathcal{A}\} = \{A \in \mathcal{A} \mid A \subset B\}.$$

On vérifie en effet facilement que $\mu|_B$ est une mesure sur $(E, \mathcal{A})$, et que $\mathcal{A}|_B$ est une tribu sur B . Il apparaît donc que $\mu|_B$, restreinte à $\mathcal{A}|_B$, est aussi une mesure sur $(B, \mathcal{A}|_B)$. Ainsi, $(B, \mathcal{A}|_B, \mu|_B)$ est un espace mesuré. Par exemple, on pourra considérer l'espace $([0, 1], \mathcal{B}(\mathbb{R})|_{[0, 1]}, \lambda|_{[0, 1]})$, qui est de masse finie (égale à 1).

Ex. 2.7. 1. Déterminer le support de la mesure $\mu = (\lambda_2)|_D$ sur $\mathbb{R}^2$, où D est le disque unité fermé.

2. A-t-on $\text{Supp}((\lambda_2)|_K) = K$ pour tout fermé K ? (Penser à K tel que $\lambda_2(K) = 0$.)

2.3 Fonctions mesurables

Les démonstrations de cette partie se trouvent dans les différentes références mentionnées en début de poly.

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

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

Proposition 2.3. • Les fonctions indicatrices 1_A où $A \in \mathcal{A}$ sont mesurables.

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

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

Un rappel sur $\liminf$ et $\limsup$ se trouve page 11.

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

Proposition 2.4. Une fonction $f : \mathbb{R}^d \rightarrow \mathbb{R}^{d'}$ est mesurable si et seulement si ses composantes le sont.

Une fonction mesurable de E vers F peut se voir comme une façon de « transporter la masse » de E vers F . Elle définit donc une nouvelle mesure sur F , appelée mesure image.

Définition (Mesure image). Soit $\varphi : (E, \mathcal{A}) \rightarrow (F, \mathcal{B})$ une application mesurable, et μ une mesure sur $(E, \mathcal{A})$. La **mesure image de μ par φ** est la mesure $\nu = \varphi_*\mu$ sur F donnée par :

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

Ex. 2.8. Entre les espaces $E = \{-2, -1, 0, +1\}$ et $F = \mathbb{N}$, munis des tribus discrètes, on considère l'application $\varphi : E \rightarrow F$ définie par $\varphi(n) = n^2$ pour tout $n \in E$. Quelle est l'image par φ de la mesure de comptage sur E ?

Ex. 2.9. 1. Quelle est la mesure image de la mesure de Lebesgue par l'application partie entière ?

2. Quelle est la mesure image de la mesure de Lebesgue par l'application valeur absolue ?

3. Quelle est la mesure image de la mesure de Lebesgue par l'application $x \mapsto x^3$? (On se contentera de calculer la mesure du segment $[a, b]$, pour tous réels $a < b$, ce qui suffit à caractériser la mesure.)

Chapitre 3

Intégration par rapport à une mesure

Dans tout ce chapitre, on se donne un espace mesuré $(E, \mathcal{A}, \mu)$.

On souhaite définir l'intégrale $\int_E f d\mu$ d'une fonction mesurable f sur l'espace E , par rapport à la mesure μ .

- Dans un premier temps, par analogie avec la notion de sommabilité (Chapitre 1), cela sera possible dès que f est à valeurs positives, auquel cas on définit $\int_E f d\mu \in [0, \infty]$, mais on dira que f est **intégrable** par rapport à μ seulement si cette intégrale est finie.
- Dans un second temps, lorsque f prend des valeurs de signe quelconque, ou même à valeurs dans $\mathbb{R}^d$ (ou $\mathbb{C}$), on pourra définir $\int f d\mu$ à condition que $|f|$ (ou $\|f\|$, dans $\mathbb{R}^d$) soit intégrable, à l'aide d'une construction qui se base sur le cas positif.

3.1 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 $n \geq 0$, $\alpha_1, \dots, \alpha_n \in \mathbb{R}^*$ distincts (les valeurs non nulles) et $A_1, \dots, A_n \in \mathcal{A}$ disjoints tels que

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

Si g est à valeurs positives (c.-à-d. si $\alpha_1, \dots, \alpha_n \geq 0$), on définit l'**intégrale de g par rapport à μ** par :

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

En particulier,

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

NB. Une **fonction en escalier** sur $\mathbb{R}$ est une fonction étagée pour laquelle les ensembles $A_1, \dots, A_n$ sont des intervalles. Remarquons qu'il y a des fonctions étagées qui ne sont pas en escalier, par exemple $g = \mathbf{1}_{\mathbb{Q}}$.

Pour le calcul de l'intégrale, on n'a en fait pas besoin de supposer les α_i distincts et les A_i disjoints :

Proposition 3.1. Pour tous $\alpha_1, \dots, \alpha_n \in [0, \infty[$ et $A_1, \dots, A_n \in \mathcal{A}$, la fonction $g = \sum_{i=1}^n \alpha_i \mathbf{1}_{A_i}$ est étagée, et

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

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

Démonstration. Vu sa définition, g est mesurable, et toute valeur de g est une somme d'éléments parmi $\alpha_1, \dots, \alpha_n$, donc il n'y a qu'un nombre fini de valeurs possibles. Ainsi, g est étagée. En définissant, pour tout $I \subset \{1, \dots, n\}$, $A_I = \bigcap_{i \in I} A_i \cap \bigcap_{i \notin I} A_i^c$, les ensembles A_I sont disjoints, et $A_i = \bigcup_{I \ni i} A_I$ pour $i = 1, \dots, n$ donc

on a, en utilisant l'additivité de μ , en intervertissant les sommes, puis en regroupant les indices I selon la valeur (notée s) de $\sum_{i \in I} \alpha_i$,

$$\begin{aligned} \sum_{i=1}^n \alpha_i \mu(A_i) &= \sum_{i=1}^n \alpha_i \sum_{I \ni i} \mu(A_I) = \sum_{I \subset \{1, \dots, n\}} \left(\sum_{i \in I} \alpha_i \right) \mu(A_I) \\ &= \sum_{s \in g(E) \setminus \{0\}} s \sum_{\substack{I \subset \{1, \dots, n\} : \\ \sum_{i \in I} \alpha_i = s}} \mu(A_I) = \sum_{s \in g(E) \setminus \{0\}} s \mu(g^{-1}(\{s\})). \end{aligned}$$

On en déduit le résultat car le membre de droite est par définition $\int g d\mu$. $\square$

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 $af + bh$ prend la valeur $a\alpha_i + b\beta_j$ sur A_{ij} , ce qui permet d'obtenir a) (on pourrait aussi utiliser la proposition précédente); 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) en sommant. $\square$

Définition (Intégrale d'une fonction positive). Soit $f : E \rightarrow [0, +\infty]$ une fonction mesurable positive. L'intégrale de f par rapport à μ est

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

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 $\int_E$.

Comme le montreront les propriétés a) et b) ci-dessous, signalons déjà à titre d'exemple que la fonction $f = \infty \mathbf{1}_A$ vérifie $\int f d\mu = 0$ si $\mu(A) = 0$, et $\int f d\mu = \infty$ si $\mu(A) > 0$. Ainsi, la définition dans le cas étagé s'étend au cas où certains α_i sont $+\infty$, avec $\alpha_i \mu(A_i) = 0$ si $\alpha_i = \infty$ et $\mu(A_i) = 0$, et $\alpha_i \mu(A_i) = \infty$ si $\alpha_i = \infty$ et $\mu(A_i) > 0$.

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). $\square$

Théorème 3.1 (Théorème de convergence monotone (TCM)). Soit $(f_n)_n$ une suite **croissante** de fonctions mesurables **positives**. Alors

$$\int \lim_n^{\uparrow} f_n d\mu = \lim_n^{\uparrow} \int f_n d\mu.$$

Démonstration. On note $f = \lim_n^{\uparrow} f_n$, qui est une fonction mesurable positive.

1^{ère} inégalité. 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 $(\int f_n d\mu)_n$ 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

$$\lim_n^{\uparrow} \int f_n d\mu \leq \int f d\mu.$$

2^{ème} inégalité. Soit h une fonction étagée telle que $0 \leq h \leq f$ et soit $\varepsilon > 0$ arbitraire. Pour tout $x \in E$, si $h(x) > 0$, on a $\lim_{n \uparrow} f_n(x) = f(x) \geq h(x) > (1 - \varepsilon)h(x)$ 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,

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

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 $\lim_{n \uparrow} \mu(A_i \cap E_n) = \mu(A_i)$. En passant à la limite dans l'inégalité précédente, on a donc

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

Comme $\varepsilon > 0$ est arbitraire, ceci donne

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

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

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

ce qui conclut la preuve. □

Par le TCM, pour calculer $\int f d\mu$, on peut considérer $\lim_{n \uparrow} \int f_n d\mu$ pour n'importe quelle suite croissante $(f_n)_n$ de fonctions mesurables positives 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. Définissons, pour tout n , $f_n(x) = \frac{1}{2^n} \lfloor 2^n \min\{f(x), n\} \rfloor$, ce qui revient à poser

$$f_n = \sum_{k=0}^{n2^n-1} \frac{k}{2^n} \mathbf{1}_{f^{-1}([\frac{k}{2^n}, \frac{k+1}{2^n}[})} + n \mathbf{1}_{f^{-1}([n, +\infty[)}.$$

Sur cette dernière écriture, il est manifeste que f_n est étagée car, f étant mesurable, les ensembles $f^{-1}([\frac{k}{2^n}, \frac{k+1}{2^n}[)$ et $f^{-1}([n, +\infty[)$ sont mesurables. De plus, pour tout $x \in E$, pour tout $n > f(x)$, on a $f_n(x) \leq f(x) < f_n(x) + 2^{-n}$, d'où la convergence de f_n vers f . Enfin, le choix d'une subdivision dyadique, c'est-à-dire en intervalles de la forme $[\frac{k}{2^n}, \frac{k+1}{2^n}[$, assure que, pour tout n , $f_n \leq f_{n+1}$. En effet, les réels $\frac{k}{2^n}$ appartenant à l'ensemble qui définit $f_n(x)$ appartiennent aussi à l'ensemble qui définit $f_{n+1}(x)$ car $\frac{k}{2^n} = \frac{2k}{2^{n+1}}$. □

Ceci permet d'étendre certaines propriétés de l'intégrale des fonctions étagées à l'intégrale de fonction mesurables positives quelconques. C'est une méthode illustrée par la propriété suivante, et que l'on retrouvera à de nombreuses reprises.

Propriété (Linéarité). *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. □

Grâce à cette linéarité, le théorème de convergence monotone admet une réécriture en termes de séries :

Corollaire 3.2 (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

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

Or, par la propriété de linéarité précédente,

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

d'où la conclusion. $\square$

Proposition 3.3 (Inégalité de Markov). Pour toute fonction mesurable **positive** f et tout réel $a > 0$,

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

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. $\square$

Corollaire 3.4. 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 $A_n = \{x \in E \mid f(x) \geq n\}$, et $A_\infty = \{x \in E \mid f(x) = \infty\}$. Pour tout n , on a $A_\infty \subset A_n$ donc $0 \leq \mu(A_\infty) \leq \mu(A_n)$, or par l'inégalité de Markov $\mu(A_n) \leq \frac{1}{n} \int f d\mu$, qui tend vers 0 quand $n \rightarrow \infty$ (vu que l'intégrale est supposée finie). On en conclut donc que $\mu(A_\infty) = 0$, ce qu'il fallait démontrer.

b) On suppose $\int f d\mu = 0$. Pour tout $n \geq 1$, on note $A_n = \{x \in E \mid f(x) \geq \frac{1}{n}\}$ et $A_\infty = \{x \in E \mid f(x) > 0\}$. La suite $(A_n)_n$ est croissante, et $A_\infty = \bigcup_n A_n$. On a donc $\mu(A_\infty) = \lim_{n \uparrow} \mu(A_n)$. Or l'inégalité de Markov donne $\mu(A_n) \leq n \int f d\mu = 0$ quel que soit n , d'où $\mu(A_\infty) = 0$, c'est-à-dire que $f = 0$ presque partout (pour μ). 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 $\int h d\mu = 0$. En appliquant la propriété de linéarité à $\max(f, g) = \min(f, g) + h$, 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, donc en particulier $\int f d\mu = \int g d\mu$. $\square$

Théorème 3.2 (Lemme de Fatou). Soit $(f_n)_{n \geq 0}$ une suite de fonction mesurables **positives**. Alors on a

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

Démonstration. On a, par le TCM,

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

Or, pour tout n et $m \geq n$, $\inf_{k \geq n} f_k \leq f_m$, 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,

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

ce qui donne le résultat. $\square$

Contre-exemples. Voici trois exemples de suites $(f_n)_n$ telles que $\lim_{n \rightarrow \infty} f_n = 0$ mais $\lim_{n \rightarrow \infty} \int f_n d\mu \neq 0$ (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** » : $f_n = 2n \mathbf{1}_{[-\frac{1}{n}, \frac{1}{n}]}$

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

3.2 Intégrale de fonctions à valeurs dans $\mathbb{R}$

Définition (Fonction intégrable). Soit $f : E \rightarrow \mathbb{R}$ une fonction mesurable. f est **intégrable par rapport à μ** (ou μ -intégrable) si

$$\int |f| d\mu < \infty.$$

On note $\mathcal{L}^1(E, \mathcal{A}, \mu)$ (ou $\mathcal{L}^1(E, \mu)$, voire $\mathcal{L}^1(E)$ ou même $\mathcal{L}^1$ si le contexte précise $(E, \mathcal{A}, \mu)$) l'espace des fonctions intégrables par rapport à μ .

Pour une fonction $f : E \rightarrow \mathbb{R}$ mesurable, on définit $f_+ = \max(0, f)$ et $f_- = \max(0, -f)$ les parties positive et négative de f . Noter que f_+, f_- sont toutes les deux des fonctions mesurables positives et que $f = f_+ - f_-$ et $|f| = f_+ + f_-$.

Définition (Intégrale d'une fonction de signe quelconque). Si $f \in \mathcal{L}^1(E, \mathcal{A}, \mu)$ est une fonction intégrable par rapport à μ , on définit

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

Noter que l'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.

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 $\int_A |f| d\mu < \infty$.

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) L'ensemble des fonctions intégrables $\mathcal{L}^1(E, \mathcal{A}, \mu)$ est un espace vectoriel, et $f \mapsto \int f d\mu$ est linéaire.

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,

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

Par linéarité de l'intégrale pour les fonctions positives, on a donc $\left| \int f d\mu \right| = \int (f_+ + f_-) d\mu = \int |f| d\mu$.

b) On le laisse en exercice (décomposer en parties positives et négatives).

c) On a $g = f + (g - f)$ 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 le sont aussi), donc $\int f_+ d\mu = \int g_+ d\mu$, de même pour f_- et g_- d'où $\int f d\mu = \int g d\mu$. $\square$

Théorème 3.5 (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) (convergence) La limite $\lim_{n \rightarrow \infty} f_n(x) = f(x)$ existe pour μ -presque tout $x \in E$;

(ii) (domination) Il existe une fonction $\varphi : E \rightarrow \mathbb{R}_+$ mesurable telle que $\int \varphi d\mu < \infty$ et

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

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

$$\lim_{n \rightarrow \infty} \int f_n d\mu = \int f d\mu \quad \text{et} \quad \lim_{n \rightarrow \infty} \int |f_n - f| d\mu = 0.$$

Démonstration. Le fait que les f_n et f soient intégrables vient des inégalités $|f_n| \leq \varphi$ et, à la limite, $|f| \leq \varphi$. Elles sont vraies 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 aussi

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

Ainsi, (*) donne $\limsup_n \int f_n d\mu \leq \int f$, et les mêmes arguments pour $\varphi + f_n$ donnent $\liminf_n \int f_n d\mu \geq \int f d\mu$. Au final,

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

ce qui montre que tous les termes sont égaux, et en particulier $\lim_{n \rightarrow \infty} \int f_n d\mu$ est égal à $\int f d\mu$. $\square$

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 :

Lemme (Approximation d'une intégrale par subdivisions verticales). Soit f une fonction intégrable. Pour toute suite de subdivisions $\ell_1^{(n)} < \dots < \ell_{N(n)}^{(n)}$ de $\mathbb{R}$ telle que $\max_{1 \leq i < N(n)} (\ell_{i+1}^{(n)} - \ell_i^{(n)}) \rightarrow 0$ et $\ell_1^{(n)} \rightarrow -\infty$, $\ell_{N(n)}^{(n)} \rightarrow +\infty$, on a

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

Démonstration. Concentrons-nous sur le cas où f est positive et prenons n grand de sorte que $\ell_1^{(n)} \leq 0$. Il s'agit de montrer que $\int f d\mu = \lim_n \int f_n d\mu$, où on a défini les fonctions étagées

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

Pour tout $x \in E$ on a, pour tout n assez grand on a $f(x) < \ell_{N(n)}^{(n)}$ (par la deuxième condition sur la subdivision) et il existe donc i tel que $f_n(x) = \ell_i^{(n)} \leq f(x) < \ell_{i+1}^{(n)}$ (rappelons que $\ell_1^{(n)} < 0$). On en déduit que

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

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

Si $f(x) < \ell_{N(n)}^{(n)}$, on a par définition $f_n(x) \leq f(x)$. Et si $f(x) \geq \ell_{N(n)}^{(n)}$, alors $f_n(x) = 0 \leq f(x)$. Ainsi, pour tout n , $0 \leq f_n(x) \leq f(x)$. Or on a supposé f intégrable, donc on peut appliquer le TCD à la suite $(f_n)_n$, dominée par $\varphi = f$. $\square$

3.3 Intégrale de fonctions à valeurs dans $\mathbb{R}^d$ ou $\mathbb{C}$

On peut aussi définir l'intégrale de fonctions à valeurs dans $\mathbb{R}^d$ (ou $\mathbb{C}$) : il suffit d'intégrer coordonnée par coordonnée (sous réserve d'intégrabilité). Les résultats des sections précédentes restent alors vrais avec les définitions suivantes (en particulier linéarité et TCD).

Définition (sur $\mathbb{R}^d$). Soit $f : E \rightarrow \mathbb{R}^d$ une fonction mesurable. On dit que f est **intégrable par rapport** μ si $\int \|f\| d\mu < \infty$, où $\|\cdot\|$ est une norme quelconque sur $\mathbb{R}^d$. On pose alors

$$\int f d\mu = \left(\int f_1 d\mu, \dots, \int f_d d\mu \right)$$

où $f_1, \dots, f_d$ sont les composantes de f : pour tout $x \in E$, $f(x) = (f_1(x), \dots, f_d(x))$.

Définition (sur $\mathbb{C}$). Soit $f : E \rightarrow \mathbb{C}$ une fonction mesurable. On dit que f est **intégrable par rapport** μ si $\int |f| d\mu < \infty$, où $|\cdot|$ est le module dans $\mathbb{C}$. On pose alors

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

où $\Re(f)$ et $\Im(f)$ sont les parties réelle et imaginaire de f . Autrement dit, si $\int |f| d\mu < \infty$,

$$\Re\left(\int f d\mu\right) = \int \Re(f) d\mu \quad \text{et} \quad \Im\left(\int f d\mu\right) = \int \Im(f) d\mu.$$

Ex. 3.1. Soit $\lambda > 0$. Calculer $\int_0^\infty e^{-\lambda t} \sin t dt$. On pourra écrire $\sin t = \Im(e^{it})$.

3.4 Exemples principaux

3.4.1 Intégrale par rapport à une mesure discrète

Rappelons la définition de la mesure de Dirac en x , notée δ_x : pour $A \in \mathcal{A}$, $\delta_x(A) = \mathbf{1}_A(x)$.

Proposition 3.6. Soit $f : E \rightarrow \mathbb{R}$ une fonction et soit $x \in \mathbb{R}$. Alors f est intégrable par rapport à δ_x et

$$\int f d\delta_x = f(x).$$

Démonstration. On a $f = f(x)\mathbf{1}_{\{x\}}$ δ_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). □

Proposition 3.7. Soit $(x_n)_{n \geq 0}$ une suite E et $(\alpha_n)_{n \geq 0}$ des réels positifs. On pose $\mu = \sum_{n \geq 0} \alpha_n \delta_{x_n}$.

Si $f : E \rightarrow \mathbb{R}_+$ est une fonction positive, on a

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

Si $f : E \rightarrow \mathbb{R}_+$ est 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

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

Ainsi, si μ_E est la mesure de comptage sur E , c'est-à-dire $\mu_E = \sum_{x \in E} \delta_x$ et si $f : E \rightarrow \mathbb{R}_+$, on a

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

Démonstration. Soit f mesurable positive. Posons $g = f\mathbf{1}_F$ 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 $g = \lim_{\uparrow} g_n$ où $g_n = \sum_{k=0}^n f(x_k)\mathbf{1}_{\{x_k\}}$ donc, par TCM,

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

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. $\square$

Ex. 3.2. En appliquant le TCM, déterminer la limite de la suite de terme général $I_n = \sum_{k=1}^{\infty} \frac{n}{nk^2 + k + 1}$.

Corollaire 3.8 (Échange de sommations). *Soit $(a_{m,n})_{m,n \in \mathbb{N}}$ une famille de réels.*

- a) Si, pour tous $m, n \in \mathbb{N}$, $a_{m,n} \geq 0$, alors $\sum_{m \in \mathbb{N}} \left(\sum_{n \in \mathbb{N}} a_{m,n} \right) = \sum_{n \in \mathbb{N}} \left(\sum_{m \in \mathbb{N}} a_{m,n} \right) \in [0, +\infty]$.
- b) Si $\sum_{m \in \mathbb{N}} \sum_{n \in \mathbb{N}} |a_{m,n}| < \infty$, alors $\sum_{m \in \mathbb{N}} \left(\sum_{n \in \mathbb{N}} a_{m,n} \right) = \sum_{n \in \mathbb{N}} \left(\sum_{m \in \mathbb{N}} a_{m,n} \right) \in \mathbb{R}$.

Démonstration. a) En notant $\mu_{\mathbb{N}}$ la mesure de comptage sur $\mathbb{N}$, on obtient par la proposition précédente

$$\sum_{m \in \mathbb{N}} \left(\sum_{n \in \mathbb{N}} a_{m,n} \right) = \sum_{m \in \mathbb{N}} \int_{\mathbb{N}} a_{m,n} d\mu_{\mathbb{N}}(n) = \int_{\mathbb{N}} \left(\sum_{m \in \mathbb{N}} a_{m,n} \right) d\mu_{\mathbb{N}}(n) = \sum_{n \in \mathbb{N}} \left(\sum_{m \in \mathbb{N}} a_{m,n} \right),$$

où l'égalité du milieu est justifiée par le théorème de convergence monotone pour les séries à termes positifs.

b) Sous l'hypothèse du corollaire, pour tout $n \in \mathbb{N}$, la série $\sum_m a_{m,n}$ converge absolument, et on a

$$\int_{\mathbb{N}} \sum_m |a_{m,n}| d\mu_{\mathbb{N}}(n) = \sum_n \sum_m |a_{m,n}| = \sum_m \sum_n |a_{m,n}| < \infty,$$

où la dernière égalité vient de a). Or la suite $(S_M)_{M \geq 0}$ définie par $S_M(n) = \sum_{m=0}^M a_{m,n}$ converge vers $S(n) = \sum_{m \in \mathbb{N}} a_{m,n}$, et vérifie $|S_M(n)| \leq \sum_{m=0}^M |a_{m,n}| \leq \sum_{m \in \mathbb{N}} |a_{m,n}|$ qui est intégrable par ce qui précède. Donc le théorème de convergence dominée donne

$$\int S_M d\mu_{\mathbb{N}} \xrightarrow{M \rightarrow \infty} \int \sum_{m \in \mathbb{N}} a_{m,n} d\mu_{\mathbb{N}}(n) = \sum_n \sum_m a_{m,n}.$$

Vu que $\int S_M d\mu_{\mathbb{N}} = \sum_n \sum_{m=0}^M a_{m,n} = \sum_{m=0}^M \sum_n a_{m,n}$ (échange avec une somme finie, simple à justifier par récurrence), la série $\sum_m \left(\sum_{n \in \mathbb{N}} a_{m,n} \right)$ converge et a pour somme la limite précédente, ce qui conclut. $\square$

Ex. 3.3. Calculer $\sum_{n=1}^{\infty} \sum_{k=1}^{\infty} \frac{1}{(k+2)^n k!}$.

On peut ainsi enfin prouver qu'une somme infinie de mesures est une mesure, ce qui avait été admis précédemment. Remarquons que la preuve du corollaire précédent n'utilise que la mesure de comptage sur $\mathbb{N}$, dont on sait déjà qu'elle est une mesure, donc l'argument n'est pas circulaire.

Corollaire 3.9 (Somme infinie de mesures). *Soit $(\mu_n)_{n \geq 0}$ une famille de mesures sur $(E, \mathcal{A})$.*

- a) La somme $\mu = \sum_{n \geq 0} \mu_n$ est une mesure ;
- b) Soit $f : E \rightarrow \mathbb{R}$ une fonction mesurable. Si f est positive, alors

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

Pour f de signe quelconque, f est intégrable par rapport à μ si et seulement si $\sum_n \int |f| d\mu_n < \infty$, et dans ce cas

$$\int f d\mu = \sum_{n \geq 0} \int f d\mu_n \in \mathbb{R}.$$

3.4.2 Intégrale par rapport à la mesure de Lebesgue (et intégrale de Riemann)

On note $\lambda = \lambda_1$ la mesure de Lebesgue sur $\mathbb{R}$.

Rappel. Soit $a < b$. 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 $\varphi \leq f \leq \psi$ et $\int_a^b (\psi(t) - \varphi(t))dt < \varepsilon$, où l'intégrale de la fonction en escalier $\psi - \varphi$ est définie de manière élémentaire (comme pour les fonctions étagées avec la mesure de Lebesgue : si $\varphi = \sum_i \alpha_i \mathbf{1}_{[x_i, x_{i+1}[}$ alors $\int_a^b \varphi(t)dt = \sum_i \alpha_i (x_{i+1} - x_i)$). Et dans ce cas on note

$$\int_a^b f(t)dt = \sup_{\substack{\varphi \text{ en escalier,} \\ \varphi \leq f}} \int_a^b \varphi(t)dt = \inf_{\substack{\psi \text{ en escalier,} \\ f \leq \psi}} \int_a^b \psi(t)dt.$$

Théorème 3.10. Si f est intégrable au sens de Riemann sur $[a, b]$, alors f est intégrable par rapport à la mesure de Lebesgue λ sur $[a, b]$, et

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

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. Les fonctions $g = \sup\{\varphi \text{ en escalier} \mid \varphi \leq f\}$ et $h = \inf\{\psi \text{ en escalier} \mid f \leq \psi\}$ sont mesurables, $g \leq f \leq h$ et $\int (h - g)d\mu = 0$ donc $g = h$ μ -presque partout. Ceci implique que f est mesurable, si on considère que les sous-ensembles des ensembles négligeables sont mesurables (on parle de tribu complétée) ; pour ce cours, on négligera cette subtilité et on prétendra f 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 (par les propriétés de l'intégrale de Riemann) $\int_a^b \varphi(t)dt \leq \int_a^b f(t)dt \leq \int_a^b \psi(t)dt$ 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 $\int_a^b \varphi(t)dt = \int_{[a, b]} \varphi d\lambda$ car φ est en escalier, et de même pour ψ . Ainsi les deux intégrales de f appartiennent au même intervalle de largeur $< \varepsilon$, donc

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

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

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

$$\int_I f = \int_I f(t)dt = \int_I f d\lambda,$$

même si f n'est pas intégrable au sens de Riemann, sans confusion possible. On utilisera aussi la notation usuelle suivante : si $a < b$, et si $f : [a, b] \rightarrow \mathbb{R}$ est mesurable positive, ou intégrable par rapport à λ ,

$$\int_{[a, b]} f(t)dt = \int_{]a, b[} f(t)dt = \int_a^b f(t)dt = - \int_b^a f(t)dt.$$

On évitera par contre en général d'utiliser cette notation pour des intégrales par rapport à une mesure quelconque, afin d'éviter des ambiguïtés (voir notamment l'exercice suivant).

Ex. 3.4. Soit $\mu = \lambda + \mu_{\mathbb{Z}}$, où $\mu_{\mathbb{Z}}$ est la mesure de comptage sur $\mathbb{Z}$. Que vaut $\int_{[0, 1]} f d\mu$? Et $\int_{]0, 1[} f d\mu$?

On peut donc, pour des intégrales au sens de Riemann, appliquer les théorèmes qu'on a vu jusqu'ici (convergence monotone, dominée, Fatou, etc.), en plus des propriétés déjà connues, telles les suivantes, que l'on redémontre pour rappel.

Proposition 3.11 (Théorème fondamental de l'analyse).

a) Soit $f : [a, b] \rightarrow \mathbb{R}$ une fonction continue. La fonction $F : x \mapsto \int_a^x f(t)dt$ est dérivable sur $[a, b]$, et sa dérivée est $F' = f$. C'est l'unique fonction dérivable sur $[a, b]$ nulle en a et dont la dérivée est f .

b) Soit $F : [a, b] \rightarrow \mathbb{R}$ une fonction de classe $\mathcal{C}^1$. On a

$$\int_a^b F'(t) dt = F(b) - F(a).$$

Notons que la dernière formule reste valide si $a > b$ avec la convention rappelée plus haut : $\int_b^a = -\int_a^b$.

Démonstration. a) Soit $x \in [a, b]$. Pour tout $h > 0$ tel que $x+h \in [a, b]$, on a $F(x+h) - F(x) = \int_a^{x+h} f(y) dy - \int_a^x f(y) dy = \int_x^{x+h} f(y) dy$ et aussi $f(x) = \frac{1}{h} \int_x^{x+h} f(x) dy$, d'où

$$\left| \frac{F(x+h) - F(x)}{h} - f(x) \right| = \frac{1}{h} \left| \int_x^{x+h} (f(y) - f(x)) dy \right| \leq \frac{1}{h} \int_x^{x+h} |f(y) - f(x)| dy.$$

Soit $\varepsilon > 0$. Par continuité de f en x , il existe $\delta > 0$ tel que pour tout $y \in [x, x+\delta]$, $|f(y) - f(x)| < \varepsilon$. Ainsi, pour tout $0 < h < \delta$,

$$\left| \frac{F(x+h) - F(x)}{h} - f(x) \right| \leq \frac{1}{h} \int_x^{x+h} \varepsilon dy = \varepsilon.$$

Ceci montre, si $x \neq b$, que F est dérivable à droite en x , de dérivée $f(x)$. On procède de même pour la dérivée à gauche en partant de $F(x) - F(x-h) = \int_{x-h}^x f$.

On a bien sûr $F(a) = 0$ et on vient de montrer $F' = f$. Montrons l'unicité. Soient F, G deux telles fonctions et posons $H = G - F$. On a donc $H(a) = 0$ et $H' = 0$. L'inégalité des accroissements finis implique alors que, pour tout $x \in [a, b]$, $|H(x)| = |H(x) - H(a)| \leq |x-a| \max_{[a,x]} |H'| = 0$, donc $H = 0$ sur $[a, b]$, c'est-à-dire $G = F$.

b) Par l'unicité prouvée en a), on a nécessairement $F(x) - F(a) = \int_a^x F'(t) dt$ pour tout $x \in [a, b]$ (les membres de gauche et de droite définissent des fonctions nulles en a et de dérivée F'). En particulier, le cas $x = b$ donne la formule annoncée. $\square$

Corollaire 3.12 (Intégration par parties (IPP) et changement de variable).

a) (Intégration par parties) Pour toutes fonctions $u, v : [a, b] \rightarrow \mathbb{R}$ de classe $\mathcal{C}^1$,

$$\int_a^b u'(t)v(t) dt = u(b)v(b) - u(a)v(a) - \int_a^b u(t)v'(t) dt.$$

b) (Changement de variable sur $\mathbb{R}$) Pour toute fonction continue $f : I \rightarrow \mathbb{R}$, et toute fonction $\varphi : [a, b] \rightarrow I$ de classe $\mathcal{C}^1$,

$$\int_{\varphi(a)}^{\varphi(b)} f(x) dx = \int_a^b f(\varphi(y)) \varphi'(y) dy.$$

Démonstration. a) Par le théorème fondamental de l'analyse appliqué à $F = uv$,

$$u(b)v(b) - u(a)v(a) = \int_a^b (uv)'(t) dt = \int_a^b (u'(t)v(t) + u(t)v'(t)) dt = \int_a^b u'(t)v(t) dt + \int_a^b u(t)v'(t) dt.$$

b) Notons F une primitive de f , c'est-à-dire que $F' = f$ (l'existence de F résulte de la proposition précédente). Par dérivation des fonctions composées, on a $(F \circ \varphi)'(y) = \varphi'(y)F'(\varphi(y)) = \varphi'(y)f(\varphi(y))$, d'où par la proposition précédente pour les fonctions F puis $F \circ \varphi$: (remarquons que l'on peut avoir $\varphi(a) > \varphi(b)$)

$$\int_{\varphi(a)}^{\varphi(b)} f(t) dt = F(\varphi(b)) - F(\varphi(a)) = \int_a^b (F \circ \varphi)'(y) dy = \int_a^b f(\varphi(y)) \varphi'(y) dy. \quad \square$$

Rappelons que l'on définit aussi l'intégrale de Riemann sur un intervalle quelconque par passage à la limite dans les bornes d'intégration, lorsque la limite existe. Si la fonction est intégrable alors on retrouve aussi l'intégrale de Lebesgue. On écrit seulement l'énoncé pour les intervalles $[a, +\infty[$ mais le résultat est général :

Proposition 3.13. Soit $f : [a, +\infty[\rightarrow \mathbb{R}$ une fonction mesurable et μ une mesure sur $\mathbb{R}$.

Si f est positive,

$$\int_{[a, +\infty[} f d\mu = \lim_{b \rightarrow +\infty} \int_{[a, b]} f d\mu.$$

Si f est intégrable par rapport à μ , alors la limite précédente reste vraie.

Démonstration. Dans le cas positif, c'est une conséquence du TCM : on a $\int_{[a,b]} f d\mu = \int f \mathbf{1}_{[a,b]} d\mu$, or $b \mapsto \int f \mathbf{1}_{[a,b]} d\mu$ est croissante et converge vers $\int f \mathbf{1}_{[a,+\infty[} d\mu$ en $+\infty$. Pour f intégrable, c'est une conséquence du TCD avec la domination $|f \mathbf{1}_{[a,b]}| \leq |f| \mathbf{1}_{[a,+\infty[}$, intégrable par hypothèse. $\square$

Ex. 3.5. Dire si les fonctions suivantes sont intégrables sur $]0, +\infty[$, $]1, +\infty[$ ou $]0,1[$ par rapport à la mesure de Lebesgue : $t \mapsto 1$, $t \mapsto e^{at}$ (où $a \in \mathbb{R}$), $t \mapsto t^{-\alpha}$ (où $\alpha \in \mathbb{R}$), $t \mapsto \ln t$, $t \mapsto \sin t$, $t \mapsto \frac{1}{t(\ln t)^\alpha}$.

Ex. 3.6. Déterminer les limites $\lim_{n \rightarrow \infty} \int_0^\infty \frac{e^{-t/n}}{1+t} dt$ et $\lim_{n \rightarrow \infty} \int_0^\infty \frac{n \sin(\frac{t}{n})}{t(1+t^2)} dt$.

3.5 Mesures à densité

On vérifie facilement avec le TCM que, si f est positive, l'application $\nu : A \mapsto \int_A f d\mu = \int f \mathbf{1}_A d\mu$ est une mesure, d'où la définition :

Définition (Mesure à densité). Soit f une fonction mesurable $E \rightarrow [0, +\infty]$ et μ une mesure sur E . On définit la mesure $\nu := f \cdot \mu$ (on notera aussi $d\nu = f d\mu$) en posant

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

Cette mesure $\nu = f \cdot \mu$ est appelée **mesure de densité f par rapport à μ** est la mesure $f \cdot \mu$.

Ex. 3.7. Soit $(\alpha_n)_{n \geq 0}$ une suite de réels positifs, et $(x_n)_{n \in \mathbb{N}}$ une suite de points de E . On considère la mesure discrète $\mu = \sum_{n=0}^\infty \alpha_n \delta_{x_n}$. Montrer que, pour toute fonction $f : E \rightarrow [0, \infty]$, la mesure $f \cdot \mu$ est la mesure discrète sur E donnée par $f \cdot \mu = \sum_{n=0}^\infty f(x_n) \alpha_n \delta_{x_n}$.

Ainsi, toute mesure à densité par rapport à une mesure discrète reste discrète ; par contre, la définition de mesures à densité par rapport à la mesure de Lebesgue fournit de nombreuses nouvelles mesures intéressantes.

Ex. 3.8. On considère la fonction $f : x \mapsto x^2 \mathbf{1}_{[-1,1]}(x)$ sur $\mathbb{R}$, et la mesure $\nu = f \cdot \lambda$. Calculer $\nu(]-\infty, x])$ pour tout $x \in \mathbb{R}$. Que vaut $\nu(\{x\})$, pour tout $x \in \mathbb{R}$?

La formule suivante donne une intuition de ce qu'est la densité par rapport à la mesure de Lebesgue λ_d (dans le cas où la densité est régulière).

Ex. 3.9. Soit f une fonction continue sur $\mathbb{R}^d$, à valeurs dans $\mathbb{R}_+$ et soit $\nu = f \cdot \lambda_d$. Montrer que

$$\text{pour tout } x \in \mathbb{R}^d, \quad f(x) = \lim_{\varepsilon \rightarrow 0^+} \frac{\mu(B(x, \varepsilon))}{\lambda_d(B(x, \varepsilon))},$$

où $B(x, \varepsilon)$ est la boule de rayon ε centrée en x . Ainsi, sur $\mathbb{R}$, on a $f(x) = \lim_{\varepsilon \downarrow 0} \frac{1}{2\varepsilon} \nu(]x - \varepsilon, x + \varepsilon])$.

Proposition 3.14 (Intégrale par rapport à une mesure à densité). Soit f une fonction mesurable $E \rightarrow [0, +\infty]$ et soit μ une mesure sur E . Soit $\nu = f \cdot \mu$ la mesure de densité f par rapport à μ .

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

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

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,

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

Ceci justifie la notation $\nu = f \cdot \mu = f(x) d\mu(x)$ ou $d\nu = f d\mu$. 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).

Ex. 3.10. Pour la mesure ν de l'exercice 3.8, calculer $\int \frac{1}{|x|} d\nu(x)$ et $\int \frac{1}{x} d\nu(x)$; on justifiera que le manque de définition de $x \mapsto \frac{1}{x}$ en 0 n'empêche pas de bien définir son intégrale.

3.6 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 $f : I \times E \rightarrow \mathbb{R}$.

Exemple. On peut interpréter la famille de fonctions $(f(t, \cdot))_{t \in \mathbb{R}_+}$ définies sur $\mathbb{R}$ par $f(t, \cdot) : x \mapsto e^{-tx} \mathbf{1}_{[0, \infty[}(x)$, comme une fonction $f : \mathbb{R}_+ \times \mathbb{R} \rightarrow \mathbb{R}$, c'est-à-dire $f : (t, x) \mapsto e^{-tx} \mathbf{1}_{[0, \infty[}(x)$.

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 paramètre). Le théorème de convergence dominée permet alors d'obtenir facilement des résultats de régularité sur F .

Théorème 3.15 (Convergence dominée/Passage à la limite sous l'intégrale). *Soit $f : (t, x) \mapsto f(t, x)$ une fonction mesurable de $I \times E$ dans $\mathbb{R}^d$ ou $\mathbb{C}$ (pour I un intervalle de $\mathbb{R}$). Soit α un élément de I ou une de ses bornes (dans $\overline{\mathbb{R}}$) et soit $g : x \mapsto g(x)$ une fonction mesurable de E dans $\mathbb{R}^d$ ou $\mathbb{C}$. On suppose que :*

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

$$\text{pour } \mu\text{-presque tout } x \in E, \quad |f(t, x)| \leq \varphi(x).$$

Alors les intégrales suivantes sont bien définies et

$$\lim_{t \rightarrow \alpha} \int f(t, x) d\mu(x) = \int g(x) d\mu(x).$$

Démonstration. L'existence des intégrales résulte de la domination (en passant à la limite, on voit que la limite est dominée par φ elle aussi). Pour prouver l'interversion entre limite et intégrale, on se ramène au TCD à l'aide de la caractérisation des limites par les suites : rappelons que pour une fonction $f : I \rightarrow \mathbb{R}$, $f(t)$ converge vers ℓ quand $t \rightarrow \alpha$ si et seulement si, pour toute suite $(t_n)_n$ dans I qui converge vers α , la suite $(f(t_n))_n$ converge vers ℓ . Soit donc $(t_n)_n$ une suite qui converge vers α . Les hypothèses du théorème permettent d'appliquer le TCD à la suite des fonctions $f_n : x \mapsto f_n(x) = f(t_n, x)$, dominée par φ , pour conclure que

$$\lim_n \int f(t_n, x) d\mu(x) = \int \lim_n f(t_n, x) d\mu(x).$$

D'après la caractérisation des limites par les suites, ceci donne le théorème. □

Ex. 3.11. On définit la fonction $I : t \mapsto I(t) = \int_0^\infty \frac{t \sin(\frac{x}{t})}{x(1+x^2)} dx$ sur $]0, +\infty[$.

Justifier que I est définie sur cet intervalle et déterminer ses limites quand $t \rightarrow \infty$ et $t \rightarrow 0^+$.

Ex. 3.12. On considère la fonction $S : t \mapsto \sum_{n=1}^\infty \frac{1+t}{1+tn^2}$ sur $]0, +\infty[$.

Justifier que S est définie sur cet intervalle et déterminer sa limite quand $t \rightarrow \infty$. Et quand $t \rightarrow 0^+$?

Continuité des intégrales à paramètre

Théorème 3.16 (Continuité sous l'intégrale). *Soit $f : (t, x) \mapsto f(t, x)$ une fonction mesurable de $I \times E$ dans $\mathbb{R}^d$ ou $\mathbb{C}$ (pour I 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 $x \in E$,

$$\text{pour } \mu\text{-presque tout } x \in E, \quad |f(t, x)| \leq \varphi(x).$$

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. Ce théorème est une conséquence directe du théorème précédent. En effet, soit $\alpha \in I$. Les hypothèses du théorème précédent sont vérifiées car pour μ -presque tout $x \in E$, $f(t, x)$ converge vers $f(\alpha, x)$ quand $t \rightarrow \alpha$ d'après la continuité. Le théorème précédent donne donc

$$\lim_{t \rightarrow \alpha} F(t) = \int \left(\lim_{t \rightarrow \alpha} f(t, x) \right) d\mu(x) = \int f(\alpha, x) dx = F(\alpha),$$

ce qui montre que F est continue en α . □

Ex. 3.13. 1. Montrer que la fonction I de l'exercice 3.11 est continue sur $]0, +\infty[$.

2. Montrer que la fonction S de l'exercice 3.12 est continue sur $[a, +\infty[$, pour tout $a > 0$. En déduire que S est continue sur $]0, +\infty[$.

Ex. 3.14. Soit $(a_n)_n$ une suite de réels telle que $\sum_n |a_n| < \infty$. On considère la fonction $F : \mathbb{R} \rightarrow \mathbb{R}$ définie par $F(x) = \sum_{n=0}^{\infty} a_n \sin(nx)$. Démontrer que F est continue sur $\mathbb{R}$.

Ex. 3.15. On considère la fonction $F : x \mapsto F(x) = \int_0^{\infty} \frac{e^{-xt}}{1+t} dt$. Démontrer que F est continue sur $]0, +\infty[$. Peut-on trouver une domination valable pour tout $x > 0$? Commencer par montrer que F est continue sur $[a, +\infty[$ où $a > 0$ est fixé.

Dérivabilité des intégrales à paramètre

Théorème 3.17 (Dérivation sous l'intégrale). Soit $f : (t, x) \mapsto f(t, x)$ une fonction de $I \times E$ dans $\mathbb{R}^d$ ou $\mathbb{C}$ (pour I un intervalle de $\mathbb{R}$). On suppose que :

- (existence de l'intégrale) 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 $t \in I$,

$$\text{pour } \mu\text{-presque tout } x \in E, \quad \left| \frac{\partial f}{\partial t}(t, x) \right| \leq \varphi(x).$$

Alors la fonction $F : t \mapsto F(t) = \int f(t, x) d\mu(x)$ est dérivable sur I et

$$\text{pour tout } t \in I, \quad F'(t) = \int \frac{\partial f}{\partial t}(t, x) d\mu(x).$$

Démonstration. Soit $\alpha \in I$. Il faut montrer que $\lim_{t \rightarrow \alpha} \frac{F(t) - F(\alpha)}{t - \alpha} = \int \frac{\partial f}{\partial t}(t, x) d\mu(x)$. Or par linéarité

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

et l'hypothèse de dérivabilité donne, pour μ -presque tout x ,

$$\lim_{t \rightarrow \alpha} \frac{f(t, x) - f(\alpha, x)}{t - \alpha} = \frac{\partial f}{\partial t}(\alpha, x).$$

La domination nécessaire pour appliquer le TCD (avec paramètre réel) est donnée par l'inégalité des accroissements finis et la troisième hypothèse : pour tout $t \in I$, pour presque tout $x \in E$,

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

et $\int \varphi d\mu < \infty$. On peut donc échanger limite et intégrale et conclure que F est dérivable en α . □

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 .

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 .

Ex. 3.16. 1. Montrer que la fonction F de l'exercice 3.15 est dérivable sur $]0, +\infty[$ et donner une expression de sa dérivée (sous forme d'intégrale).

2. Vérifier que F est solution de l'équation différentielle $y' = y - \frac{1}{x}$.

3. Montrer que F est de classe $\mathcal{C}^\infty$ sur $]0, +\infty[$ et donner une expression de ses dérivées successives.

Donnons un exemple de fonction définie par une intégrale à paramètre qui jouera un rôle dans de nombreux autres cours :

Définition (Transformée de Fourier). Soit $f : \mathbb{R} \rightarrow \mathbb{C}$ une fonction mesurable, intégrable par rapport à la mesure de Lebesgue. La **transformée de Fourier** de f est la fonction $\mathcal{F}f = \widehat{f} : \mathbb{R} \rightarrow \mathbb{C}$ définie par :

$$\text{pour tout } \xi \in \mathbb{R}, \quad \widehat{f}(\xi) = \int f(t)e^{-it\xi} dt.$$

Notons que $\widehat{f}$ est bien définie car $\int |f(t)e^{-it\xi}| dt = \int |f(t)| dt < \infty$, par intégrabilité de f .

Ex. 3.17. Calculer la transformée de Fourier de la fonction $f : \mathbb{R} \rightarrow \mathbb{R}$ définie par $f(x) = e^{-|x|}$.

Ex. 3.18. Montrer les propriétés suivantes :

1. La fonction $\widehat{f}$ est continue sur $\mathbb{R}$;
2. Si $\int t|f(t)| dt < \infty$, alors $\widehat{f}$ est de classe $\mathcal{C}^1$ et $(\widehat{f})'(\xi) = -i\widehat{g}(\xi)$, où $g : t \mapsto tf(t)$;
3. Si f est de classe $\mathcal{C}^1$ et f' est intégrable, alors $(\widehat{f'})(\xi) = i\xi\widehat{f}(\xi)$ pour tout $\xi \in \mathbb{R}$. (Intégrer par parties.)

3.7 Espaces de fonctions intégrables

Soit $(E, \mathcal{A}, \mu)$ un espace mesuré. On rappelle la définition de l'espace (vectoriel) des fonctions intégrables

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

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

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

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

Remarque. On peut aussi considérer le cas “ $p = \infty$ ”, qui correspond à

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

l'espace des fonctions « bornées presque partout » sur E .

On va se concentrer plus bas sur les cas $p = 1$ et $p = 2$, qui sont les plus importants.

Identifier les fonctions égales presque partout. Rappelons qu'avoir $\int |f| d\mu = 0$ implique que $f = 0$ μ -presque partout, mais pas forcément que $f = 0$! (Par exemple, $f(x) = \sin(2\pi x)e^{-|x|}$ est d'intégrale nulle par rapport à la mesure $\mu_{\mathbb{Z}}$ de comptage sur $\mathbb{Z}$.) Pour contourner ce « problème », on peut identifier les fonctions égales μ -presque partout. Ceci mène à la définition suivante :

Définition. On note $f \sim g$ si $f = g$ presque partout. Alors, la relation $f \sim g$ est une relation d'équivalence, et on définit $L^p(E, \mathcal{A}, \mu) = \mathcal{L}^p(E, \mathcal{A}, \mu) / \sim$.

Autrement dit, on obtient $L^p(E, \mathcal{A}, \mu)$ en identifiant entre elles les fonctions égales μ -presque partout (les éléments de $L^p(E, \mathcal{A}, \mu)$ ne sont pas vraiment des fonctions mais des classes de fonctions égales entre elles μ -presque partout). Par exemple, l'élément $0 \in L^p(E, \mathcal{A}, \mu)$ est la classe des fonctions nulles presque partout : si $f = 0$ μ -presque partout, dans L^p on identifiera f à la fonction nulle (même si f n'est pas la fonction nulle).

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 $L^p(E, \mathcal{A}, \mu)$ étaient des fonctions, c'est-à-dire que l'on pourra confondre un élément de L^p (une classe de fonctions) et l'un quelconque de ses représentants (l'une de ces fonctions), à ceci près que si $f, g \in L^p(E, \mathcal{A}, \mu)$ sont égales presque partout, alors $f = g$ par définition.

3.7.1 Espace $L^1(E, \mathcal{A}, \mu)$

Pour $f \in L^1(E, \mathcal{A}, \mu)$, on note

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

On déduit des propriétés de l'intégrale que $\|\cdot\|_1$ est une norme sur $L^1(E, \mathcal{A}, \mu)$. En effet :

- pour tout $\lambda \in \mathbb{R}$, pour tout $f \in L^1(E)$, $\|\lambda f\|_1 = |\lambda| \|f\|_1$;
- pour tous $f, g \in L^1(E)$, $\|f + g\|_1 \leq \|f\|_1 + \|g\|_1$;
- pour tout $f \in L^1(E)$, $\|f\|_1 = 0$ si et seulement si $f = 0 \in L^1(E)$ (noter qu'ici on a utilisé le fait que dans L^1 on a identifié la fonction nulle μ -presque partout à $0 \in L^1$).

Si une suite $(f_n)_n$ dans $L^1(E, \mathcal{A}, \mu)$ converge vers $f \in L^1(E, \mathcal{A}, \mu)$ pour la norme $\|\cdot\|_1$, c'est-à-dire si $\lim_{n \rightarrow \infty} \|f_n - f\|_1 = 0$, on notera $f_n \xrightarrow[n \rightarrow \infty]{L^1(E, \mathcal{A}, \mu)} f$, ou plus simplement $f_n \xrightarrow[n \rightarrow \infty]{L^1} f$.

Théorème 3.18. *L'application $f \mapsto \|f\|_1$ est une norme sur $L^1(E, \mathcal{A}, \mu)$. De plus, muni de cette norme, l'espace vectoriel $L^1(E, \mathcal{A}, \mu)$ est complet.*

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 $\sup_{p \geq 0} \|f_{n+p} - f_n\|_1 \xrightarrow[n \rightarrow \infty]{} 0$, alors il existe $f \in L^1$ telle que $\|f_n - f\|_1 \xrightarrow[n \rightarrow \infty]{} 0$. On rappelle aussi qu'un espace vectoriel normé complet sur $\mathbb{R}$ ou $\mathbb{C}$ est appelé **espace de Banach**.

On donne quelques résultats utiles de densité :

Proposition 3.19. 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 C^∞ à support compact sont denses dans $L^1(\mathbb{R}^d)$.

Ex. 3.19. (Lemme de Lebesgue) Montrer que, pour toute $f \in L^1(\mathbb{R})$,

$$\int f(x) \sin(tx) dx \xrightarrow[t \rightarrow \infty]{} 0.$$

On utilisera au choix la densité dans $L^1(\mathbb{R})$ des fonctions en escalier ou de classe C^∞ à support compact.

3.7.2 Espace $L^2(E, \mathcal{A}, \mu)$

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

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

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 $(|f| - |g|)^2 \geq 0$), et on note

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

qui s'avère être un produit scalaire sur $L^2(E, \mathcal{A}, \mu)$.

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

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

De plus, $\|f\|_2 = 0$ implique $f = 0 \in L^2(E, \mathcal{A}, \mu)$, donc $\langle \cdot, \cdot \rangle$ est définie positive sur $L^2(E, \mathcal{A}, \mu)$; noter qu'on a ici de nouveau utilisé le fait que dans L^2 on a identifié la fonction nulle μ -presque partout à $0 \in L^2$.

Théorème 3.20. *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 $f \mapsto \|f\|_2$. De plus, muni de ce produit scalaire, l'espace vectoriel $L^2(E, \mathcal{A}, \mu)$ est complet.*

On rappelle qu'un espace vectoriel normé complet (espace de Banach) dont la norme découle d'un produit scalaire $\|f\| = \langle f, f \rangle^{1/2}$ est appelé **espace de Hilbert**.

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

$$\langle f, g \rangle^2 = \left(\int fg d\mu \right)^2 \leq \int f^2 d\mu \cdot \int g^2 d\mu = \|f\|_2^2 \cdot \|g\|_2^2.$$

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

Proposition 3.21. 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 3.22. Soit μ une mesure sur E de masse finie, $\mu(E) < \infty$. Alors $f \in L^2(E) \implies f \in L^1(E)$.

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 $g = 1$:

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

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

Ex. 3.20. On munit $\mathbb{R}$ de la mesure de Lebesgue. Donner un exemple de fonction f telle que

1. On a $f \in L^1(\mathbb{R})$ mais $f \notin L^2(\mathbb{R})$.
2. On a $f \in L^2(\mathbb{R})$ mais $f \notin L^1(\mathbb{R})$.

Ex. 3.21. (Espaces ℓ^p) On définit $\ell^p = \mathcal{L}^p(\mathbb{N}, \mathcal{P}(\mathbb{N}), \mu_{\mathbb{N}})$, où $\mu_{\mathbb{N}}$ est la mesure de comptage sur $\mathbb{N}$. C'est donc l'espace des fonctions $a : \mathbb{N} \rightarrow \mathbb{R}$, autrement dit des suites $(a_n)_{n \in \mathbb{N}}$ à valeurs réelles, telles que

$$\sum_{n \in \mathbb{N}} |a_n|^p < \infty.$$

1. Quels sont les ensembles négligeables pour $\mu_{\mathbb{N}}$? En déduire qu'il n'y a pas besoin de passer par l'étape "identification des fonctions égales presque partout", autrement dit que $L^p = \ell^p$.
2. Montrer que $\ell^1 \subset \ell^2$ et, plus généralement, que $\ell^p \subset \ell^q$ si $0 \leq p \leq q$.

Chapitre 4

Intégration sur un espace produit, changement de variables

Dans tout le chapitre, $(E, \mathcal{A}, \mu)$ et $(F, \mathcal{B}, \nu)$ désignent deux espaces mesurés. On supposera que μ et ν sont σ -finies : il existe une suite croissante $(E_n)_n$ de parties mesurables de E telle que

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

et de même pour $(F, \mathcal{B}, \nu)$.

Pour une fonction $f : E \times F \rightarrow \mathbb{R}$, on peut considérer la double intégration $\int_E \left(\int_F f(x, y) d\nu(y) \right) d\mu(x)$. 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 la mesure de Lebesgue sur $\mathbb{R}^2$ à partir de la mesure de Lebesgue sur $\mathbb{R}$ (en partant du fait que l'aire d'un rectangle est le produit des *longueurs* de ses côtés), et qui jouera aussi un rôle important dans la notion d'indépendance en probabilités.

4.1 Produit d'espaces mesurés et mesure produit

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 (Tribu produit). La **tribu produit** de $\mathcal{A}$ et $\mathcal{B}$ est la tribu $\mathcal{A} \otimes \mathcal{B}$ engendrée par les ensembles $A \times B$ (appelés *pavés*) 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 4.1. 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 par récurrence une tribu produit de d tribus.

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 peut vérifier que, si $C \in \mathcal{A} \otimes \mathcal{B}$, alors les tranches sont mesurables (dans les espaces respectifs).

Définition-Théorème (Mesure produit). Il existe une unique mesure m sur $(E \times F, \mathcal{A} \otimes \mathcal{B})$ telle que

$$\forall A \in \mathcal{A}, \forall B \in \mathcal{B}, \quad m(A \times B) = \mu(A)\nu(B),$$

(avec ici $\infty \cdot 0 = 0 \cdot \infty = 0$). La mesure m est appelée **mesure produit** de μ par ν , notée $m = \mu \otimes \nu$.

De plus, pour tout $C \in \mathcal{A} \otimes \mathcal{B}$,

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

L'existence vient de la dernière formule, qui décrit une intégration « par tranches » : la mesure de C est l'intégrale des mesures de ses tranches, horizontales ou verticales. Dans le cas de la mesure de Lebesgue sur $\mathbb{R}^d$, le produit redonne les mesures déjà connues :

Proposition 4.2. *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.

Ex. 4.1. Soit $f : \mathbb{R} \rightarrow \mathbb{R}$ une fonction mesurable positive. Démontrer que l'aire de la partie de $\mathbb{R}^2$ comprise entre l'axe des abscisses et le graphe de f est égale à l'intégrale de f par rapport à la mesure de Lebesgue :

$$\int_{\mathbb{R}} f(x) dx = \lambda_2(\{(x, y) \in \mathbb{R}^2 \mid 0 \leq y \leq f(x)\}).$$

Plus généralement, pour toute fonction $f : E \rightarrow \mathbb{R}_+$ mesurable, justifier que pour toute mesure μ sur E ,

$$\int_E f(x) d\mu(x) = (\mu \otimes \lambda)(\{(x, y) \in \mathbb{R}^2 \mid 0 \leq y \leq f(x)\}).$$

Ex. 4.2. En écrivant $\lambda_3 = \lambda_2 \otimes \lambda_1$, calculer le volume d'un cône plein C , dans $\mathbb{R}^3$, de sommet $O = (0, 0, 0)$ et de base $A \times \{h\}$ où $A \in \mathcal{B}(\mathbb{R}^2)$ et $h > 0$ (h est donc la hauteur du cône) :

$$C = \left\{ (x, y, z) \in \mathbb{R}^3 \mid 0 \leq z \leq h \text{ et } (x, y) \in \frac{z}{h} A \right\}.$$

4.2 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 4.3 (Fubini–Tonelli). *Pour toute fonction mesurable positive $f : E \times F \rightarrow [0, +\infty]$,*

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

Démonstration. La formule est vraie si $f = \mathbf{1}_C$ où $C \in \mathcal{A} \otimes \mathcal{B}$: c'est le théorème précédent. Par suite, la linéarité de l'intégrale montre qu'elle est vraie pour toute fonction étagée positive. Finalement, si f est mesurable positive, la formule s'obtient avec plusieurs applications du TCM en écrivant f comme limite croissante d'une suite $(f_n)_n$ de fonctions étagées positives (sur $E \times F$). L'autre égalité s'obtient de même en échangeant l'ordre d'intégration.

Pour les détails, on renvoie à l'une des références mentionnée en début de poly. $\square$

En conséquence de ce théorème, lorsque f est mesurable positive, on pourra noter

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

c'est-à-dire sans les parenthèses. Ceci justifie la notation $d\mu d\nu$ pour la mesure produit $d(\mu \otimes \nu)$.

Ex. 4.3. On définit la fonction $\Gamma : a \mapsto \int_0^\infty t^{a-1} e^{-t} dt$. Soit $a > 0$. En écrivant que $t^a = \int_0^\infty au^{a-1} \mathbf{1}_{(0 \leq u \leq t)} du$ dans la définition de $\Gamma(a+1)$ (et en le justifiant), utiliser le théorème de Fubini–Tonelli pour démontrer que $\Gamma(a+1) = a\Gamma(a)$.

En appliquant le théorème de Fubini–Tonelli à $(x, y) \mapsto f(x)g(y)$, on a le résultat suivant :

Proposition 4.4 (Produit de mesures à densité). *Soit $f : E \rightarrow \mathbb{R}_+$ et $g : F \rightarrow \mathbb{R}_+$ mesurables. Alors, la mesure produit de la mesure $f \cdot \mu$ de densité f par rapport μ avec la mesure $g \cdot \nu$ de densité g par rapport à ν est la mesure de densité $f \otimes g$ par rapport à $\mu \otimes \nu$, où $(f \otimes g)(x, y) = f(x)g(y)$. Autrement dit*

$$(f \cdot \mu) \otimes (g \cdot \nu) = (f \otimes g) \cdot (\mu \otimes \nu).$$

On peut aussi écrire la conclusion de la proposition sous 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 $f(x)g(y)dxdy$.

On peut aussi traiter le cas d'une fonction f de signe quelconque (comme d'habitude, en écrivant $f = f_+ - f_-$). On obtient le résultat suivant :

Théorème 4.5 (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

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

Remarque. Par le théorème de Fubini–Tonelli, la condition équivaut à

$$\int_E \left(\int_F |f(x, y)| d\nu(y) \right) d\mu(x) < \infty \quad \text{ou} \quad \int_F \left(\int_E |f(x, y)| d\mu(x) \right) d\nu(y) < \infty.$$

Ex. 4.4. 1. Justifier que l'intégrale $F(x) = \int_0^\infty \frac{\sin(t)}{t} e^{-tx} dt$ est bien définie.

2. En écrivant $\frac{1}{t} = \int_0^\infty e^{-tu} du$, mettre $F(x)$ sous la forme d'une intégrale double, et justifier l'interversion des intégrales en utilisant le théorème précédent.

3. En déduire la valeur de $F(x)$. On pourra faire appel à l'Exercice 3.1.

4.3 Changements de variables

4.3.1 Mesure image et formule de transfert

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

Définition (Mesure image). *La **mesure image** de μ par φ est la mesure $\varphi_*\mu$ sur F donnée par :*

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

Théorème 4.6 (Théorème de transfert). *Soit $\varphi : (E, \mathcal{A}) \rightarrow (F, \mathcal{B})$ une application mesurable et μ une mesure sur $(E, \mathcal{A})$.*

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

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

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

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

Ce théorème est une formule de changement de variable : elle énonce que si, dans une intégrale en la variable x , on pose $y = \varphi(x)$, alors on remplace la mesure $d\mu(x)$ par la mesure $d(\varphi_*\mu)(y)$. Cette formule n'est pas utile en pratique si on ne connaît pas la mesure image $\varphi_*\mu$. On va cependant voir que l'on peut la calculer lorsque μ est la mesure de Lebesgue et φ est régulière. De plus, on utilisera très souvent cette formule en probabilités car on supposera justement que la mesure image est connue (ce sera la loi d'une variable aléatoire).

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. $\square$

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

Changement de variable par une application linéaire

Proposition 4.7. 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 $B = [0,1]^d$, cela donne l'interprétation suivante du déterminant de M : c'est le volume du parallélépipède engendré par les vecteurs colonnes de M .

Cette proposition permet de faire des changements de variables linéaires : si $f : \mathbb{R}^d \rightarrow \mathbb{R}$ est positive, ou intégrable par rapport à λ_d , le théorème de transfert donne

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

Ex. 4.5. 1. Montrer que si $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ est positive ou intégrable, $\int_{\mathbb{R}^2} f(x+y, x-y) dx dy = \frac{1}{2} \int_{\mathbb{R}^2} f(s,t) ds dt$.
2. En utilisant ce changement de variable, calculer l'intégrale sur $\mathbb{R}^2$ de la fonction $f(x,y) = (x^2 - y^2)^{-2} \mathbf{1}_{\{x \geq 1+|y|\}}$.

Pour montrer la proposition, on aura besoin dans la suite du résultat suivant, qui est intéressant en soi :

Lemme. Si μ est une mesure sur $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ invariante par translation (c'est-à-dire $\mu(a+A) = \mu(A)$ pour tous $A \in \mathcal{B}(\mathbb{R}^d)$ et $a \in \mathbb{R}^d$) et telle que $\mu([0,1]^d) < \infty$, alors il existe une constante $C \geq 0$ telle que $\mu = C\lambda_d$.

Démonstration. On écrit la preuve pour $\mathbb{R}^2$, mais le principe est général. Par un résultat (admis) du premier chapitre, il suffit de montrer que, pour tout pavé fermé P de $\mathbb{R}^2$, $\mu(P) = C\lambda_2(P)$ ce qui, par l'invariance par translation, se ramène à montrer que, pour tous $a, b > 0$,

$$\mu([0,a] \times [0,b]) = Cab.$$

Posons $C = \mu([0,1]^2)$ (qui est fini par hypothèse). Pour tout $n \geq 1$, le carré semi-ouvert $[0,1[^2$ est l'union disjointe de n^2 carrés semi-ouverts disjoints de côté $\frac{1}{n}$, qui ont tous la même mesure pour μ du fait de l'hypothèse d'invariance par translation, donc cette mesure est $\mu([0, \frac{1}{n}[^2) = \frac{C}{n^2}$. Alors pour tous entiers $i, j \geq 0$, $\mu([0, \frac{i}{n}[\times [0, \frac{j}{n}[) = ij \frac{C}{n^2} = C \frac{i}{n} \frac{j}{n}$, en décomposant de même ce rectangle en ij carrés semi-ouverts disjoints de côté $\frac{1}{n}$.

Soit ensuite $a, b > 0$. Pour tout n , on a l'encadrement $[0, a_n[\times [0, b_n[\subset [0, a] \times [0, b] \subset [0, a_n + \frac{1}{n}] \times [0, b_n + \frac{1}{n}]$, où $a_n = \lfloor \frac{na}{n} \rfloor$ et $b_n = \lfloor \frac{nb}{n} \rfloor$ (approximations de a et b de la forme $\frac{k}{n}$, par en-dessous) d'où, par monotonie des mesures,

$$Ca_nb_n \leq \mu([0,a] \times [0,b]) \leq C\left(a_n + \frac{1}{n}\right)\left(b_n + \frac{1}{n}\right).$$

Cela donne l'égalité annoncée en passant à la limite, car $a_n \rightarrow a$ et $b_n \rightarrow b$. $\square$

Démonstration de la Proposition 4.7. Si M n'est pas inversible, l'image de M est incluse dans un hyperplan H ; or les hyperplans ont une mesure nulle pour λ_d , donc $\lambda_d(M(B)) \leq \lambda_d(H) = 0$ pour tout $B \in \mathcal{B}(\mathbb{R}^d)$. Puisque $\det M = 0$ dans ce cas, la formule est vraie.

Supposons M inversible. Alors $\lambda_M : A \mapsto \lambda_M(A) = \lambda_d(M(A))$ est une mesure sur $\mathbb{R}^d$ (c'est la mesure image $(M^{-1})_*\lambda_d$), qui est invariante par translation (par linéarité de M et invariance par translation de λ_d). De plus,

$\lambda_M([0,1]^d) < \infty$ puisque $M([0,1]^d)$ est borné. En vertu du lemme précédent, λ_M est multiple de la mesure de Lebesgue et il existe donc une constante $C_M \geq 0$ telle que, pour tout $A \in \mathcal{B}(\mathbb{R}^d)$, $\lambda_d(M(A)) = C_M \lambda_d(A)$. Vu que $\lambda_d([0,1]^d) = 1$, on a $C_M = \lambda_M([0,1]^d)$.

Commençons par voir que $C_M = 1$ si M est orthogonale (${}^tMM = I_d$). En effet, dans ce cas M et M^{-1} préservent les distances, donc M préserve les boules centrées en 0 : si B est la boule de rayon 1, alors $M(B) = B$, d'où $\lambda_d(B) = C_M \lambda_d(B)$. Comme $\lambda_d(B) > 0$ on conclut que $C_M = 1$ si M est orthogonale.

Ensuite, montrons que $C_M = \det M$ si M est diagonale. Si les coefficients diagonaux sont $a_1, \dots, a_d$, alors $M([0,1]^d) = [0, a_1] \times [0, a_2] \times \dots \times [0, a_d]$ donc $C_M = \lambda_M([0,1]^d) = |a_1| \dots |a_d| = |\det M|$.

De façon générale, si M est inversible, on va utiliser le fait que, pour toutes applications linéaires M et M' ,

$$C_{MM'} = C_M C_{M'}.$$

En effet, pour $A = [0,1]^d$, $C_{MM'} = \lambda_d(MM'(A)) = \lambda_d(M(M'(A))) = C_M \lambda_d(M'(A)) = C_M C_{M'} \lambda_d(A) = C_M C_{M'}$. Il existe une matrice orthogonale O et une matrice symétrique S telles que $M = OS$ (c'est la décomposition polaire : si S est une racine carrée symétrique de la matrice symétrique définie positive tMM , on vérifie que $O = MS^{-1}$ est orthogonale). Ainsi, on a $C_M = C_O C_S = C_S$: comme $\det M = \det S$, on s'est ramené à montré $C_S = \det S$ dans le cas d'une matrice symétrique.

On peut alors diagonaliser S en base orthonormale : il existe P orthogonale et D diagonale telles que $S = PDP^{-1}$. Ainsi,

$$C_S = C_P C_D C_{P^{-1}} = C_D = \det D,$$

par le cas des matrices orthogonales et diagonales. Comme $\det Sx = \det D$, cela conclut la démonstration. $\square$

Changement de variable par un $\mathcal{C}^1$ -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 $\varphi = (\varphi_i)_{1 \leq i \leq d}$ 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 :

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

D'après 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 $J_\varphi(x) \neq 0$). Puisque le facteur $|J_{\varphi^{-1}}(y)|$ dépend du point y près duquel on considère la mesure, cela suggère que la mesure image est (partout) la mesure à densité $|J_{\varphi^{-1}}(z)| d\lambda_d(z)$.

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 4.8 (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) *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 .)

b) *En particulier, si f est positive, alors*

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

c) *Si f est intégrable sur D , alors $u \mapsto f(\varphi(u)) |J_\varphi(u)|$ est intégrable sur U et la première égalité est vérifiée. De même, si $f \circ \varphi$ est intégrable sur U , alors $y \mapsto f(y) |J_{\varphi^{-1}}(y)|$ est intégrable sur D et la deuxième égalité est vérifiée.*

Ex. 4.6. On rappelle la définition de la fonction $\Gamma : a \mapsto \int_0^\infty x^{a-1} e^{-x} dx$ sur $]0, +\infty[$. Pour $a, b > 0$, écrire le produit $\Gamma(a)\Gamma(b)$ comme une intégrale sur $]0, +\infty[^2$ en les variables x et y , et effectuer le changement de variables $x = uv$, $y = u(1-v)$. En déduire que

$$\Gamma(a)\Gamma(b) = \Gamma(a+b)B(a,b)$$

où $B(a,b)$ est la fonction Bêta d'Euler $B(a,b) = \int_0^1 v^{a-1}(1-v)^{b-1} dv$.

Ex. 4.7. Soit $a, b > 1$. Calculer l'aire de l'ouvert D délimité par les courbes d'équation $y = ax$, $y = x/a$, $y = b/x$ et $y = 1/(bx)$ et contenant le point $(1,1)$. On posera $x = u/v$ et $y = uv$.

Deuxième partie

— Probabilités —

Chapitre 5

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.

5.1 Espaces de probabilités

Définition (Espace de probabilités). Une **espace de probabilités** (ou *espace probabilisé*) 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**.
- L'ensemble 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** (p.s.).

Une interprétation en est la suivante :

- Ω représente l'ensemble de toutes les éventualités possibles dans l'expérience aléatoire considérée (on parle aussi de l'ensemble des *réalisations* possibles du hasard).
- $\mathcal{A}$ 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 par exemple des considérations de symétrie ou un calcul lié aux propriétés physiques de 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).

On pourra appliquer toutes les propriétés des mesures, comme la σ -additivité, la continuité croissante ou décroissante, la sous-additivité, le principe d'inclusion-exclusion : on renvoie à la Section 2.2 (en prenant $\mu = P$).

Remarque. Malgré l'importance théorique de l'espace de probabilités $(\Omega, \mathcal{A}, P)$, 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.

Dans toute la suite, on se donne un espace de probabilités $(\Omega, \mathcal{A}, P)$.

5.2 Variables aléatoires

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

On précisera parfois variable aléatoire à valeurs dans E , et on parle de **variable aléatoire réelle** si l'espace d'arrivée est $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$.

La définition suivante est fondamentale.

Définition (Loi d'une variable aléatoire). 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 (et on abrège parfois $X \sim \mu$) si la loi de X est μ .

Notation pratique. 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 , $\{f(X) \geq 0\} = \{\omega \in \Omega \mid f(X(\omega)) \geq 0\}$.

Proposition 5.1. Si μ est une mesure de probabilité sur un ensemble $(E, \mathcal{E})$, il existe un espace de probabilités $(\Omega, \mathcal{A}, P)$ et une variable aléatoire $X : \Omega \rightarrow E$ dont la loi est μ .

Démonstration. Le plus simple est de choisir $(\Omega, \mathcal{A}, P) = (E, \mathcal{E}, \mu)$, et $X = \text{Id}_E$. C'est ce qu'on appelle l'espace canonique. $\square$

NB. En raison de ce résultat, toute mesure de probabilité peut être vue comme la loi d'une certaine variable aléatoire et on aura donc tendance à utiliser le mot « loi » pour parler d'une mesure de probabilité, sans qu'elle soit nécessairement associée à une variable aléatoire particulière.

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.*

5.2.1 Variables aléatoires discrètes

Définition (v.a. discrète). Une variable aléatoire $X : \Omega \rightarrow E$ est **discrète** s'il existe un ensemble dénombrable $D \subset E$ tel que, presque sûrement, $X \in D$.

Par abus de langage, on dira que X est à valeurs dans D .

Cette définition dépend seulement de la loi de X , et revient à dire que P_X est une mesure discrète. En effet, pour tout $B \in \mathcal{E}$,

$$P_X(B) = P(X \in B) = P(\{X \in B \cap D\} \sqcup \{X \in B \cap D^c\}) = P(X \in B \cap D) + P(X \in B \cap D^c)$$

et $P(X \in B \cap D^c) \leq P(X \in D^c) = 0$ donc

$$P_X(B) = P(X \in B \cap D) = P\left(\bigsqcup_{x \in B \cap D} \{X = x\}\right) = \sum_{x \in B \cap D} P(X = x),$$

c'est-à-dire que

$$P_X = \sum_{x \in D} p_x \delta_x, \quad \text{où } p_x = P(X = x) \text{ pour } x \in D.$$

Remarque. Si X est une variable discrète, se donner la loi de X revient donc à se donner les valeurs $p_x = P(X = x)$ pour $x \in D$, dont la somme vaut 1 :

$$1 = P(X \in D) = \sum_{x \in D} P(X = x) = \sum_{x \in D} p_x.$$

Ex. 5.1. On lance deux dés à six faces. Donner la loi de la somme des deux résultats.

Définition (Lois discrètes usuelles I).

- 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 $|E| = n$, la **loi uniforme sur E** est la loi d'une variable aléatoire X à valeurs dans E telle que

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

C'est un choix naturel dans les situations où les différentes issues de l'expérience jouent des rôles symétriques.

- 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 peut voir X comme le résultat du lancer d'une pièce biaisée ayant probabilité p de tomber sur pile.

5.2.2 Variables aléatoires continues (ou à densité).

Définition (v.a. à densité). Une variable aléatoire X à valeurs dans $\mathbb{R}^d$ est **continue**, ou **à densité**, si sa loi admet une densité f par rapport à la mesure de Lebesgue. La fonction f est appelée la **densité** de X . Autrement dit, X a pour densité f si la fonction f est mesurable positive et vérifie, pour tout $A \in \mathcal{B}(\mathbb{R}^d)$,

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

et donc en particulier $\int f(x) dx = P(X \in \mathbb{R}^d) = 1$.

Ainsi, une fonction mesurable $f : \mathbb{R}^d \rightarrow \mathbb{R}$ est une densité si et seulement si f est positive et $\int f(x) dx = 1$.

Propriétés. Si X est une variable aléatoire de densité f , alors

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

Démonstration. a) On a $\lambda_d(\{x\}) = 0$, donc $f \mathbf{1}_{\{x\}} = 0$ presque partout, d'où

$$P(X = x) = \int_{\{x\}} f(t) dt = \int f(t) \mathbf{1}_{\{x\}}(t) dt = 0.$$

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 .

c) L'ensemble $B = \{x \in \mathbb{R}^d \mid f(x) > g(x)\}$ vérifie $P(X \in B) = \int \mathbf{1}_B(x) f(x) dx = \int \mathbf{1}_B(x) g(x) dx$ d'où

$$0 = \int \mathbf{1}_B(x) (f(x) - g(x)) dx,$$

or la fonction dans l'intégrale est positive, donc elle est nulle presque partout. Comme $\mathbf{1}_B(x) (f(x) - g(x)) > 0$ si $x \in B$, on en déduit que $\lambda_d(B) = 0$. On a donc $f(x) \leq g(x)$ pour presque tout x . De même par symétrie $g(x) \leq f(x)$ pour presque tout x , donc $f = g$ presque partout. $\square$

Définition (Lois à densité usuelles).

- 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.)

• Plus généralement, si $A \in \mathcal{B}(\mathbb{R}^d)$ est tel que $0 < \lambda_d(A) < \infty$, la **loi uniforme sur A** est la loi (notée $\mathcal{U}(A)$) de densité

$$f(x) = \frac{1}{\lambda_d(A)} \mathbf{1}_A(x).$$

(Si X suit la loi uniforme sur A , alors $X \in A$ 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 (voir plus loin).

La loi $\mathcal{N}(0,1)$, appelée loi normale standard, a donc pour densité $\frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$.

Ex. 5.2. Soit $r > 0$. On considère Z une variable aléatoire de loi uniforme dans le disque D de centre $(0,0)$ et de rayon r dans $\mathbb{R}^2$. On note R la distance entre $(0,0)$ et Z .

1. Montrer que, pour tous a, b tels que $0 \leq a \leq b \leq r$, $P(R \in [a, b]) = \int_a^b \frac{2t}{r^2} dt$.

2. En déduire que R suit la loi de densité $f : t \mapsto f(t) = \frac{2t}{r^2} \mathbf{1}_{[0, r]}(t)$.

Ex. 5.3. Soit U une variable aléatoire de loi uniforme sur $[0, 1]$. On pose $X = \max(U, \frac{1}{2})$.

1. Calculer $P(X = \frac{1}{2})$, et en déduire que X n'a pas de densité.

2. Justifier que X n'est pas discrète.

5.3 Espérance

Définition (Espérance). Soit X une variable aléatoire réelle. Son **espérance** est son intégrale :

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

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 a en particulier, si $B \in \mathcal{A}$

si $B \in \mathcal{A}$

$$E[\mathbf{1}_B] = P(B).$$

De même, pour toute constante $c \in \mathbb{R}$, $E[c] = c$.

Remarque. Les variables aléatoires bornées sont intégrables : si $|X| \leq M$ p.s., alors $E[|X|] \leq E[M] = M < \infty$.

Interprétation de l'espérance. L'espérance $E[X]$ est la **moyenne** de la variable aléatoire X : c'est une généralisation d'un barycentre à coefficients positifs car c'est une « somme » (intégrale) des valeurs de X pondérées par la probabilité P , et P attribue des poids positifs et de masse totale égale à 1. Ainsi, l'espérance renseigne sur l'ordre de grandeur des valeurs prises par X .

Une autre interprétation est la suivante : si X est le gain à un jeu de hasard, alors $E[X]$ est le prix que le casino doit faire payer au joueur pour rendre le jeu *équitable* : supposons que le joueur rejoue indéfiniment au même jeu (de façon *indépendante*), alors si le prix à payer était plus grand que $E[X]$, la fortune du joueur tendrait vers $-\infty$ (il serait donc ruiné au bout d'un certain temps), tandis que pour un prix plus petit, sa fortune tendrait vers $+\infty$ (et c'est le casino qui serait ruiné, en imaginant que le joueur a pu emprunter temporairement de quoi couvrir ses pertes éventuelles). *Ceci sera justifié par la loi des grands nombres*. De façon plus courante, l'espérance sert donc aussi à calculer des prix d'actifs financiers.

Un théorème important qui permet de faire la plupart des calculs d'espérance est le théorème de transfert (voir page p.37), qui s'écrit dans le cadre des espaces de probabilités comme suit :

Proposition 5.2 (Théorème de transfert). Soit X une variable aléatoire à valeurs dans $(E, \mathcal{E})$ et φ une fonction mesurable $E \rightarrow \mathbb{R}$ telle que $E[\varphi(X)]$ est bien définie. Alors

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

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 E , on a, par le théorème de transfert, pour toute fonction $\varphi : E \rightarrow \mathbb{R}$ positive (ou telle que $\varphi(X)$ est intégrable),

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

- 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),

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

Ex. 5.4. Étudier la bonne définition des espérances suivantes et les calculer si elles ont un sens :

1. $E[X]$, où X suit la loi uniforme sur $\{1, 2, \dots, 6\}$: $P(X = k) = \frac{1}{6}$ pour $k = 1, \dots, 6$;
2. $E[X]$ et $E[X^2]$ où, pour des réels $a < b$, X suit la loi uniforme sur $[a, b]$, de densité $\frac{1}{b-a} \mathbf{1}_{[a, b]}$;
3. $E[X]$ où X suit la loi de Cauchy, de densité $x \mapsto \frac{1}{\pi(1+x^2)}$;
4. $E[s^X]$ où $s \in \mathbb{R}$ et, pour un réel $\lambda > 0$, X suit la loi de Poisson de paramètre λ : X est à valeurs dans $\mathbb{N}$ et pour tout $n \in \mathbb{N}$, $P(X = n) = e^{-\lambda} \frac{\lambda^n}{n!}$;
5. $E[X]$ où $X = \max(U, \frac{1}{2})$ et U suit la loi uniforme sur $[0, 1]$.

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

Proposition 5.3 (Inégalité de Markov). Soit X une variable aléatoire réelle **positive**. Pour tout $a > 0$,

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

Corollaire 5.4. Soit X une variable aléatoire réelle **positive**.

- si $E[X] < \infty$, alors $X < \infty$ presque sûrement. (Attention, la réciproque est fausse !)
- Si $E[X] = 0$, alors $X = 0$ presque sûrement.

Proposition 5.5 (Convergence monotone). 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 $E\left[\sum_{n \geq 0} X_n\right] = \sum_{n \geq 0} E[X_n]$.

Ex. 5.5. Soit X une variable aléatoire à valeurs dans $[0, +\infty[$. Pour tout $n \in \mathbb{N}$, on définit $X_n = \min(n, X)$. Montrer que $\lim_{n \rightarrow \infty} E[X_n] = E[X]$.

Proposition 5.6 (Convergence dominée). Soit $(X_n)_{n \geq 0}$ une suite de variables aléatoires réelles, et X une variable aléatoire réelle. Si $\lim_{n \rightarrow \infty} X_n = X$ p.s. et s'il existe Z intégrable telle que $|X_n| \leq Z$ p.s., alors

$$\lim_{n \rightarrow \infty} E[X_n] = E[X].$$

Ex. 5.6. Soit X une variable aléatoire à valeurs dans $\mathbb{R}$, intégrable. Pour tout $n \in \mathbb{N}$, on définit $X_n = X \mathbf{1}_{\{|X| < n\}}$. Montrer que $\lim_{n \rightarrow \infty} E[X_n] = E[X]$.

On pourra utiliser les théorèmes de Fubini ou encore les théorèmes pour les intégrales (espérances) à paramètre.

Ex. 5.7. Soit X une variable aléatoire réelle, et posons $F(t) = E[e^{-t|X|}]$ pour $t \in \mathbb{R}_+$.

1. Montrer que F est bien définie, continue et dérivable sur $\mathbb{R}_+$.
2. Déterminer la limite de $F(t)$ quand $t \rightarrow \infty$.

5.4 Variance

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 , les espaces des v.a. réelles respectivement intégrables et de carré intégrable. Rappelons que :

- Si $X, Y \in L^2$ alors $XY \in L^1$ (car $|XY| \leq \frac{1}{2}(X^2 + Y^2)$).
- L'application qui à $X, Y \in L^2$ associe $\langle X, Y \rangle = E[XY]$ définit un produit scalaire. En particulier, l'inégalité de Cauchy-Schwarz s'écrit $E[XY]^2 \leq E[X^2]E[Y^2]$.
- Si $X \in L^2$ alors $X \in L^1$ (grâce à l'inégalité de Cauchy-Schwarz, $E[|X|] \leq E[|X|^2]^{1/2}E[1]^{1/2} = E[X^2]^{1/2}$). Autrement dit, les v.a. de carré intégrable sont intégrables.

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

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

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 le point a) ci-dessous :

Propriétés. Soit X une variable aléatoire réelle, de carré intégrable.

- pour tout réel a , $\text{Var}(aX) = a^2 \text{Var}(X)$ et $\sigma(aX) = |a|\sigma(X)$.
- $\text{Var}(X) = E[X^2] - E[X]^2$.

Ex. 5.8. Soit X une v.a. de la loi uniforme sur $[a, b]$, pour des réels $a < b$. Calculer $\text{Var}(X)$.

Ex. 5.9. Soit X une variable aléatoire de carré intégrable.

- Montrer que $Z = \frac{X - E[X]}{\sigma(X)}$ est centrée ($E[Z] = 0$) et standardisée ($\text{Var}(Z) = 1$).
- Si $m \in \mathbb{R}$ et $\sigma \in]0, \infty[$ sont tels que $\frac{X - m}{\sigma}$ est centrée et standardisée, montrer que $E[X] = m$ et $\text{Var}(X) = \sigma^2$.
- Si X suit la loi $\mathcal{N}(m, \sigma^2)$, montrer que $\frac{X - m}{\sigma} \sim \mathcal{N}(0, 1)$. Calculer l'espérance et la variance de la loi $\mathcal{N}(0, 1)$ et en déduire celles de X .

Proposition 5.7 (Inégalité de Bienaymé-Tchebychev). Soit X est une variable aléatoire de carré intégrable. Alors, pour tout $a > 0$

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

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, en appliquant l'inégalité de Markov à la variable aléatoire positive $(X - E[X])^2$, on a

$$P(|X - E[X]|^2 \geq a^2) \leq \frac{1}{a^2} E[|X - E[X]|^2].$$

La conclusion vient de la définition de la variance. □

Ex. 5.10. Si X est une variable aléatoire de carré intégrable, et $\alpha \geq 1$, montrer que l'inégalité de Tchebychev s'écrit aussi sous la forme suivante :

$$P(|X - E[X]| \geq \alpha \sigma(X)) \leq \frac{1}{\alpha^2},$$

et en déduire deux intervalles (non aléatoires) contenant X avec probabilité au moins 75 %, et au moins 99 %.

5.5 Vecteurs aléatoires de $\mathbb{R}^d$

Une variable aléatoire X à valeurs dans $\mathbb{R}^d$ est aussi appelée *vecteur aléatoire*.

Définition (Loi jointe et lois marginales). Soit $X = (X_1, \dots, X_d)$ un vecteur aléatoire.

- Les **lois marginales** de X sont : la loi de X_1 , la loi de $X_2, \dots$, la loi de X_d .
- La loi de X est appelée la **loi jointe** de $X_1, \dots, X_d$.

Attention ! Il ne suffit pas de connaître les lois marginales pour connaître la loi jointe.

Ex. 5.11. Soient deux variables aléatoires Y et Z indépendantes et de loi $\mathcal{E}(\lambda)$. On définit les couples $(X_1, X_2) = (Y, Z)$ et $(X'_1, X'_2) = (Y, Y)$. Vérifier qu'ils ont mêmes lois marginales mais qu'ils n'ont pas la même loi jointe.

Dans le cas de variables aléatoires discrètes ou à densité, on peut déduire les lois marginales à partir des lois jointes (mais pas l'inverse).

Proposition 5.8. Soit $X = (X_1, \dots, X_d)$ une variable aléatoire discrète à valeurs dans $E = E_1 \times \dots \times E_d$. Alors les lois marginales sont données, pour $i = 1, \dots, d$, par

$$\forall x_i \in E_i, \quad P(X_i = x_i) = \sum_{x_1 \in E_1} \dots \sum_{x_{i-1} \in E_{i-1}} \sum_{x_{i+1} \in E_{i+1}} \dots \sum_{x_d \in E_d} P(X = (x_1, \dots, x_d)).$$

Par exemple, si (X, Y) est une variable aléatoire discrète à valeurs dans $E \times F$, les lois marginales de X et Y sont données par

$$\text{pour } x \in E, \quad P(X = x) = \sum_{y \in F} P((X, Y) = (x, y)), \quad \text{pour } y \in F, \quad P(Y = y) = \sum_{x \in E} P((X, Y) = (x, y)).$$

Proposition 5.9. Soit $X = (X_1, \dots, X_d)$ une variable aléatoire continue à valeurs dans $\mathbb{R}^d$ de densité f . Alors $X_1, \dots, X_d$ ont aussi des densités $f_1, \dots, f_d$, données, pour $i = 1, \dots, d$, par

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

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

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

Ex. 5.12. Soit $Z = (X, Y)$ une variable aléatoire de loi uniforme sur le disque unité $D = \{(x, y) \mid x^2 + y^2 \leq 1\}$. Déterminer les lois marginales de X et Y .

Interprétations probabilistes

Le tableau suivant résume les correspondances entre la théorie de l'intégration et les probabilités :

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 $\bar{A}$)
$A \subset B$	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) $E[X] = \int_{\Omega} X dP$

Chapitre 6

Indépendance

Dans ce chapitre, on introduit la notion d'indépendance d'événements et d'indépendance de variables aléatoires.

6.1 Probabilité conditionnelle et indépendance d'événements

6.1.1 Probabilité conditionnelle

Définition (Probabilité conditionnelle d'événement). 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.

Ex. 6.1. Mon voisin a deux enfants. Si on apprend que l'un est un garçon, quelle est la probabilité que l'autre est une fille ?

On note que $P|_B : A \mapsto P(A | B)$ 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

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

Proposition 6.1. 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 telle 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**)

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

b) (**Formule de Bayes**) pour tout $1 \leq n \leq N$,

$$P(B_n | A) = \frac{P(B_n)P(A | B_n)}{\sum_{k=1}^N P(B_k)P(A | B_k)}.$$

Démonstration. a) La première formule se déduit du fait que $A = \bigsqcup_n (A \cap B_n)$, et la deuxième vient de $X = \sum_n X \mathbf{1}_{B_n}$. À noter que si $N = \infty$, la deuxième formule requiert le TCM (si $X \geq 0$) ou le TCD (si X est intégrable, avec domination par $|X|$).

b) La formule de Bayes se déduit directement de $P(B_n | A) = \frac{P(B_n \cap A)}{P(A)}$, en utilisant ensuite $P(B_n \cap A) = P(B_n)P(A | B_n)$ pour le numérateur et a) pour le dénominateur. $\square$

Ex. 6.2. Un laboratoire pharmaceutique a mis au point un test capable de détecter la présence d'une maladie qui touche en moyenne 1 personne sur 1000. Ce test n'est pas infallible : si le patient est malade, le test est positif avec probabilité 99 %, et si le patient est sain, le test est négatif avec probabilité 99,8 %. On effectue le test sur un patient choisi au hasard. Il est positif. Quelle est la probabilité que le patient soit malade ? Commenter.

6.1.2 Événements indépendants

Définition (Pour 2 événements). 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 à n événements :

Définition (Pour n événements). Les événements $A_1, \dots, A_n$ sont (mutuellement) **indépendants** si

$$\text{pour tous } i_1 < \dots < i_k, \quad P(A_{i_1} \cap \dots \cap A_{i_k}) = P(A_{i_1}) \cdots P(A_{i_k}).$$

Autrement dit, pour vérifier que A, B, C sont indépendants, il faut vérifier que l'on a $P(A \cap B \cap C) = P(A)P(B)P(C)$ mais aussi que $P(A \cap B) = P(A)P(B)$, $P(A \cap C) = P(A)P(C)$, $P(B \cap C) = P(B)P(C)$. Attention, il ne suffit pas de vérifier la première égalité, ou les trois suivantes.

Ex. 6.3. On considère un lancer de deux pièces équilibrées. Montrer que les événements A = « la 1^{re} pièce tombe sur pile », B = « la 2^e pièce tombe sur pile » et C = « les deux pièces tombent du même côté » sont indépendants 2 à 2, mais pas mutuellement indépendants.

6.2 Indépendance de variables aléatoires

Définition (Indépendance de v.a.). 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

$$\text{pour tous } B_1 \in \mathcal{E}_1, \dots, B_n \in \mathcal{E}_n \quad P(\{X_1 \in B_1\} \cap \dots \cap \{X_n \in B_n\}) = P(X_1 \in B_1) \cdots P(X_n \in B_n).$$

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 ».

Ex. 6.4. (*Paradoxe des anniversaires*) On se donne $X_1, \dots, X_n$ des v.a. de loi uniforme sur $\{1, \dots, 365\}$, qui représentent les dates d'anniversaire de n personnes ; on suppose qu'elles sont mutuellement indépendantes. Quelle est la probabilité qu'au moins 2 personnes aient leur anniversaire le même jour ? En donner une minoration en utilisant l'inégalité $1 - x \leq e^{-x}$ pour $x \in \mathbb{R}$. L'évaluer pour $n = 23$.

Définition (Loi binomiale). 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

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

Une interprétation importante est la suivante : la loi $\mathcal{B}(n, p)$ est la loi du nombre de piles obtenus en n lancers indépendants d'une pièce de monnaie qui tombe sur pile avec probabilité p :

Proposition 6.2. 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 . Remarquons que S est le nombre de $k \in \{1, \dots, n\}$ tel que $X_k = 1$.

Démonstration. On note tout d'abord que, comme $X_1, \dots, X_n$ sont à valeurs dans $\{0, 1\}$, leur somme S est à valeurs dans $\{0, \dots, n\}$. Soit $k \in \{0, \dots, n\}$. On décompose selon les valeurs de $X_1, \dots, X_n$: on a

$$\{S = k\} = \bigsqcup_{\substack{i_1, \dots, i_n \in \{0, 1\} : \\ i_1 + \dots + i_n = k}} \{X_1 = i_1, \dots, X_n = i_n\}$$

et cette union est finie (donc dénombrable) et disjointe donc

$$P(S = k) = \sum_{\substack{i_1, \dots, i_n \in \{0,1\} : \\ i_1 + \dots + i_n = k}} P(X_1 = i_1, \dots, X_n = i_n).$$

Pour tous $i_1, \dots, i_n \in \{0,1\}$ tels que $i_1 + \dots + i_n = k$, par indépendance des $X_1, \dots, X_n$ on a

$$P(X_1 = i_1, \dots, X_n = i_n) = p^k (1-p)^{n-k}.$$

Par ailleurs, le nombre de n -uplets $(i_1, \dots, i_n)$ tels que $i_1 + \dots + i_n = k$ est égal au nombre de parties à k éléments d'un ensemble qui en possède n (car il suffit de choisir quels indices valent 1), c.-à-d. $\binom{n}{k}$. Donc $P(S = k) = \binom{n}{k} p^k (1-p)^{n-k}$ comme annoncé. $\square$

Définition (Loi de Poisson). 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

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

Cette loi apparaît souvent comme loi d'un nombre d'événements « rares » qui se produisent parmi une « longue » suite de tirages indépendants. Un énoncé plus précis est le suivant, qui résulte d'un calcul simple :

Proposition 6.3. Soit $(p_n)_{n \geq 1}$ une suite de réels dans $[0,1]$ telle que $\lim_{n \rightarrow \infty} np_n = \lambda > 0$. Supposons que, pour tout n , S_n suive la loi $\mathcal{B}(n, p_n)$, alors

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

Ex. 6.5. Démontrer la proposition précédente, c'est-à-dire : pour tout $k \in \mathbb{N}$, $\lim_{n \rightarrow \infty} \binom{n}{k} p^k (1-p)^{n-k} = e^{-\lambda} \frac{\lambda^k}{k!}$.

6.2.1 Indépendance et loi produit

La définition d'indépendance rappelle la mesure produit. En notant $P_{(X_1, \dots, X_n)}$ la loi (jointe) du vecteur $(X_1, \dots, X_n)$, la définition d'indépendance de variables aléatoires nous dit en effet que les v.a. $X_1, \dots, X_n$ sont indépendantes si pour tous $B_1 \in \mathcal{E}_1, \dots, B_n \in \mathcal{E}_n$

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

Ceci nous donne le résultat suivant :

Théorème 6.4 (Indépendance et loi produit). Les variables aléatoires $X_1, \dots, X_n$ sont indépendantes si et seulement si la loi jointe du vecteur $(X_1, \dots, X_n)$ est le produit des lois marginales 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 toute fonction f mesurable positive ou telle que $f(X_1, \dots, X_n)$ est intégrable,

$$E[f(X_1, \dots, X_n)] = \int \cdots \int f(x_1, \dots, x_n) dP_{X_1}(x_1) \cdots dP_{X_n}(x_n),$$

où l'ordre d'intégration est quelconque. En particulier, pour toutes fonctions $f_1, \dots, f_n$ mesurables positives ou telles que $f_1(X_1), \dots, f_n(X_n)$ sont intégrables,

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

Démonstration. Le premier point vient de l'identité au dessus du théorème. Le second point est la traduction directe des théorèmes de Fubini (Fubini–Tonelli dans le cas positif, Fubini–Lebesgue dans le cas intégrable). La dernière formule est simplement le deuxième point appliqué au cas particulier $f(x_1, \dots, x_n) = f_1(x_1) \cdots f_n(x_n)$: dans ce cas, la formule se simplifie à l'aide de la linéarité de l'intégrale. $\square$

On en déduit le résultat suivant pour les vecteurs aléatoires à densité.

Proposition 6.5. Soient $X_1, \dots, X_d$ des variables aléatoires réelles.

a) Supposons que pour tout $i = 1, \dots, d$, la v.a. X_i ait pour densité f_i et que les v.a. $X_1, \dots, X_d$ soient indépendantes. Alors $(X_1, \dots, X_d)$ est un vecteur aléatoire à densité, de densité donnée par la fonction

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

b) Inversement, supposons que $(X_1, \dots, X_d)$ soit un vecteur aléatoire à densité, dont la densité s'écrit sous la forme

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

Alors les v.a. $X_1, \dots, X_d$ sont des v.a. à densité indépendantes, et pour $i = 1, \dots, d$, la densité de X_i est proportionnelle à g_i et est donc

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

6.2.2 Covariance

Définition (Covariance). Soient X, Y deux variables aléatoires réelles de carré intégrable. La **covariance** de X et Y est

$$\text{Cov}(X, Y) = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])].$$

Proposition 6.6. a) Pour toute variable aléatoire X , $\text{Cov}(X, X) = \text{Var}(X)$.

b) Pour toutes variables aléatoires X et Y de carré intégrable, $\text{Cov}(X, Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$.

c) Pour toutes variables aléatoires $X_1, \dots, X_n$ de carré intégrable,

$$\text{Var}(X_1 + \cdots + X_n) = \sum_{i=1}^n \text{Var}(X_i) + 2 \sum_{1 \leq i < j \leq n} \text{Cov}(X_i, X_j).$$

Démonstration. a) est immédiat, et b) s'obtient en développant le produit et en utilisant la linéarité de l'espérance. Pour c), même principe :

$$\begin{aligned} \text{Var}(X_1 + \cdots + X_n) &= \mathbb{E} \left[\left(\sum_{i=1}^n X_i - \mathbb{E} \left[\sum_{i=1}^n X_i \right] \right)^2 \right] = \mathbb{E} \left[\left(\sum_{i=1}^n (X_i - \mathbb{E}[X_i]) \right)^2 \right] \\ &= \mathbb{E} \left[\sum_{1 \leq i, j \leq n} (X_i - \mathbb{E}[X_i])(X_j - \mathbb{E}[X_j]) \right] = \sum_{1 \leq i, j \leq n} \text{Cov}(X_i, X_j) \\ &= \sum_{i=1}^n \text{Cov}(X_i, X_i) + 2 \sum_{1 \leq i < j \leq n} \text{Cov}(X_i, X_j) \end{aligned}$$

(par la symétrie $\text{Cov}(X_i, X_j) = \text{Cov}(X_j, X_i)$), d'où la formule. □

Covariance et indépendances. Une conséquence du théorème p.53 plus haut, liant indépendance et loi produit est que si X et Y sont indépendantes et positives (ou intégrables), alors $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$. Autrement dit, si deux v.a. X, Y de carré intégrable sont indépendantes, alors $\text{Cov}(X, Y) = 0$.

Ex. 6.6 – La réciproque est fausse. Soient U, V deux variables aléatoires de carré intégrable, indépendantes et de même loi. On pose $X = U + V$ et $Y = U - V$. Montrer que $\text{Cov}(X, Y) = 0$, mais qu'en général X et Y ne sont pas indépendantes (prendre par exemple U, V de loi $\mathcal{B}(p)$).

On a alors la conséquence suivante du point c) de la proposition précédente :

Corollaire 6.7. Si $X_1, \dots, X_n$ sont des variables aléatoires réelles, de carrés intégrables et (deux à deux) indépendantes, alors

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

Ex. 6.7. Calculer l'espérance et la variance de la loi binomiale $\mathcal{B}(n, p)$.

6.3 Propriétés de l'indépendance

Commençons par élargir la notion d'indépendance à une famille infinie.

Définition (Indépendance d'une famille infinie). *On dit qu'une famille infinie d'événements ou de variables aléatoires est indépendante si toute sous-famille finie est indépendante.*

Remarque. La définition de l'indépendance de variables aléatoires peut aussi s'introduire en définissant la notion plus fondamentale d'indépendance d'une famille de tribus, qui sera utile en *Probabilités 2*. C'est en réalité cette notion qui permet de démontrer les propriétés suivantes.

On donne deux opérations générales qui préservent l'indépendance à partir d'une première famille de variables indépendantes :

Propriétés. Soit $(X_n)_{n \geq 1}$ une suite de variables aléatoires indépendantes.

a) (**Indépendance par fonction mesurable**) Pour toutes fonctions mesurables $(f_n)_{n \geq 1}$, les variables aléatoires $(f_n(X_n))_{n \geq 1}$ sont indépendantes.

b) (**Indépendance par paquets**) Soit $(i_n)_{n \geq 1}$ une suite strictement croissante d'indices entiers et soient

$$Y_1 = (X_1, \dots, X_{i_1}), Y_2 = (X_{i_1+1}, \dots, X_{i_2}), \dots, Y_n = (X_{i_{n-1}+1}, \dots, X_{i_n}), \dots$$

Alors les variables $(Y_n)_{n \geq 1}$ sont indépendantes.

Le point b) montre que si on regroupe $X_1, \dots, X_n$ en paquets disjoints, ces paquets sont indépendants. Combiné avec le point a), cela montre que 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.

L'indépendance entre événements et entre variables aléatoires sont liées :

Proposition 6.8. Une famille $(A_n)_{n \geq 1}$ d'événements est indépendante si et seulement si leurs fonctions indicatrices $(\mathbf{1}_{A_n})_{n \geq 1}$ forment une famille indépendante.

Les opérations sur les ensembles correspondant à des opérations sur les fonctions indicatrices ($\mathbf{1}_{A^c} = 1 - \mathbf{1}_A$, $\mathbf{1}_{A \cap B} = \mathbf{1}_A \mathbf{1}_B$ et $\mathbf{1}_{A \cup B} = \mathbf{1}_A + \mathbf{1}_B - \mathbf{1}_A \mathbf{1}_B$), les propriétés précédentes s'adaptent donc aux événements : si A, B, C, D sont des événements indépendants, alors $A^c \cap B$ et $C \cup D$ sont indépendants, de même que $A^c \cup (B \cap D^c)$ et C , par exemple.

Donnons pour finir une dernière variable aléatoire importante, dont l'interprétation repose sur la notion de suite (infinie) d'événements indépendants.

Définition (Loi géométrique). 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

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

Pour $p = 1$, on interprète ceci comme $P(X = 1) = 1$.

Une interprétation importante de la loi $\mathcal{G}(p)$ est comme étant la loi de l'instant où l'on obtient pile pour la première fois dans une suite de lancers indépendants d'une pièce qui a probabilité p de tomber sur pile : c'est la loi du « temps de premier succès » dans une suite d'expériences indépendantes ayant chacune pour probabilité de succès p . Ceci s'exprime par la proposition suivante.

Proposition 6.9. Soit $p \in]0, 1]$. Si $(X_n)_{n \geq 1}$ sont des variables aléatoires indépendantes toutes de loi $\mathcal{B}(p)$, alors la variable aléatoire

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

est finie presque sûrement, et suit la loi géométrique de paramètre p .

Démonstration. Vérifions d'abord $N < \infty$ p.s., c'est-à-dire que p.s., il existe $n \geq 1$ tel que $X_n = 1$. La suite d'événements $A_n = \{X_1 = 0, \dots, X_n = 0\}$ est décroissante, donc

$$\lim_{n \rightarrow \infty} P(A_n) = P\left(\bigcap_{n=1}^{\infty} A_n\right).$$

Or $P(A_n) = (1-p)^n \rightarrow 0$ car $1-p < 1$ vu l'hypothèse, et $\bigcap_n A_n = \{\forall n \geq 1, X_n = 0\} = \{N = \infty\}$, donc l'égalité précédente donne $P(N = \infty) = 0$, c.-à-d. $P(N < \infty) = 1$.

Ainsi, N prend ses valeurs dans $\mathbb{N}^*$ et, pour tout $k \geq 1$, on a $\{N = k\} = \{X_1 = 0, \dots, X_{k-1} = 0, X_k = 1\}$. Ainsi, par indépendance des $X_1, \dots, X_k$, on a

$$P(N = k) = P(X_1 = 0) \cdots P(X_{k-1} = 0)P(X_k = 1) = (1-p)^{k-1}p,$$

ce qui conclut. On aurait aussi pu déduire $P(N < \infty) = 1$ à partir du calcul de $P(N = k)$, car

$$P(N < \infty) = P\left(\bigcup_{k=1}^{\infty} \{N = k\}\right) = \sum_{k=1}^{\infty} P(N = k) = \sum_{k=1}^{\infty} (1-p)^{k-1}p = \frac{1}{1-(1-p)}p = 1. \quad \square$$

Ex. 6.8 – Absence de mémoire. Soit X une v.a. de loi $\mathcal{G}(p)$ pour $p \in]0,1[$.

1. Montrer que pour tout $k \in \mathbb{N}$, $P(X > k) = (1-p)^k$.

2. En déduire que pour tous $k, \ell \in \mathbb{N}$ on a $P(X > k + \ell \mid X > k) = P(X > \ell)$. Interpréter.

Récapitulatif de lois importantes

Variables aléatoires discrètes importantes

		$P(X = k)$	$E(X)$	$\text{Var}(X)$
Bernoulli	$p \in [0,1]$	$\begin{cases} p & \text{si } k = 1, \\ 1-p & \text{si } k = 0. \end{cases}$	p	$p(1-p)$
Binomiale	$n \in \mathbb{N}^*, p \in [0,1].$	$\binom{n}{k} p^k (1-p)^{n-k}, k \in \{0,1,\dots,n\}$	np	$np(1-p)$
Poisson	$\lambda \in]0,\infty[$	$e^{-\lambda} \frac{\lambda^k}{k!}, k \in \mathbb{N}$	λ	λ
Géométrique	$p \in]0,1]$	$p(1-p)^{k-1}, k \in \mathbb{N}^*$	$\frac{1}{p}$	$\frac{1-p}{p^2}$
Uniforme	$\{1, \dots, m\}$	$\frac{1}{m}, k \in \{1, \dots, m\}$	$\frac{m+1}{2}$	$\frac{m^2-1}{12}$

Variables aléatoires à densité importantes

		densité $f_X(x)$	$E(X)$	$\text{Var}(X)$
Uniforme	$[a,b]$	$\frac{1}{b-a} \mathbf{1}_{]a,b[}(x)$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$
Exponentielle	$\lambda \in]0,\infty[$	$\lambda e^{-\lambda x} \mathbf{1}_{\mathbb{R}_+}(x)$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$
Normale	$m \in \mathbb{R}, \sigma \in]0,\infty[$	$\frac{1}{\sqrt{2\pi}\sigma^2} e^{-\frac{(x-m)^2}{2\sigma^2}}$	m	σ^2

Chapitre 7

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 7.1. 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 tout pavé fermé A de $\mathbb{R}^d$, $P(X \in A) = P(Y \in A)$;
- (iv) pour toute fonction mesurable positive $f : \mathbb{R}^d \rightarrow \mathbb{R}_+$, $E[f(X)] = E[f(Y)]$;
- (v) pour toute fonction C^∞ à support compact f , $E[f(X)] = E[f(Y)]$.

Souvent, on utilisera plutôt le calcul d'une fonction associée à X et qui caractérise sa loi : fonction de répartition, fonction génératrice ou fonction caractéristique. Passons en revue ces différentes méthodes.

7.1 Fonction de répartition (variables aléatoires réelles)

Définition (Fonction de répartition). Soit X une variable aléatoire réelle. Sa **fonction de répartition** est la fonction

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

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 .

Théorème 7.2 (La fonction de répartition caractérise la loi). La loi d'une variable aléatoire réelle 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.

Proposition 7.3. a) F_X est croissante, continue à droite, et $\lim_{t \rightarrow -\infty} F_X(t) = 0$, $\lim_{t \rightarrow +\infty} F_X(t) = 1$.

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

Propriété. Pour tout $t \in \mathbb{R}$, la limite de F_X en t à gauche est $F_X(t^-) = P(X < t)$. Ainsi, F_X est continue en t si et seulement si $P(X = t) = 0$.

Les cas des variables discrètes et continues sont particulièrement importants (mais ce ne sont pas les seuls!).

Proposition 7.4 (Cas discret). 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 $P(X = x)$.

Proposition 7.5 (Cas à densité). Si X est une variable aléatoire de densité f_X , on a

$$F_X(x) = \int_{-\infty}^x f_X(t) dt.$$

Ainsi, si f_X est continue par morceaux, F_X est la primitive de f_X nulle en $-\infty$; de plus, 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 $f_X = F_X'$.

Ex. 7.1. Déterminer et représenter la fonction de répartition des variables aléatoires suivantes :

1. X , de loi uniforme sur $\{1, 2, \dots, 6\}$;
2. $X = \max(U, 1/2)$, où U est de loi uniforme sur $[0, 1]$.
3. $X = \max(U, V)$, où U, V sont indépendantes et de loi uniforme sur $[0, 1]$.

7.2 Fonction génératrice (variables aléatoires entières)

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

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

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

Théorème 7.6 (La fonction génératrice caractérise la loi). La loi d'une variable aléatoire X à valeur dans $\mathbb{N}$ est déterminée par sa fonction génératrice : 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 peut aussi appliquer le théorème de dérivation sous l'intégrale pour le justifier) : pour tout $s \in]-1, 1[$, pour tout $k \geq 0$,

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

d'où $G_X^{(k)}(0) = k! \mathbb{P}(X = k)$ 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 , $\mathbb{P}(X = k) = \mathbb{P}(Y = k)$, ce qui implique que X et Y ont même loi (puisqu'elles sont à valeurs dans $\mathbb{N}$). $\square$

On peut remarquer que la dérivée G_X' est croissante sur $[0, 1[$, donc admet toujours une limite en 1 par la gauche (dans $\mathbb{R}$), même si G_X n'est pas dérivable en 1. On la note $G_X'(1^-) = \lim_{s \uparrow 1} G_X'(s) \in [0, \infty]$, et il s'avère qu'elle est égale à l'espérance de X :

Proposition 7.7. On a $G_X'(1^-) = \mathbb{E}[X] \in [0, \infty]$. Plus généralement, la dérivée k -ème de G_X vérifie

$$G_X^{(k)}(1^-) = \lim_{s \uparrow 1} G_X^{(k)}(s) = \mathbb{E}[X(X-1) \cdots (X-k+1)] \in [0, \infty].$$

Démonstration. Comme observé précédemment, on peut dériver terme à terme sur $]-1, 1[$: pour tout $s \in]-1, 1[$,

$$G_X'(s) = \sum_{k=1}^{\infty} k s^{k-1} \mathbb{P}(X = k) = \mathbb{E}[X s^{X-1} \mathbf{1}_{\{X \geq 1\}}].$$

Or, presque sûrement, $s \mapsto X s^{X-1} \mathbf{1}_{\{X \geq 1\}}$ est croissante, positive, et converge vers X quand $s \rightarrow 1^-$. Par le théorème de convergence monotone, on en déduit la proposition. (Formellement, il faut appliquer le théorème à une suite; pour cela, comme la croissance de G_X' assure que sa limite en 1^- existe, c'est aussi la limite de $G_X'(s_n)$ où $s_n = 1 - \frac{1}{n}$, et on peut alors raisonner sur la suite $Z_n = X s_n^{X-1} \mathbf{1}_{\{X \geq 1\}}$ comme précédemment). $\square$

Le cas de la dérivée k -ème est identique. $\square$

Ex. 7.2. Soit X une v.a. de loi $\mathcal{G}(p)$ pour $p \in]0,1[$.

1. Calculer la fonction génératrice G_X . Quel est son rayon de convergence ?

2. Calculer G'_X et G''_X . En déduire $E[X]$ et $\text{Var}(X)$.

Proposition 7.8. Soit $X_1, \dots, X_n$ des variables aléatoires à valeurs dans $\mathbb{N}$ **indépendantes**. Alors, pour tout $s \in [-1,1]$, on a $G_{X_1+\dots+X_n}(s) = G_{X_1}(s) \cdots G_{X_n}(s)$.

Démonstration. On a, pour tout s ,

$$G_{X_1+\dots+X_n}(s) = E[s^{X_1+\dots+X_n}] = E[s^{X_1} \cdots s^{X_n}] = E[s^{X_1}] \cdots E[s^{X_n}] = G_{X_1}(s) \cdots G_{X_n}(s). \quad \square$$

Ex. 7.3. Soit X, Y deux variables aléatoires indépendantes, de lois $\mathcal{P}(\lambda)$ et $\mathcal{P}(\mu)$.

1. Calculer la fonction génératrice de X et Y . Retrouver la valeur de $E[X]$.

2. En déduire la loi de $Z = X + Y$

7.3 Méthode de la fonction test (variables à densité)

Une méthode générale pour déterminer une densité est d'utiliser le théorème de transfert (qui correspond au point (iv) de la Proposition 7.1 en début de chapitre) :

Proposition 7.9 (Méthode de la fonction test). Une variable aléatoire X à valeurs dans $\mathbb{R}^d$ est à densité et de densité f si et seulement si pour toute fonction $h : \mathbb{R}^d \rightarrow \mathbb{R}_+$ mesurable positive (appelée fonction test), on a la formule de transfert $E[h(X)] = \int h(x)f(x)dx$.

La méthode de la fonction test se déroule de la manière suivante :

- 1) On se donne une fonction test $h : \mathbb{R} \rightarrow \mathbb{R}_+$ mesurable positive quelconque ;
- 2) Après des manipulations, on écrit $E[h(Y)]$ sous la forme $\int h(y)f(y)dy$, pour une fonction f à déterminer ;
- 3) On conclut grâce à la méthode de la fonction test que $f = f_Y$ est la densité de la variable aléatoire Y .

Exemple. Si X est une v.a. réelle continue de densité f alors pour $a > 0$ et $b \in \mathbb{R}$, on peut déterminer la densité de $Y = aX + b$. En effet, on peut utiliser le théorème de transfert pour X , pour écrire

$$E[h(Y)] = E[h(aX + b)] = \int_{-\infty}^{+\infty} h(ax + b)f(x)dx.$$

En utilisant le changement de variable $y = ax + b$ ($x = \frac{y-b}{a}$, $dx = \frac{1}{a}dy$), on en conclut que

$$E[h(Y)] = \int_{-\infty}^{+\infty} h(y)\frac{1}{a}f\left(\frac{y-b}{a}\right)dy = \int_{-\infty}^{+\infty} h(y)f_Y(y)dy,$$

avec $f_Y(y) = \frac{1}{a}f\left(\frac{y-b}{a}\right)$. Ainsi, la méthode de la fonction test montre que $Y = aX + b$ a pour densité $\frac{1}{a}f\left(\frac{y-b}{a}\right)$.

Ex. 7.4. Soit $\lambda, a > 0$. Montrer que si X suit la loi $\mathcal{E}(\lambda)$, alors $e^{-\lambda X}$ suit la loi $\mathcal{U}([0,1])$.

Image d'un vecteur aléatoire à densité par un $\mathcal{C}^1$ -difféomorphisme. Si $X = (X_1, \dots, X_d)$ est un vecteur aléatoire de 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 $\varphi : U \rightarrow D$ (où D 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 4.8 de changement de variable et de la Proposition 7.9 ci-dessus. On calcule la densité avec la méthode de la fonction test : pour toute fonction mesurable positive $h : \mathbb{R}^d \rightarrow \mathbb{R}_+$, en appliquant le théorème de transfert à la fonction $h \circ \varphi$ on obtient

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

à l'aide du théorème de changement de variable pour la dernière égalité. Ainsi, la méthode de la fonction test montre que $(Y_1, \dots, Y_d)$ a pour densité

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

(C'est bien une densité : cette fonction est positive et son intégrale vaut 1, prendre par exemple $h = 1$ ci-dessus.)

La formule n'est pas à retenir, il vaut mieux faire le calcul à chaque fois (ce qui réduit le risque d'erreur).

Ex. 7.5. Pour $c > 0$, on note $\gamma(c)$ la loi gamma de paramètre c , qui a pour densité

$$f_c : x \mapsto f_c(x) = \frac{1}{\Gamma(c)} x^{c-1} e^{-x} \mathbf{1}_{\mathbb{R}_+}(x).$$

Soit $a, b > 0$ et X, Y deux variables aléatoires indépendantes de lois $\gamma(a)$ et $\gamma(b)$. On pose $U = \frac{X}{X+Y}$ et $V = X+Y$.

1. Calculer la loi du couple (U, V) .

2. Les variables aléatoires U et V sont-elles indépendantes? Quelles sont leurs lois?

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

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

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

où $\cdot$ désigne le produit scalaire usuel de $\mathbb{R}^d$ (ou le produit usuel si $d = 1$) et $i^2 = -1$.

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

Remarque. Si X est une variable réelle de pour densité f , alors pour tout $t \in \mathbb{R}$,

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

Ainsi, Φ_X est la (conjuguée de la) transformée de Fourier de f (voir la définition p.32).

Théorème 7.10 (La fonction caractéristique caractérise la loi). La loi d'un vecteur aléatoire X à valeur dans $\mathbb{R}^d$ est déterminé par sa fonction caractéristique : 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.

Les propriétés suivantes s'obtiennent via les théorèmes de continuité et dérivation sous l'intégrale.

Proposition 7.11. 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 $\Phi'_X(0) = i \mathbb{E}[X]$.
- c) De manière générale, si $\mathbb{E}[|X|^m] < \infty$, alors Φ_X est de classe $\mathcal{C}^m$ sur $\mathbb{R}$ et $\Phi_X^{(m)}(0) = i^m \mathbb{E}[X^m]$.

Ex. 7.6. Soit Z une variable aléatoire de loi $\mathcal{N}(0,1)$, c'est-à-dire de densité $x \mapsto \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$.

- 1. Justifier que Φ_Z est dérivable et donner une expression intégrale de $\Phi'_Z(t)$.
- 2. En intégrant par parties, montrer que Φ_Z est solution de l'équation différentielle $y' = -ty$.
- 3. En déduire la fonction Φ_Z , puis la fonction caractéristique d'une variable X de loi $\mathcal{N}(m, \sigma^2)$.

Proposition 7.12. Soit $X_1, \dots, X_n$ des variables aléatoires indépendantes à valeurs dans $\mathbb{R}^d$, alors

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

Démonstration. On a, pour tout $t \in \mathbb{R}^d$, en utilisant l'indépendance des X_i ,

$$\Phi_{X_1 + \dots + X_n}(t) = \mathbb{E}[e^{it(X_1 + \dots + X_n)}] = \mathbb{E}[e^{itX_1} \cdots e^{itX_n}] = \mathbb{E}[e^{itX_1}] \cdots \mathbb{E}[e^{itX_n}] = \Phi_{X_1}(t) \cdots \Phi_{X_n}(t). \quad \square$$

Ex. 7.7. Soit X, Y deux variables aléatoires indépendantes, de lois respectives $\mathcal{N}(m_X, \sigma_X^2)$ et $\mathcal{N}(m_Y, \sigma_Y^2)$. Déterminer la loi de $Z = X + Y$ à l'aide des fonctions caractéristiques.

Chapitre 8

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.

8.1 Suites d'événements : deux outils

8.1.1 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, par sous-additivité,

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

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.

8.1.2 Lemme de Borel-Cantelli

On considère $(A_n)_{n \geq 1}$ une suite d'événements et on note $N = \sum_{n \geq 1} \mathbf{1}_{A_n}$, qui est le nombre d'événements A_n qui se réalisent (à valeurs dans $\mathbb{N} \cup \{+\infty\}$).

Théorème 8.1 (Lemme de Borel-Cantelli). Soit $(A_n)_{n \geq 1}$ une suite d'événements.

a) On suppose que $\sum_{n \geq 1} P(A_n) < \infty$. Alors

$$p.s. \quad N = \sum_{n \geq 1} \mathbf{1}_{A_n} < +\infty,$$

c'est-à-dire : p.s. il n'y a qu'un nombre fini d'événements A_n qui se réalisent.

b) On suppose que les événements $(A_n)_n$ sont (deux à deux) indépendants et que $\sum_{n \geq 1} P(A_n) = \infty$.

Alors

$$p.s. \quad N = \sum_{n \geq 1} \mathbf{1}_{A_n} = +\infty,$$

c'est-à-dire : p.s. il y a une infinité d'événements A_n qui se réalisent.

Remarque. On peut écrire l'événement $\{N = \infty\}$ formellement de la manière suivante :

$$\{N = \infty\} = \{\text{il existe une infinité d'indices } n \text{ tels que } A_n \text{ se réalise}\} = \bigcap_{n \geq 1} \bigcup_{k \geq n} A_k.$$

Cet événement est noté $\limsup_n A_n := \bigcap_{n \geq 1} \bigcup_{k \geq n} A_k$ (voir l'analogie avec la $\limsup$ d'une suite réelle p.11). De manière analogue, on définit l'événement $\liminf_n A_n := \bigcup_{n \geq 1} \bigcap_{k \geq n} A_k$ qui correspond à l'événement « à partir d'un certain rang, A_n se réalise ».

Démonstration. a) Par le théorème de convergence monotone pour les séries positives, on a

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

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

b) Rappelons que $\{N = \infty\} = \bigcap_{n \geq 1} \bigcup_{k \geq n} A_k$, où la première intersection est décroissante : on a donc

$$P(N = \infty) = P\left(\bigcap_{n \geq 1} \bigcup_{k \geq n} A_k\right) = \lim_{n \rightarrow \infty} P\left(\bigcup_{k \geq n} A_k\right).$$

Montrons maintenant une propriété utile :

Lemme. Pour toute suite $(A_n)_{n \geq 1}$ d'événements indépendants, $P\left(\bigcap_{n \geq 1} A_n\right) = \prod_{n \geq 1} P(A_n)$.

Démonstration. Pour tout $m \geq 1$, on note $B_m = A_1 \cap \dots \cap A_m$. Alors la suite $(B_m)_{m \geq 1}$ est décroissante, donc

$$\lim_{m \rightarrow \infty} P(B_m) = P\left(\bigcap_{m \geq 1} B_m\right) = P\left(\bigcap_{n \geq 1} A_n\right).$$

Or, vu l'indépendance des A_n , $P(B_m) = P(A_1) \cdots P(A_m)$, qui tend vers $\prod_{n \geq 1} P(A_n)$ quand $m \rightarrow \infty$, d'où le résultat. $\square$

En utilisant le lemme (pour la suite $(A_k^c)_{k \geq n}$), on obtient

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

Ainsi, grâce à 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. On en déduit que $P\left(\bigcup_{k \geq n} A_k\right) = 1$, donc $P(N = \infty) = 1$, comme annoncé. $\square$

Remarque. On a dans tous les cas $E[N] = \sum_{n=1}^{\infty} P(A_n)$, mais dans le cas où $E[N] = +\infty$ on ne peut pas en déduire directement que $N = \infty$ p.s. : l'hypothèse d'indépendance est cruciale !

Ex. 8.1. (*Démonstration alternative de b)*) On note $N_k = \sum_{n=1}^k \mathbf{1}_{A_n}$ et $m_k = \sum_{n=1}^k P(A_n) = E[N_k]$; on suppose que $\lim_{k \rightarrow \infty} m_k = +\infty$.

1. Calculer $\text{Var}(N_k)$ et montrer que $\text{Var}(N_k) \leq m_k$.
2. Montrer que $P(N \leq \frac{1}{2}m_k) \leq P(N_k \leq \frac{1}{2}m_k) \leq 4/m_k$. (*Utiliser l'inégalité de Bienaymé-Tchebychev.*)
3. Montrer que $P(N < +\infty) = \lim_{k \rightarrow \infty} P(N \leq \frac{1}{2}m_k)$ et conclure.

Ex. 8.2. Soit $(X_n)_{n \geq 1}$ une suite de variables aléatoires indépendantes, de loi $\mathcal{B}(p)$, où $p \in]0,1[$.

1. Montrer que la suite contient presque sûrement une infinité de 1.
2. Montrer que la suite contient presque sûrement une infinité de fois la séquence "01".
3. Montrer que la suite contient presque sûrement une infinité de fois n'importe quelle séquence finie.

8.2 À propos des suites infinies de variables aléatoires

Si $X_1, X_2, \dots$ sont des variables aléatoires réelles alors 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 jointe décrit notamment les dépendances possibles entre $X_1, \dots, X_d$). Pour voir $(X_n)_{n \geq 1}$ comme une « suite aléatoire », c'est à-dire comme une variable aléatoire à valeurs dans $\mathbb{R}^{\mathbb{N}}$, il faut munir l'espace $E = \mathbb{R}^{\mathbb{N}}$ des suites réelles d'une tribu appropriée.

Définition (Tribu cylindrique). 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 :

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

On peut vérifier alors que $(X_n)_{n \geq 1}$ est une variable aléatoire à valeurs dans $(E, \mathcal{B})$ si et seulement si, pour tout $n \geq 1$, X_n est une variable aléatoire réelle.

En conséquence de la définition de $\mathcal{B}$, pour décrire la loi de la suite $(X_n)_{n \geq 1}$, il suffit de donner la loi de $(X_1, \dots, X_n)$ pour tout $n \geq 1$. En particulier, si la suite $(X_n)_{n \geq 1}$ est indépendante, alors (par définition) les variables aléatoires $X_1, \dots, X_n$ sont indépendantes pour tout n , donc la loi du vecteur $(X_1, \dots, X_n)$ est la loi produit des lois de $X_1, \dots, X_n$, de sorte que la loi de la suite $(X_n)_{n \geq 1}$ 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 infinie de variables aléatoires indépendantes de loi uniforme sur $[0,1]$ (ce cas permettrait d'ailleurs d'obtenir un cas plus général).

Proposition 8.2. *Il existe un espace de probabilités $(\Omega, \mathcal{A}, P)$ sur lequel sont définies des variables aléatoires indépendantes $(X_n)_{n \geq 1}$, toutes de loi $\mathcal{U}([0,1])$.*

La démonstration de cette proposition est fortement basée sur le lemme suivant, qui construit déjà une suite infinie de tirages à pile ou face équilibrés indépendants.

Lemme. a) 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

$$U = \overline{0, X_1 X_2 \cdots} = \sum_{n=1}^{\infty} \frac{X_n}{2^n}. \quad (*)$$

(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 $\mathcal{B}(\frac{1}{2})$.

b) 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. $\square$

Démonstration de la Proposition 8.2. Le lemme permet de définir une suite infinie $(X_n)_{n \geq 1}$ de variables aléatoires indépendantes de loi $\mathcal{B}(\frac{1}{2})$. On peut utiliser cette suite pour définir une suite infinie $(Y_n)_{n \geq 1}$ de variables aléatoires indépendantes de loi $\mathcal{U}([0,1])$. En effet, il suffit d'utiliser le fait que $\mathbb{N}^*$ et $\mathbb{N}^* \times \mathbb{N}^*$ sont en bijection l'un avec l'autre (par dénombrabilité) : on peut donc ré-indexer la suite $(X_n)_{n \in \mathbb{N}^*}$ en $(X_{i,j})_{(i,j) \in \mathbb{N}^* \times \mathbb{N}^*}$. En posant $Y_j = \sum_{i=1}^{\infty} 2^{-i} X_{i,j}$, on obtient que $Y_j \sim \mathcal{U}([0,1])$ par la proposition précédente et les variables $(Y_j)_{j \geq 1}$ sont indépendantes grâce à l'indépendance par paquets. $\square$

Notation. 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).

8.3 Convergence d'une suite de variables aléatoires

Dans cette section, on considère une suite $(X_n)_{n \geq 1}$ de variables aléatoires réelles et X une variable aléatoire réelle, **définies sur le même espace de probabilités** $(\Omega, \mathcal{A}, P)$. Les définitions suivantes s'étendent aisément au cas de vecteurs aléatoires.

8.3.1 Convergence en probabilité

Définition. La suite $(X_n)_{n \geq 1}$ **converge vers X en probabilité**, ce que l'on note aussi $X_n \xrightarrow{P} X$, si

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

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

Propriété. Si $X_n \xrightarrow{P} X$ et $X_n \xrightarrow{P} X'$, alors $X = X'$ p.s.

Ex. 8.3. Montrer que $X_n \xrightarrow{P} X$ si et seulement si $\lim_{n \rightarrow \infty} E[|X_n - X| \wedge 1] = 0$, où $x \wedge y = \min(x, y)$.

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

Proposition 8.3 (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 $E[(X_n)^2] < \infty$). Alors on a la convergence en probabilité suivante :

$$\frac{X_1 + \cdots + X_n}{n} \xrightarrow{P} E[X_1].$$

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

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

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

et

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

car $X_1, \dots, X_n$ sont indépendantes. Comme $X_1, \dots, X_n$ ont même loi, cela donne

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

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

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

Supposons que $(X_n)_{n \geq 1}$ soit une suite de v.a. i.i.d. de loi $\mathcal{B}(\frac{1}{2})$ (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

$$\frac{X_1 + \cdots + X_n}{n} \xrightarrow{P} E[X_1] = \frac{1}{2}.$$

Or les X_i valent 0 ou 1, donc $X_1 + \cdots + X_n$ est aussi le nombre d'indices entre 1 et n pour lesquels $X_i = 1$, de sorte que $\frac{X_1 + \cdots + X_n}{n}$ 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 $\frac{1}{2}$ 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 $\frac{1}{2}$; ce serait une « convergence presque sûre ».

8.3.2 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 $(X_n)_{n \geq 1}$ **converge vers X presque sûrement**, noté $X_n \xrightarrow{p.s.} X$, si, presque sûrement, $(X_n)_{n \geq 0}$ converge vers X , c'est-à-dire si

$$P\left(\lim_{n \rightarrow \infty} X_n = X\right) = 1.$$

Autrement dit $(X_n)_{n \geq 1}$ 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 8.4. On a $X_n \xrightarrow{p.s.} X$ si et seulement si

$$\text{pour tout } \delta > 0 \quad \lim_{n \rightarrow \infty} P\left(\sup_{k \geq n} |X_k - X| > \delta\right) = P(|X_n - X| > \delta \text{ infiniment souvent}) = 0.$$

Démonstration. Notons tout d'abord que $\{\sup_{k \geq n} |X_k - X| > \delta\} = \bigcup_{k \geq n} \{|X_k - X| > \delta\}$ de sorte qu'on a la limite (décroissante) suivante :

$$\lim_{n \rightarrow \infty} P\left(\sup_{k \geq n} |X_k - X| > \delta\right) = P\left(\bigcap_{n \geq 1} \bigcup_{k \geq n} |X_k - X| > \delta\right) = P(|X_n - X| > \delta \text{ une infinité de fois}).$$

Supposons $X_n \xrightarrow{p.s.} X$ et fixons $\delta > 0$. On a p.s. $|X_n - X| \leq \delta$ pour n grand, donc l'événement contraire, « pour tout n , il existe $k \geq n$ tel que $|X_k - X| > \delta$ », est négligeable :

$$0 = P\left(\bigcap_{n \geq 1} \bigcup_{k \geq n} |X_k - X| > \delta\right),$$

ce qui montre la première implication.

Montrons la réciproque : supposons que pour tout $\delta > 0$ on ait $P\left(\bigcap_{n \geq 1} \bigcup_{k \geq n} |X_k - X| > \delta\right) = 0$, c'est-à-dire, en passant au complémentaire, $P\left(\bigcup_{n \geq 1} \bigcap_{k \geq n} |X_k - X| > \delta\right) = 1$. Alors, en prenant $\delta \in \mathbb{Q}$ on en déduit (par intersection dénombrable d'événements presque sûrs)

$$1 = P\left(\bigcap_{\delta > 0, \delta \in \mathbb{Q}} \bigcup_{n \geq 1} \bigcap_{k \geq n} |X_k - X| \leq \delta\right) = P(\forall \delta > 0, \delta \in \mathbb{Q}, \text{ on a } |X_n - X| \leq \delta \text{ à partir d'un certain rang}).$$

On en conclut bien que $X_n \xrightarrow{p.s.} X$. □

Corollaire 8.5. a) Si $X_n \xrightarrow{p.s.} X$, alors $X_n \xrightarrow{P} X$;

b) Si $\sum_{n \geq 0} P(|X_n - X| > \delta) < \infty$ pour tout $\delta > 0$, alors $X_n \xrightarrow{p.s.} X$;

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

Démonstration. On va utiliser la proposition précédente.

a) Comme $\sup_{k \geq n} |X_k - X| \leq |X_n - X|$, on a donc $P(|X_n - X| > \delta) \leq P(\sup_{k \geq n} |X_k - X| > \delta)$. Ainsi, le terme de droite tend vers 0 si $X_n \xrightarrow{p.s.} X$ et donc on a bien $X_n \xrightarrow{P} X$.

b) Le lemme de Borel–Cantelli montre que, pour tout $\delta > 0$, $P(|X_n - X| > \delta \text{ une infinité de fois}) = 0$. La proposition précédente montre que $X_n \xrightarrow{p.s.} X$.

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

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

Alors, pour tout $\delta > 0$,

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

donc le b) s'applique à la suite $(X_{n_k})_{k \geq 1}$, qui converge donc p.s. vers X . □

Théorème 8.6 (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 telles que $E[|X_1|] < \infty$). Alors on a

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

Démonstration. On peut supposer dans la démonstration que $E[X_1] = 0$ (car ensuite, on peut écrire $X_n = (X_n - E[X_1]) + E[X_1]$ pour s'y ramener). 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, etc.), on peut donner une preuve très courte qui s'apparente à celle de la loi faible des grands nombres L^2 .

Notons $S_n = X_1 + \cdots + X_n$. En développant $(S_n)^4$ puis en regroupant les termes du type $E[X_i(X_j X_k X_l)] = E[X_i]E[X_j X_k X_l] = 0$, $E[(X_i)^4] = E[(X_1)^4]$, $E[(X_i)^2(X_j)^2] = E[(X_i)^2]E[(X_j)^2] = E[(X_1)^2]^2$, on obtient

$$\begin{aligned} E[(S_n)^4] &= E\left[\sum_{1 \leq i,j,k,l \leq n} X_i X_j X_k X_l\right] = \sum_{1 \leq i,j,k,l \leq n} E[X_i X_j X_k X_l] \\ &= 3n(n-1)E[(X_1)^2]^2 + nE[(X_1)^4] \underset{n \rightarrow \infty}{\sim} 3n^2 E[(X_1)^2]^2. \end{aligned}$$

Ainsi (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. Ainsi, p.s. le terme général $\frac{(S_n)^4}{n^4}$ tend vers 0 et $\lim_{n \rightarrow \infty} \frac{S_n}{n} = 0$ p.s., ce que l'on voulait démontrer. $\square$

Pour des tirages à pile ou face, ce résultat traduit l'observation : pour (presque) toute suite de tirages, la proportion de piles converge vers $\frac{1}{2}$. 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.

8.3.3 Convergences L^1 et L^2

On peut aussi définir pour les variables aléatoires les normes L^1 et L^2 et donc les convergences associées :

Définition. La suite $(X_n)_{n \geq 1}$ **converge vers X dans L^1** , noté $X_n \xrightarrow{L^1} X$, si $\|X_n - X\|_1 \xrightarrow{n \rightarrow \infty} 0$, c'est-à-dire si

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

La suite $(X_n)_{n \geq 1}$ **converge vers X dans L^2** , noté $X_n \xrightarrow{L^2} X$, si $\|X_n - X\|_2 \xrightarrow{n \rightarrow \infty} 0$, c'est-à-dire si

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

Noter que pour avoir une convergence $X_n \xrightarrow{L^p} X$ ($p = 1$ ou $p = 2$), les variables aléatoires X_n et X doivent être dans L^p . On a aussi la propriété suivante de convergence des espérances. *Attention, la réciproque est fautive !*

Propriétés. • Si $X_n \xrightarrow{L^1} X$, alors $\lim_{n \rightarrow \infty} E[X_n] = E[X]$.
• Si $X_n \xrightarrow{L^2} X$, alors $\lim_{n \rightarrow \infty} E[(X_n)^2] = E[X^2]$.

Démonstration. Pour le premier point, on a simplement $|E[X_n] - E[X]| = |E[X_n - X]| \leq E[|X_n - X|] \xrightarrow{n \rightarrow \infty} 0$.

Pour le deuxième point, on peut observer que pour n'importe quelle norme $\|x\| - \|y\| \leq \|x - y\|$ (cela découle de l'inégalité triangulaire). Ainsi, cela montre que si $\|X_n\|_2 \xrightarrow{n \rightarrow \infty} \|X\|_2$, d'où le résultat en passant au carré. $\square$

Proposition 8.7. Si $X_n \xrightarrow{L^2} X$, alors $X_n \xrightarrow{L^1} X$.

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 en prenant $n \rightarrow \infty$. $\square$

Proposition 8.8. Si $X_n \xrightarrow{L^p} X$ (où $p = 1$ ou 2), alors $X_n \xrightarrow{P} X$.

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

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

En revanche, aucune implication n'existe en général entre convergence presque sûre et la convergence L^1 ou convergence L^2 , comme le montrent les exercices suivants.

Ex. 8.4. Soit U une variable aléatoire de loi uniforme sur $[0,1]$. On pose $X_n = n \mathbf{1}_{\{U \leq \frac{1}{n}\}}$.

1. Montrer que $X_n \xrightarrow{p.s.} 0$.

2. Montrer que X_n ne converge pas dans L^1 vers 0.

Ex. 8.5. Soit $(X_n)_{n \geq 1}$ une suite de v.a. indépendante, où $X_n \sim \mathcal{B}(\frac{1}{n})$.

1. Montrer que $X_n \xrightarrow{L^2} 0$.

2. Montrer que $(X_n)_{n \geq 0}$ ne converge pas p.s. vers 0.

On a par contre la propriété suivante qui résulte directement du TCD.

Proposition 8.9. a) Si $X_n \xrightarrow{n \rightarrow \infty} X$ p.s., et si $|X_n| \leq Z$ pour tout n , où $E[Z] < \infty$, alors $X_n \xrightarrow{L^1} X$.
b) Si $X_n \xrightarrow{n \rightarrow \infty} X$ p.s., et si $|X_n| \leq Z$ pour tout n , où $E[Z^2] < \infty$, alors $X_n \xrightarrow{L^2} X$.

8.4 Convergence en loi

8.4.1 Définition et propriétés

Définition. La suite $(X_n)_{n \geq 1}$ converge vers X en loi (ou en distribution), noté $X_n \xrightarrow{(d)} X$, si

$$\forall \varphi : \mathbb{R} \rightarrow \mathbb{R} \text{ continue bornée} \quad \lim_{n \rightarrow \infty} E[\varphi(X_n)] = E[\varphi(X)].$$

Remarque. Notons 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 $X_n \xrightarrow{(d)} \mu$ si $X_n \xrightarrow{(d)} X$ 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 différente des précédentes.

Remarque. Le fait que φ soit continue bornée est essentiel dans la définition. D'une part, la continuité assure que la convergence en loi est compatible avec la convergence de suites réelles (si $x_n \rightarrow x$ et φ n'est pas continue en x on n'aura pas $\varphi(x_n) \rightarrow \varphi(x)$). D'autre part, la bornitude de φ assure que les espérances $E[\varphi(X_n)]$, $E[\varphi(X)]$ sont bien définies. En particulier, la convergence en loi n'implique pas la convergence $E[X_n] \rightarrow E[X]$!

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

Théorème 8.10. 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, $E[\varphi(X_n)] \xrightarrow{n \rightarrow \infty} E[\varphi(X)]$.

(iii) pour toute fonction $\varphi : \mathbb{R} \rightarrow \mathbb{R}$ continue à support compact, $E[\varphi(X_n)] \xrightarrow{n \rightarrow \infty} E[\varphi(X)]$.

(iv) en tout point $t \in \mathbb{R}$ où la fonction de répartition F_X est continue (c'est-à-dire tel que $P(X = t) = 0$),

$$F_{X_n}(t) \xrightarrow{n \rightarrow \infty} F_X(t).$$

(v) pour tout réel t , $\Phi_{X_n}(t) \xrightarrow{n \rightarrow \infty} \Phi_X(t)$.

Le point (iv) s'avère notamment utile quand X_n est définie comme min ou max de v.a. indépendantes. L'équivalence de (v) avec (i) est un théorème important (et difficile) appelée **théorème de Lévy**. Le point (v) s'avère notamment utile notamment quand X_n est définie comme somme de v.a. indépendantes.

La convergence en loi est plus faible que toutes les autres, en effet (rappelons que la convergences p.s. et la convergence L^1 impliquent toutes les deux la convergence en probabilité) :

Proposition 8.11. a) Si $X_n \xrightarrow{P} X$, alors $X_n \xrightarrow{(d)} X$.

b) Dans le cas où X est constante égale à $c \in \mathbb{R}$, $X_n \xrightarrow{P} c$ équivaut à $X_n \xrightarrow{(d)} c$.

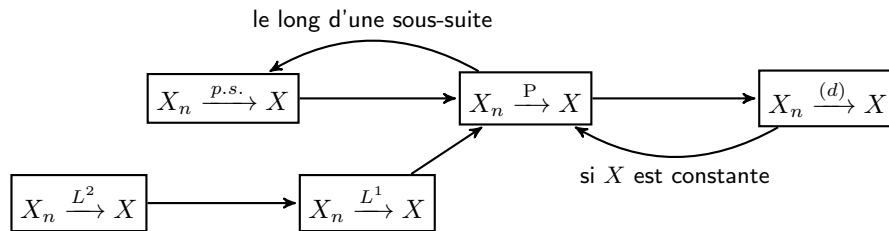
On laisse la démonstration en exercice. Signalons un cas particulier :

Proposition 8.12. Si les v.a. X_n sont à valeurs dans $\mathbb{Z}$, alors $X_n \xrightarrow{(d)} X$ si et seulement si

$$\forall x \in \mathbb{Z} \quad P(X_n = x) \xrightarrow{n \rightarrow \infty} P(X = x).$$

Remarque. Par cette proposition, la propriété citée lors de la définition de la loi de Poisson p. 53 peut se voir comme une convergence en loi : si X_n suit la loi $\mathcal{B}(n, p_n)$ où $np_n \xrightarrow{n \rightarrow \infty} \lambda > 0$, alors $X_n \xrightarrow{(d)} \mathcal{P}(\lambda)$.

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



8.4.2 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

$$\Phi(t) = e^{itm - t^2 \sigma^2 / 2}.$$

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

Théorème 8.13 (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 $\sigma^2 = \text{Var}(X_1)$. Alors on a

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

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 $\sigma > 0$.

Autres formulations possibles. Signalons une autre écriture de cet énoncé :

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

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

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

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) du Théorème 8.10 :

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

(On peut appliquer (iv) en tout $t \in \mathbb{R}$ car la loi normale a une fonction de répartition continue.)

Démonstration. On utilise l'équivalence (v) (théorème de Lévy) du Théorème 8.10. 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 ,

$$\Phi_{\frac{X_1 + \dots + X_n}{\sqrt{n}}}(t) = E \left[e^{it \frac{X_1 + \dots + X_n}{\sqrt{n}}} \right] = E \left[e^{it \frac{X_1}{\sqrt{n}}} \dots e^{it \frac{X_n}{\sqrt{n}}} \right] = E \left[e^{it \frac{X_1}{\sqrt{n}}} \right] \dots E \left[e^{it \frac{X_n}{\sqrt{n}}} \right] = E \left[e^{it \frac{X_1}{\sqrt{n}}} \right]^n$$

car $X_1, \dots, X_n$ sont indépendantes et de même loi. On en déduit donc que

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

Notons maintenant (voir Section 7.4) 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 $\frac{t}{\sqrt{n}} \rightarrow 0$, 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

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

autrement dit

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

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