

0. RAPPELS DE THÉORIE DE L'INTÉGRATION

Notations

Soit E un ensemble.

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

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

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

1 Intégrale de Riemann

L'intégrale introduite en L1 ou en classe prépa est l'intégrale dite « de Riemann », qui se définit (sur un segment) comme limite des sommes de Riemann, à supposer que celles-ci convergent à mesure que la subdivision se raffine.

1.1 Définition

- Soit $a < b$ des réels et $n \in \mathbb{N}^*$. Une **subdivision de $[a,b]$ en n sous-intervalles** est un $(n+1)$ -uplet $\sigma = (a_0, \dots, a_n)$, où $a = a_0 < a_1 < \dots < a_{n-1} < a_n = b$; le **pas** de cette subdivision σ est le réel $\max_{0 \leq i < n} (a_{i+1} - a_i)$. Une **subdivision pointée de $[a,b]$ en n sous-intervalles** est un couple (σ, τ) où $\sigma = (a_0, a_1, \dots, a_n)$ est une subdivision de $[a,b]$ en n sous-intervalles et $\tau = (t_1, \dots, t_n)$ est un n -uplet de points de $[a,b]$ tels que, pour $i = 1, \dots, n$, $t_i \in [a_{i-1}, a_i]$. Soit $f : [a,b] \rightarrow \mathbb{R}$ une fonction. Pour toute subdivision pointée $(\sigma, \tau) = ((a_0, \dots, a_n), (t_1, \dots, t_n))$ de $[a,b]$, on définit la **somme de Riemann**

$$S(f, (\sigma, \tau)) = \sum_{i=0}^{n-1} f(t_i)(a_{i+1} - a_i).$$

f est **intégrable au sens de Riemann** (sur $[a,b]$) s'il existe un réel s tel que, pour toute suite $(\sigma_n, \tau_n)_{n \in \mathbb{N}}$ de subdivisions pointées dont le pas tend vers 0, $S(f, (\sigma_n, \tau_n))$ converge vers s . La limite commune s est alors appelée l'**intégrale** de f (au sens de Riemann), et on note $s = \int_a^b f(t) dt$.

On utilise parfois une définition apparemment plus forte mais en fait équivalente : pour tout $\varepsilon > 0$, il existe $\delta > 0$ tel que, pour toute subdivision pointée (σ, τ) de $[a,b]$ de pas inférieur à δ , $|S(f, (\sigma, \tau)) - s| < \varepsilon$.

- Remarquons que les fonctions intégrables au sens de Riemann sur $[a,b]$ sont bornées sur $[a,b]$. On peut néanmoins introduire une **intégrale généralisée** qui donne un sens à certaines intégrales de fonctions non bornées, ou sur un intervalle non borné : on définit ainsi pour $\alpha < b$ ou $\alpha = -\infty$,

$$\int_{] \alpha, b]} f = \lim_{a \rightarrow \alpha^+} \int_a^b f$$

lorsque cette limite existe, et de même pour des intervalles de la forme $[a, \beta[$ ou $] \alpha, \beta[$ avec $-\infty \leq \alpha < \beta \leq \infty$. On peut également étendre la définition aux fonctions à valeurs dans \mathbb{C} ou dans \mathbb{R}^d , en intégrant séparément chaque composante :

$$\int (f + ig) = \int f + i \int g \quad \text{et} \quad \int \begin{pmatrix} f_1 \\ \vdots \\ f_n \end{pmatrix} = \begin{pmatrix} \int f_1 \\ \vdots \\ \int f_n \end{pmatrix}$$

Enfin, on peut l'étendre aux fonctions définies sur \mathbb{R}^d en introduisant une notion de subdivision d'un domaine en pavés et en introduisant les sommes de Riemann correspondantes.

• Signalons une caractérisation des fonctions intégrables au sens de Riemann sur $[a, b]$: ce sont les fonctions bornées sur $[a, b]$ dont l'ensemble des points de discontinuité est négligeable. (cf. Gourdon par exemple)

1.2 Limitations

Cette définition de l'intégrale correspond à l'intuition : on approche l'aire sous la courbe par l'aire d'une union de rectangles, et l'intégrale obtenue à la limite peut se voir comme une somme de quantités infinitésimales « $f(x)dx$ » correspondant à l'aire de chacun de ces rectangles.

Toutefois, cette définition se prête mal aux généralisations. Citons quelques-unes de ses limitations :

- la définition sur les intervalles non bornés est possible (intégrale généralisée, voir ci-dessus), mais indirecte
- certaines fonctions bornées très irrégulières ne sont pas intégrables (comme la fonction indicatrice $\mathbf{1}_{\mathbb{Q}}$)
- les théorèmes d'échange entre limite simple et intégrale sont peu satisfaisants (la notion naturelle pour l'intégrale de Riemann est la convergence uniforme, très restrictive)
- la définition ne s'étend pas facilement à d'autres espaces de départ (il faut pouvoir les subdiviser en parties « simples à mesurer » comme les pavés de \mathbb{R}^d).

L'intégrale de Lebesgue peut être vue comme une façon de répondre à cette dernière limitation. Elle s'avère répondre aux précédentes aussi, et même fournir un cadre mathématique à la théorie des probabilités.

Pour l'intégrale de Riemann, tout part de l'idée que l'on peut facilement définir l'intégrale des fonctions en escalier sur $[a, b]$,

$$\varphi = \sum_{i=1}^n \alpha_i \mathbf{1}_{[a_{i-1}, a_i]},$$

par

$$\int_a^b \varphi = \sum_{i=1}^n \alpha_i (a_i - a_{i-1}) = \sum_{i=1}^n \alpha_i \text{longueur}([a_{i-1}, a_i]),$$

puis l'on dit que $f : [a, b] \rightarrow \mathbb{R}$ est intégrable si les intégrales de fonctions en escalier qui encadrent f peuvent être rendues arbitrairement proches.

Pour intégrer des fonctions $f : E \rightarrow \mathbb{R}$, en l'absence de notion d'« intervalle » sur E , il est dès lors naturel d'essayer d'étendre le principe précédent en partant de « fonctions simples » qui soient de la forme

$$\varphi = \sum_{i=1}^n \alpha_i \mathbf{1}_{A_i},$$

où A_1, \dots, A_n sont des parties disjointes de E , pour lesquelles on définirait

$$\int_E \varphi = \sum_{i=1}^n \alpha_i \text{mesure}(A_i),$$

pour une « mesure » à préciser (sur \mathbb{R} , ce serait la longueur), et de là on définirait $\int_E f$ de même que sur \mathbb{R} .

Si cette idée, à l'origine de l'intégrale de Lebesgue, est simple, la mise en œuvre de la notion de mesure s'avère plus délicate que l'on pourrait s'y attendre. En effet, même sur \mathbb{R} , définir la longueur d'une partie quelconque est une tâche impossible si l'on souhaite que la longueur soit additive (pour A et B disjoints, la « longueur » de $A \cup B$ devrait être la somme de celles de A et B). On ne va donc considérer comme « mesurables » que certaines parties de \mathbb{R} , dont l'ensemble vérifie des propriétés de stabilité qui mènent à la définition de **tribu**.

2 Tribu, mesure, intégrale de Lebesgue

2.1 Espace mesurés

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

2.1.1 Tribus

Définition

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

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

Les conséquences suivantes sont aussi importantes que la définition :

Propriétés

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

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

Exemples.

- $\mathcal{P}(E)$ est la **tribu discrète** sur E .
- $\{\emptyset, E\}$ est la **tribu grossière** sur E .
- La plupart des exemples importants viennent de la définition suivante, qui découle du fait qu'une intersection quelconque de tribus est une tribu :

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}** .

- Sur \mathbb{R}^d , la **tribu borélienne** est la tribu engendrée par les ouverts. On la note $\mathcal{B}(\mathbb{R}^d)$. On peut vérifier que $\mathcal{B}(\mathbb{R}^d)$ est aussi la tribu engendrée par les intervalles de \mathbb{R}^d . Ses éléments sont les **boréliens**. Cette définition s'étend à d'autres espaces topologiques : la tribu borélienne d'un espace topologique (E, \mathcal{T}) est la tribu engendrée par l'ensemble \mathcal{T} des ouverts.

Tout sous-ensemble de \mathbb{R} construit à partir d'intervalles à l'aide des opérations de passage au complémentaire, d'union dénombrable et d'intersection dénombrable, est un borélien. Ceci recouvre tous les sous-ensembles de \mathbb{R} que l'on construit explicitement. Il existe toutefois des parties de \mathbb{R} non boréliennes.

2.1.2 Mesures

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

Définition

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

- (i) $\mu(\emptyset) = 0$
- (ii) pour toute suite $(A_n)_{n \in \mathbb{N}}$ de parties mesurables **disjointes**, $\mu\left(\biguplus_{n \in \mathbb{N}} A_n\right) = \sum_{n \in \mathbb{N}} \mu(A_n)$.

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

Les conséquences suivantes sont aussi importantes que la définition :

Propriétés

- a) Si $A_1, \dots, A_n \in \mathcal{A}$ sont disjoints, alors $\mu(A_1 \uplus \dots \uplus A_n) = \mu(A_1) + \dots + \mu(A_n)$.
- b) Si $A, B \in \mathcal{A}$ et $A \subset B$, alors $\mu(A) \leq \mu(B)$ et, si $\mu(A) < \infty$, alors $\mu(B \setminus A) = \mu(B) - \mu(A)$.
- c) Pour tous $A, B \in \mathcal{A}$, et $\mu(A \cap B) < \infty$, alors $\mu(A \cup B) = \mu(A) + \mu(B) - \mu(A \cap B)$.
- d) Si $(A_n)_n$ est une suite croissante de parties mesurables, alors $\mu\left(\bigcup_n A_n\right) = \limup_n \mu(A_n)$.
- e) Si $(A_n)_n$ est une suite décroissante de parties mesurables, et $\mu(A_0) < \infty$, alors $\mu\left(\bigcap_n A_n\right) = \limdown_n \mu(A_n)$.
- f) Pour toute suite $(A_n)_n$ de parties mesurables, $\mu\left(\bigcup_n A_n\right) \leq \sum_n \mu(A_n)$. (**sous-additivité**)

Exemples.

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

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

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

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

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

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

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

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

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

Définition-Théorème

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

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

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

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

Propriétés

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

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

où $a + A = \{a + x \mid x \in A\}$.

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

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

où $tA = \{tx \mid x \in A\}$.

On aura constamment besoin de caractériser des mesures, autrement dit de montrer que deux mesures sont égales entre elles. Pour cela, il suffit de les comparer sur une famille qui engendre la tribu, à condition qu'elle soit stable par intersections finies. Citons juste un cas particulier important :

Proposition

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

Finissons par quelques définitions relatives aux ensembles de mesure nulle.

Définition

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

- Si $A \in \mathcal{A}$ est tel que $\mu(A) = 0$, on dit que A est **négligeable**.
On peut préciser « μ -négligeable », ou « négligeable pour la mesure μ », si la mesure μ n'est pas claire d'après le contexte.
- Si une propriété P_x est vraie pour tout $x \in A$, où A^c est négligeable pour la mesure μ , on dit que P_x est vraie pour **presque tout** x , ou encore que P est vraie **presque partout**.
On peut préciser « μ -presque partout », ou « presque partout pour la mesure μ », si la mesure μ n'est pas claire d'après le contexte. Sauf précision contraire, sur \mathbb{R}^d , « presque partout » fait référence à la mesure de Lebesgue λ_d .

Par sous-additivité des mesures, l'union d'une famille dénombrable d'ensembles négligeables est négligeable.

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 vérifie facilement que l'on a aussi

$$(\text{Supp}(\mu))^c = \bigcup_{\substack{O \text{ ouvert,} \\ \mu(O)=0}} O,$$

ce qui montre que le support de μ est un fermé. Par densité des rationnels, on peut aussi restreindre la réunion précédente à l'ensemble *dénombrable* des boules $B(x, r)$ où $x \in \mathbb{Q}^d$, $r \in \mathbb{Q} \cap]0, +\infty[$ et telles que $\mu(B(x, r)) = 0$, d'où il résulte par sous-additivité que $\text{Supp}(\mu)^c$ est μ -négligeable. Le support de μ est donc le plus petit fermé dont le complémentaire est μ -négligeable : c'est le plus petit fermé qui « porte toute la masse » de μ .

2.1.3 Fonctions mesurables

Définition

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

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

On rappelle que si $f : \mathbb{R}^d \rightarrow \mathbb{R}^{d'}$ est continue, pour tout ouvert B de $\mathbb{R}^{d'}$, $f^{-1}(B)$ est ouvert. De plus, on vérifie que l'ensemble des parties $B \in \mathcal{B}(\mathbb{R}^{d'})$ telles que $f^{-1}(B)$ est mesurable forme une tribu, ce qui entraîne que :

Proposition

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

Remarques sur $\overline{\mathbb{R}}$. Il sera très souvent utile de considérer des fonctions à valeurs dans la droite réelle achevée $\overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty, +\infty\} = [-\infty, +\infty]$, ou dans $\overline{\mathbb{R}}_+ = \mathbb{R}_+ \cup \{+\infty\} = [0, +\infty]$.

- $\overline{\mathbb{R}}$ (et $\overline{\mathbb{R}}_+$) est un espace métrique compact. On peut donc définir sa tribu borélienne. En fait, $\mathcal{B}(\overline{\mathbb{R}})$ est composée des ensembles de la forme A , $A \cup \{+\infty\}$, $A \cup \{-\infty\}$ et $A \cup \{-\infty, +\infty\}$ pour $A \in \mathcal{B}(\mathbb{R})$.
- Ainsi, une fonction $f : E \rightarrow \overline{\mathbb{R}}$ est mesurable si et seulement si $f^{-1}(\{+\infty\})$, $f^{-1}(\{-\infty\})$ et $f^{-1}(A)$ sont des ensembles mesurables pour tout $A \in \mathcal{B}(\mathbb{R})$.
- Une fonction mesurable $f : E \rightarrow \mathbb{R}$ est donc mesurable aussi en tant que fonction $E \rightarrow \overline{\mathbb{R}}$.
- On utilisera les opérations naturelles suivantes sur $\overline{\mathbb{R}}$: $\infty + \infty = \infty$, $\infty + a = \infty$ pour tout $a \in \mathbb{R}$, $\infty \times a = \infty$ si $a > 0$ et $\infty \times a = -\infty$ si $a < 0$, $\infty \times \infty = \infty$, $\infty \times (-\infty) = -\infty$. En revanche, on ne définit pas les opérations $\infty - \infty$ et $0 \times \infty$ en général. On sera amené plus bas à choisir $0 \times \infty = \infty \times 0 = 0$ dans la définition de l'intégrale, et *uniquement* dans cette définition.

Propriétés

L'espace des fonctions mesurables de (E, \mathcal{A}) dans $(\overline{\mathbb{R}}, \mathcal{B}(\overline{\mathbb{R}}))$ est stable par :

- a) addition (si $f, g : E \rightarrow \overline{\mathbb{R}}$ sont mesurables, alors $f + g$ aussi)
- b) multiplication (si $f, g : E \rightarrow \overline{\mathbb{R}}$ sont mesurables, alors fg aussi)
- c) passage au sup et à l'inf (si, pour tout n , $f_n : E \rightarrow \overline{\mathbb{R}}$ est mesurable, alors $\sup_n f_n$ et $\inf_n f_n$ aussi)
- d) valeur absolue (si $f : E \rightarrow \overline{\mathbb{R}}$ est mesurable, alors $|f|$ aussi)
- e) passage à la lim inf, lim sup et donc à la limite (si, pour tout n , $f_n : E \rightarrow \overline{\mathbb{R}}$ est mesurable, alors $\liminf_n f_n$ et $\limsup_n f_n$ sont mesurables; et si $f_n(x) \rightarrow f(x)$ pour tout $x \in E$, alors f est mesurable)

Toute fonction $\mathbb{R}^d \rightarrow \overline{\mathbb{R}}$ obtenue à partir de fonctions continues par ces opérations est donc mesurable. Ceci inclut de fait toutes les fonctions que l'on peut construire. Enfin,

Proposition

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

2.2 Intégration par rapport à une mesure

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

2.2.1 Intégrale de fonctions mesurables positives

Définition

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

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

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

Ces fonctions forment un espace vectoriel.

Définition

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

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

(avec $0 \cdot \infty = 0$, c'est-à-dire que si $\alpha_i = 0$ et $\mu(A_i) = \infty$, on prend $\alpha_i \mu(A_i) = 0$ dans cette définition).

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

En particulier,

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

Propriétés

Soit g, h des fonctions étagées positives.

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

Définition

Soit $f : E \rightarrow [0, +\infty]$ mesurable. L'**intégrale de f par rapport à μ** est

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

NB. On utilise aussi les notations suivantes : $\int f d\mu = \int f(x)d\mu(x) = \int f(x)\mu(dx)$ et on peut spécifier \int_E . Dans la suite, de même que dans cette définition, une fonction « mesurable positive » est à valeurs dans $[0, +\infty]$. Notons que par la propriété b) ci-dessous, si $f = \infty \cdot \mathbf{1}_A$ (autrement dit, $f(x) = \infty$ si $x \in A$ et $f(x) = 0$ sinon) avec $\mu(A) = 0$ alors $\int f d\mu = “\infty \cdot 0” = 0$. Ainsi, pour $\alpha \in \mathbb{R}_+ \cup \{+\infty\}$, $\int \alpha \mathbf{1}_A d\mu = \alpha \mu(A)$ avec “ $0 \cdot \infty = \infty \cdot 0 = 0$ ”.

Propriétés

Soit f, g des fonctions mesurables positives.

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

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

Théorème (Théorème de convergence monotone (TCM))

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

$$\int \limup_n f_n d\mu = \limup_n \int f_n d\mu.$$

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

Lemme

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

Propriétés

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

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

Corollaire (Théorème de convergence monotone pour les séries positives)

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

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

Proposition (Inégalité de Markov)

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

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

Corollaire

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

a) Si $\int f d\mu < \infty$, alors $f < \infty$ presque partout.

b) $\int f d\mu = 0$ si, et seulement si $f = 0$ presque partout.

c) Si $f = g$ presque partout, alors $\int f d\mu = \int g d\mu$.

Théorème (Lemme de Fatou)

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

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

2.2.2 Fonctions intégrables

Définition

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

On pose alors

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

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

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

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

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

Propriétés

a) Pour toute $f \in \mathcal{L}^1(E, \mathcal{A}, \mu)$, $\left| \int f d\mu \right| \leq \int |f| d\mu$

b) $\mathcal{L}^1(E, \mathcal{A}, \mu)$ est un espace vectoriel, et $f \mapsto \int f d\mu$ est une application linéaire de $\mathcal{L}^1(E, \mathcal{A}, \mu)$ dans \mathbb{R} .

c) Pour $f, g \in \mathcal{L}^1(E, \mathcal{A}, \mu)$, si $f \leq g$, alors $\int f d\mu \leq \int g d\mu$.

d) Pour $f, g \in \mathcal{L}^1(E, \mathcal{A}, \mu)$, si $f = g$ presque partout, alors $\int f d\mu = \int g d\mu$.

Théorème (Théorème de convergence dominée (TCD))

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

(i) $f_n(x) \rightarrow f(x)$ pour presque partout $x \in E$;

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

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

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

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

On peut alors donner une formule « concrète » de calcul de $\int f d\mu$ par approximation en subdivisant l'intervalle d'arrivée, de même que les sommes de Riemann subdivisent l'intervalle de départ :

Corollaire

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

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

on a

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

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

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

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

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

2.2.3 Exemples principaux

Intégrale par rapport à une mesure atomique (On dit que x est un **atome** de μ si $\{x\} \in \mathcal{A}$ et $\mu(\{x\}) > 0$)

Proposition

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

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

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

$$\int f d\mu = \sum_n \alpha_n f(x_n) \in [0, +\infty].$$

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

$$\int f d\mu = \sum_n \alpha_n f(x_n) \in \mathbb{R}.$$

En particulier, si μ_E est la mesure de comptage sur E et $f : E \rightarrow \mathbb{R}_+$, $\int f d\mu_E = \sum_{x \in E} f(x)$.

Intégrale par rapport à la mesure de Lebesgue (lien avec l'intégrale de Riemann) On note $\lambda = \lambda_1$.

Soit $a < b$.

Notons que toute fonction mesurable bornée $f : [a, b] \rightarrow \mathbb{R}$ est intégrable sur $[a, b]$ par rapport à la mesure de Lebesgue. En effet, si $|f(x)| \leq M$ pour tout $x \in [a, b]$, alors

$$\int_{[a, b]} |f| d\lambda \leq \int_{[a, b]} M d\lambda = M \int_{[a, b]} d\lambda = M\lambda([a, b]) = M|b - a| < \infty.$$

Théorème

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

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

(l'intégrale de droite étant l'intégrale au sens de Riemann)

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

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

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

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

Intégrale par rapport à une mesure à densité Soit (E, \mathcal{A}) un espace mesurable. On vérifie facilement que, si f est positive, $A \mapsto \int_A f d\mu$ est une mesure, d'où la définition :

Définition

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

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

NB. A est négligeable pour $f \cdot \mu$ dès que A est négligeable pour μ (ou que f est nulle sur A). Ceci caractérise les mesures à densité, cf. le théorème de Radon-Nikodym.

Proposition

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

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

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

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

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

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

2.2.4 Intégrales dépendant d'un paramètre

Les résultats suivants se déduisent rapidement du théorème de convergence dominée.

Théorème (Théorème de continuité sous l'intégrale)

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

On suppose que :

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

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

Alors la fonction

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

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

Théorème (Théorème de dérivation sous l'intégrale)

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

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

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

Alors la fonction

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

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

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

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

Deuxième remarque : Pour montrer que F est de classe $\mathcal{C}^2, \mathcal{C}^3, \dots$, on peut appliquer le théorème plusieurs fois. Pour montrer que F est de classe \mathcal{C}^∞ , on peut montrer par récurrence que F est de classe \mathcal{C}^n pour tout n , ou alors, si c'est le cas, montrer directement que F est analytique :

Théorème (Théorème d'holomorphic sous l'intégrale)

Soit $f : (z,x) \mapsto f(z,x)$ une fonction de $U \times E$ dans \mathbb{C} , où U est un ouvert de \mathbb{C} . On suppose que :

- **(mesurabilité)** pour tout $z \in U$, $x \mapsto f(z,x)$ est mesurable ;
- **(holomorphic)** pour μ -presque tout $x \in E$, $z \mapsto f(z,x)$ est holomorphe sur U , de dérivée notée $\frac{\partial f}{\partial z}$;
- **(domination de f !)** il existe une fonction $\varphi : E \rightarrow \mathbb{R}_+$ mesurable telle que $\int \varphi d\mu < \infty$ et

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

Alors la fonction

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

est holomorphic sur U , et, pour tout $z \in U$, la fonction $x \mapsto \frac{\partial f}{\partial z}(z,x)$ est intégrable et

$$F'(z) = \int \frac{\partial f}{\partial z}(z,x) d\mu(x).$$

2.3 Intégration sur un espace produit

Soit (E, \mathcal{A}, μ) et (F, \mathcal{B}, ν) deux espaces mesurés. On suppose dans cette partie que μ et ν sont σ -finies : il existe une suite croissante $(E_n)_n$ de parties de E telle que $E = \bigcup_n E_n$ et, pour tout n , $\mu(E_n) < \infty$ (de même pour ν).

2.3.1 Produit d'espaces mesurés

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

Définition

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

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

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

Proposition

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

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

Théorème

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

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

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

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

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

Proposition

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

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

2.3.2 Théorèmes de Fubini

Intégrer par rapport à $\mu \otimes \nu$ revient à intégrer par rapport à μ puis à ν , ou le contraire :

Théorème (Théorème de Fubini-Tonelli)

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

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

En conséquence de ce théorème, on pourra noter $\int_E \int_F f(x, y) d\nu(y) d\mu(x)$ sans parenthèses lorsque f est mesurable positive. En décomposant une fonction f de signe quelconque en $f = f_+ - f_-$, on obtient facilement :

Théorème (Théorème de Fubini-Lebesgue)

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

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

on a

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

2.4 Changements de variables

Mesure image On définit ici une façon de « transporter » une mesure d'un espace à un autre, par une fonction. Cette opération sera notamment importante en probabilités pour définir la loi d'une variable aléatoire. Soit $\varphi : (E, \mathcal{A}) \rightarrow (F, \mathcal{B})$ une application mesurable. On rappelle que μ est une mesure sur (E, \mathcal{A}) .

Définition

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

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

Théorème (Théorème de transfert)

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

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

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

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

Changements de variables dans \mathbb{R}^d

• Cas linéaire

Proposition

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

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

Autrement dit, si M est inversible,

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

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

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

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

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

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

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

a) Si f est positive, alors

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

et

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

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

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

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

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

I. FONDEMENTS DES PROBABILITÉS

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

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

1 Définitions

1.1 Espace de probabilités

Définition

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

$$\mathbb{P}(\Omega) = 1.$$

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

Une interprétation en est la suivante :

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

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

1.2 Variables aléatoires

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

Définition

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

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

La définition suivante est fondamentale.

Définition

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

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

On dit que X **suit la loi** μ (et on note parfois $X \sim \mu$) si la loi de X est μ .

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

$$\mathbb{P}_X(B) = \mathbb{P}(X \in B).$$

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

Si X est une variable aléatoire à valeurs dans \mathbb{R}^d , le **support** de X , noté $\text{Supp}(X)$, est le support de sa loi. C'est donc le plus petit fermé F tel que $X \in F$ p.s..

Lemme

Soit μ une loi de probabilités sur un espace (E, \mathcal{E}) . Il existe un espace de probabilités $(\Omega, \mathcal{F}, \mathbb{P})$ et une variable aléatoire $X : \Omega \rightarrow E$ tels que X a pour loi μ .

Démonstration : Il suffit de prendre $(\Omega, \mathcal{A}, \mathbb{P}) = (E, \mathcal{E}, \mu)$, et $X = \text{Id}$. ■

Variables discrètes et continues

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

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

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

Autrement dit, $\mathbb{P}_X = \sum_{x \in E} p_x \delta_x$. Donner la loi de X revient donc à calculer les valeurs p_x pour $x \in E$.

• **Variables aléatoires continues (ou à densité).** Dans le cas où X est à valeurs dans \mathbb{R}^d et la loi de X admet une densité f par rapport à la mesure de Lebesgue, on dit que X est une variable aléatoire **continue**, ou **à densité**, de densité f . Autrement dit, X a pour densité f si f est une fonction mesurable positive qui vérifie, pour tout $A \in \mathcal{B}(\mathbb{R}^d)$,

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

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

Propriétés

Si X est une variable aléatoire ayant pour densité f , alors

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

Démonstration : a) On a $\lambda_d(\{x\}) = 0$ et donc $\mathbb{P}(X = x) = \int_{\{x\}} f(t) dt = \lambda(\{x\})f(x) = 0$.

b) Si f est nulle sur $B \in \mathcal{B}(\mathbb{R}^d)$ (autrement dit, $f(x) = 0$ pour tout $x \in B$), alors $\mathbb{P}(X \in B) = \int_B f(t) dt = 0$ 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) Notons $A = \{x \in \mathbb{R}^d \mid f(x) > g(x)\}$. On a $\int \mathbf{1}_A(x) f(x) dx = \mathbb{P}(X \in A) = \int \mathbf{1}_A(x) g(x) dx$ car f et g sont des densités pour X , et ces intégrales sont finies (même, inférieures à 1) donc on peut soustraire :

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

or la fonction intégrée est strictement positive sur A et nulle ailleurs, d'où $\lambda(A) = 0$: presque partout, $f \leq g$. Par symétrie, presque partout $g \leq f$. Donc $f = g$ presque partout. ■

Tribu engendrée

Définition

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

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

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

Proposition

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

1.3 Espérance

Définition

Soit X une variable aléatoire réelle (ou à valeurs dans $\overline{\mathbb{R}}$). Son **espérance** est son intégrale par rapport à \mathbb{P} :

$$\mathbb{E}[X] = \int_{\Omega} X(\omega) d\mathbb{P}(\omega),$$

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 $\mathbb{E}[X] \in [0, \infty]$)
- si X est intégrable, c'est-à-dire $\mathbb{E}[|X|] = \int |X| d\mathbb{P} < \infty$.

On interprète $\mathbb{E}[X]$ comme la moyenne de la variable aléatoire X .

On a en particulier

$$\mathbb{E}[\mathbf{1}_A] = \mathbb{P}(A).$$

et, pour toute constante $c \in \mathbb{R}$, $\mathbb{E}[c] = c\mathbb{P}(\Omega) = c$.

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

Proposition (Théorème de transfert)

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

$$\mathbb{E}[\varphi(X)] = \int_G \varphi(x) d\mathbb{P}_X(x).$$

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

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

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

• 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),

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

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

Proposition

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

- (Inégalité de Markov) Pour tout $a > 0$, $\mathbb{P}(X \geq a) \leq \frac{\mathbb{E}[X]}{a}$.
- Si $\mathbb{E}[X] < \infty$, alors $X < \infty$ presque sûrement.
- Si $\mathbb{E}[X] = 0$, alors $X = 0$ presque sûrement.

Proposition

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

- (TCM) Si la suite $(X_n)_n$ est croissante et converge vers X , alors $\lim_{n \uparrow} \mathbb{E}[X_n] = \mathbb{E}[X]$.
- (TCM pour les séries) On a $\mathbb{E}\left[\sum_{n \geq 0} X_n\right] = \sum_{n \geq 0} \mathbb{E}[X_n]$.

Proposition (Théorème de convergence dominée)

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

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

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

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

Définition

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

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

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 $\mathbb{E}[X]$, de même que l'écart-type. L'écart-type de X a l'intérêt d'être homogène à X , c'est-à-dire que $\sigma(aX) = a\sigma(X)$ pour $a \geq 0$ (tandis que $\text{Var}(aX) = a^2 \text{Var}(X)$), de sorte qu'il pourra être pertinent de comparer les valeurs de $\sigma(X)$ à celles de $X - \mathbb{E}[X]$.

Proposition (Inégalité de Tchebychev)

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

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

Démonstration : Soit $a > 0$. Presque sûrement, on a $|X - \mathbb{E}[X]| \geq a$ si et seulement si $(X - \mathbb{E}[X])^2 \geq a^2$. Donc $\mathbb{P}(|X - \mathbb{E}[X]| \geq a) = \mathbb{P}((X - \mathbb{E}[X])^2 \geq a^2)$ et l'inégalité de Markov permet de conclure. ■

Remarque. L'intégration par rapport à une mesure de probabilité généralise la notion de combinaison linéaire convexe (qui correspond au cas d'une mesure à support fini $\mu = \sum_i \lambda_i \delta_{x_i}$). Ceci se traduit par la proposition :

Proposition

Soit X une variable aléatoire dans \mathbb{R}^d intégrable, à valeurs dans un convexe C . Alors

a) $\mathbb{E}[X] \in C$.

b) (Inégalité de Jensen) Pour toute fonction convexe $\varphi : C \rightarrow \mathbb{R}$, on a $\mathbb{E}[\varphi(X)_-] < \infty$ de sorte que l'on peut définir $\mathbb{E}[\varphi(X)] = \mathbb{E}[\varphi(X)_+] - \mathbb{E}[\varphi(X)_-]$, et on a

$$\varphi(\mathbb{E}[X]) \leq \mathbb{E}[\varphi(X)].$$

Démonstration : Pour simplifier, on ne prouve a) que si C est fermé. Alors C est égal à l'intersection des demi-espaces fermés qui le contiennent. Or par linéarité si un demi-espace contient X p.s., alors il contient aussi $\mathbb{E}[X]$, d'où la conclusion. On pourrait aussi utiliser la loi forte des grands nombres.

Passons à b). Par convexité de φ , son graphe admet un hyperplan d'appui au point $\mathbb{E}[X]$: il existe $a \in \mathbb{R}^d$, $b \in \mathbb{R}$ tels que, pour tout $x \in C$, $\varphi(x) \geq \langle a, x \rangle + b$, avec égalité en $x_0 = \mathbb{E}[X]$. On a donc p.s. $\varphi(X) \geq \langle a, X \rangle + b$. En particulier, $\varphi(X)_- \leq |a||X| + |b|$ d'où l'intégrabilité de $\varphi(X)_-$. Et, en prenant l'espérance, $\mathbb{E}[\varphi(X)] \geq \langle a, \mathbb{E}[X] \rangle + b$, où le terme de droite vaut $\varphi(\mathbb{E}[X])$ par le cas d'égalité. ■

1.4 Interprétations probabilistes

Intégration	Probabilités
espace mesuré (E, \mathcal{A}, μ)	espace de probabilités $(\Omega, \mathcal{A}, \mathbb{P})$
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 ($\mathbb{P}(A) = 1$)
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) $\mathbb{E}[X] = \int_{\Omega} X d\mathbb{P}$

2 Indépendance

2.1 Probabilité conditionnelle

Définition

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

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

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

On note que $A \mapsto \mathbb{P}(A|B)$ est une probabilité sur (Ω, \mathcal{A}) : c'est la mesure de densité $\frac{\mathbf{1}_B}{\mathbb{P}(B)}$ 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

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

Proposition

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

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

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

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

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

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

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

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

2.2 Événements indépendants

Définition

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

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

On généralise la définition :

Définition

Les événements A_1, \dots, A_n sont **indépendants** si, pour $k = 2, \dots, n$, pour tous $1 \leq i_1 < \dots < i_k \leq n$,

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

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

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

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

2.3 Variables aléatoires, tribus indépendantes

Définition

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

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

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

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

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

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

Théorème

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

$$\mathbb{P}_{(X_1, \dots, X_n)} = \mathbb{P}_{X_1} \otimes \dots \otimes \mathbb{P}_{X_n}.$$

On a alors, pour toute fonction $f : E_1 \times \dots \times E_n \rightarrow \mathbb{R}$ mesurable positive ou telle que $f(X_1, \dots, X_n)$ est intégrable,

$$\mathbb{E}[f(X_1, \dots, X_n)] = \int \cdots \int f(x_1, \dots, x_n) d\mathbb{P}_{X_1}(x_1) \cdots d\mathbb{P}_{X_n}(x_n),$$

où les n intégrales peuvent se calculer dans un ordre 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,

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

Et inversement, si, pour toutes fonctions f_1, \dots, f_n mesurables positives, ou bornées,

$$\mathbb{E}[f_1(X_1) \cdots f_n(X_n)] = \mathbb{E}[f_1(X_1)] \cdots \mathbb{E}[f_n(X_n)],$$

alors X_1, \dots, X_n sont indépendantes.

Le second point est la traduction du théorème de Fubini.

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

Définition

Si X et Y sont des variables aléatoires de carré intégrable, leur **covariance** est

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

et leur **corrélation** est

$$\text{Corr}(X, Y) = \frac{\text{Cov}(X, Y)}{\sigma(X)\sigma(Y)} \in [-1, 1],$$

(où les bornes sont conséquences de l'inégalité de Cauchy-Schwarz).

Corollaire

Si X et Y sont des variables aléatoires de carré intégrables indépendantes, alors elles sont **non corrélées** :

$$\text{Cov}(X, Y) = 0.$$

Si X_1, \dots, X_n sont des variables aléatoires de carré intégrables et **indépendantes**, alors

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

Démonstration : Le premier point vient directement de ce qui précède, et le deuxième en résulte puisque qu'un développement donne de façon générale

$$\text{Var}\left(\sum_{i=1}^n X_i\right) = \mathbb{E}\left[\left(\sum_{i=1}^n (X_i - \mathbb{E}[X_i])\right)^2\right] = \sum_{i=1}^n \text{Var}(X_i) + \sum_{1 \leq i \neq j \leq n} \text{Cov}(X_i, X_j)$$

Remarque. La définition d'indépendance de variables aléatoires peut aussi s'introduire en définissant l'indépendance d'une famille de tribus :

Définition

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

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

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

Propriétés

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

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

b) (« **Indépendance par paquets** ») Si (I_1, \dots, I_k) est une partition de $\{1, \dots, n\}$, alors les variables aléatoires

$$(X_i)_{i \in I_1}, \dots, (X_i)_{i \in I_k}$$

sont indépendantes.

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

L'indépendance par paquets s'étend au cas des tribus : si $\mathcal{G}_1, \dots, \mathcal{G}_n$ sont des tribus indépendantes, et (I_1, \dots, I_k) est une partition de $\{1, \dots, n\}$, alors les tribus

$$\sigma(\mathcal{G}_i; i \in I_1), \dots, \sigma(\mathcal{G}_i; i \in I_k)$$

sont indépendantes, où $\sigma(\mathcal{G}_i; i \in I_1)$ est la tribu engendrée par les tribus \mathcal{G}_i pour $i \in I_1$, etc.

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

Les propriétés a) et b) ci-dessus s'étendent alors à une famille infinie $(X_i)_{i \in I}$ (avec une partition $(I_j)_{j \in J}$ de I).

2.4 Cas des variables à densité dans \mathbb{R}^d

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

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

On s'intéresse au cas des variables à densité.

Proposition

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

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

Démonstration : Soit $1 \leq i \leq d$. Pour toute fonction mesurable $g : \mathbb{R} \rightarrow \mathbb{R}_+$, on a

$$\mathbb{E}[g(X_i)] = \int_{\mathbb{R}^d} g(x_i) d\mathbb{P}_{(X_1, \dots, X_d)}(x_1, \dots, x_d) = \int_{\mathbb{R}^d} g(x_i) f(x_1, \dots, x_d) dx_1 \cdots dx_d = \int_{\mathbb{R}} g(x_i) f_i(x_i) dx_i$$

où la dernière égalité vient du théorème de Fubini-Tonelli. Ceci montre que X_i a pour densité f_i .

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

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

Par le théorème de Fubini-Tonelli, pour deux fonctions mesurables $f : E \rightarrow \mathbb{R}_+$ et $g : F \rightarrow \mathbb{R}_+$, la mesure produit des mesures de densité f sur E et de densité g sur F est la mesure de densité $f \otimes g : (x, y) \mapsto f(x)g(y)$. On en déduit que l'indépendance entre les marginales se traduit par le fait que la densité du vecteur aléatoire est à variables séparées, c'est-à-dire un produit de fonctions ne dépendant que d'une variable :

Proposition

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

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

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

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

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

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

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

Image par un \mathcal{C}^1 -difféomorphisme. Soit $X = (X_1, \dots, X_d)$ un vecteur aléatoire ayant pour densité f . On suppose que, presque sûrement, $X \in U$, où U est un ouvert de \mathbb{R}^d . Ceci équivaut à dire que f est nulle presque partout hors de U .

Soit $\varphi : U \rightarrow V$ un \mathcal{C}^1 -difféomorphisme (où V est un ouvert de \mathbb{R}^d). Alors la variable aléatoire

$$Y = (Y_1, \dots, Y_d) = \varphi(X_1, \dots, X_d)$$

admet une densité.

C'est une conséquence du théorème de changement de variable, qui permet aussi de calculer cette densité : pour toute fonction mesurable positive $g : \mathbb{R}^d \rightarrow \mathbb{R}$, on a

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

La dernière expression est $\mathbb{E}[g(Z_1, \dots, Z_d)]$ où (Z_1, \dots, Z_d) est un vecteur aléatoire (quelconque) ayant pour densité

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

Comme l'égalité $\mathbb{E}[g(Y_1, \dots, Y_d)] = \mathbb{E}[g(Z_1, \dots, Z_d)]$ vaut pour toute fonction mesurable positive g , on en déduit que (Y_1, \dots, Y_d) et (Z_1, \dots, Z_d) ont même loi, donc la fonction ci-dessus est aussi la densité de (Y_1, \dots, Y_d) :

$$f_{(Y_1, \dots, Y_d)} : (y_1, \dots, y_d) \mapsto f(\varphi^{-1}(y_1, \dots, y_d)) |J_{\varphi^{-1}}(y_1, \dots, y_d)|.$$

Loi de la somme de 2 variables aléatoires indépendantes à densité. Soit X, Y deux variables aléatoires réelles indépendantes ayant des densités f_X et f_Y . On définit $Z = X + Y$. Pour toute fonction mesurable positive $g : \mathbb{R} \rightarrow \mathbb{R}$, on a

$$\begin{aligned} \mathbb{E}[g(Z)] &= \mathbb{E}[g(X + Y)] = \int_{\mathbb{R}^2} g(x + y) d\mathbb{P}_{(X, Y)}(x, y) = \int_{\mathbb{R}^2} g(x + y) f_X(x) f_Y(y) dx dy \\ &= \int_{\mathbb{R}} \left(\int_{\mathbb{R}} g(x + y) f_Y(y) dy \right) f_X(x) dx = \int_{\mathbb{R}} \left(\int_{\mathbb{R}} g(z) f_Y(z - x) dz \right) f_X(x) dx \\ &= \int_{\mathbb{R}} g(z) \left(\int_{\mathbb{R}} f_X(x) f_Y(z - x) dx \right) dz, \end{aligned}$$

en utilisant le changement de variable $y \mapsto z = x + y$, et le théorème de Fubini-Tonelli. Ceci montre que Z a pour densité la fonction

$$f_Z : z \mapsto \int_{\mathbb{R}} f_X(x) f_Y(z-x) dx,$$

autrement dit le produit de convolution de f_X et f_Y : si X et Y sont indépendantes, $f_{X+Y} = f_X * f_Y$.

3 Lois classiques

Lois discrètes

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

$$\mathbb{P}(X = x) = 1.$$

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

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

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

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

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

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

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

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

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

Proposition

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

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

suit la loi binomiale de paramètres n et p .

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

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

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

Proposition

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

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

suit la loi géométrique de paramètre p .

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

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

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

Proposition

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

de paramètres n et p_n , alors

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

Lois continues

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

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

- En général, si $A \in \mathcal{B}(\mathbb{R}^d)$ avec $\lambda_d(A) > 0$, la **loi uniforme sur** A est la loi sur \mathbb{R}^d (notée $\mathcal{U}(A)$) de densité

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

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

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

Cette loi peut se voir comme une extension de la loi géométrique au cas continu (quand on aura la définition appropriée, on pourra dire que c'est la limite de la loi de $\frac{1}{n}N(\frac{\lambda}{n})$ quand $n \rightarrow \infty$, où $N(p)$ suit la loi $\mathcal{G}(p)$), et s'utilise de même pour modéliser des temps d'attente, pour des phénomènes « sans vieillissement » :

Proposition

Soit X une variable aléatoire à valeurs dans $]0, +\infty[$. Les propriétés suivantes sont équivalentes :

- (i) il existe $\lambda > 0$ tel que X suit la loi $\mathcal{E}(\lambda)$;
- (ii) la loi de X est « sans mémoire » : pour tous $s, t > 0$, $\mathbb{P}(X > s + t) = \mathbb{P}(X > s)\mathbb{P}(X > t)$.

La propriété (ii) s'écrit aussi $\mathbb{P}(X > s + t | X > s) = \mathbb{P}(X > t)$, ce qui s'interprète ainsi : avoir déjà attendu un temps s ne renseigne pas sur le temps qu'il reste à attendre.

- 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

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

L'appellation est justifiée en vérifiant que m est la moyenne et σ^2 la variance de cette loi.

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

Cette loi peut se voir comme une limite de la loi binomiale : ce sera la limite de la loi de $\frac{1}{\sqrt{n}}(2S_n - n)$ quand $n \rightarrow \infty$, si S_n suit la loi $\mathcal{B}(n, \frac{1}{2})$. Elle apparaît donc comme la loi approximative de la différence (normalisée) entre le nombre de piles et de faces sur un grand nombre de tirages d'une pièce équilibrée. De façon beaucoup plus générale, le rôle fondamental de la loi normale viendra du Théorème Central Limite.

II. REPRÉSENTATIONS FONCTIONNELLES DES LOIS DE PROBABILITÉ

On souhaite disposer d'une représentation des lois de variables aléatoires sur \mathbb{R}^d par des objets plus « simples » et qui soit adaptée à certains calculs. Précisons cela. On rappelle la proposition suivante :

Proposition

Soit X et Y des variables aléatoires à valeurs dans le même espace (E, \mathcal{B}) . Les propriétés suivantes sont équivalentes :

- (i) X et Y ont même loi (autrement dit, $\mathbb{P}_X = \mathbb{P}_Y$);
- (ii) pour tout $B \in \mathcal{B}$, $\mathbb{P}(X \in B) = \mathbb{P}(Y \in B)$;
- (iii) pour toute fonction mesurable positive $f : E \rightarrow \mathbb{R}$, $\mathbb{E}[f(X)] = \mathbb{E}[f(Y)]$;
- (iv) pour toute fonction mesurable bornée $f : E \rightarrow \mathbb{R}$, $\mathbb{E}[f(X)] = \mathbb{E}[f(Y)]$.

Démonstration : L'équivalence entre (i) et (ii) relève de la simple définition de \mathbb{P}_X . On voit immédiatement que (iii) et (iv) impliquent (ii) en considérant $f = \mathbf{1}_B$. On déduit (iv) de (iii) en appliquant (iii) à f_+ et f_- , qui sont positives et telles que $f = f_+ - f_-$. Il reste à montrer que (ii) implique (iii).

Supposons (ii) vrai. Alors (iii) est vérifié pour les fonctions indicatrices $f = \mathbf{1}_B$ où $B \in \mathcal{B}$. Par linéarité, (iii) est donc vérifié pour les fonctions étagées positives. Enfin, toute fonction mesurable positive est limite d'une suite croissante de fonctions étagées positives, ce qui permet d'obtenir (iii) à l'aide du théorème de convergence monotone. ■

Comme (ii) est une spécialisation de (iii) ou (iv) au cas $f = \mathbf{1}_B$, les énoncés (iii) et (iv) n'ont pas d'utilité pour montrer que X et Y ont même loi. Cependant, dans le cas où $E = \mathbb{R}^d$, on va voir qu'il suffit de vérifier (iii) ou (iv) *seulement* pour les fonctions f appartenant à une classe beaucoup plus restreinte. Par exemple :

Proposition

Soit X et Y des variables aléatoires à valeurs dans un même espace $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$. Les propriétés suivantes sont équivalentes :

- (i) X et Y ont même loi (autrement dit, $\mathbb{P}_X = \mathbb{P}_Y$);
- (ii') pour tout pavé ouvert $P =]a_1, b_1[\times \dots \times]a_d, b_d[$, $\mathbb{P}(X \in P) = \mathbb{P}(Y \in P)$;
- (iii') pour toute fonction $f : \mathbb{R}^d \rightarrow \mathbb{R}$ positive, de classe \mathcal{C}^∞ et à support compact, $\mathbb{E}[f(X)] = \mathbb{E}[f(Y)]$.

Démonstration : Par la proposition précédente, (i) implique directement (ii') et (iii').

• Supposons (ii'). On peut déduire (ii) du lemme de classe monotone (voir exercices) ; donnons une preuve utilisant un autre lemme important. D'abord, l'égalité $\mathbb{P}(X \in P) = \mathbb{P}(Y \in P)$ s'étend aux pavés semi-ouverts $P = [a_1, b_1[\times \dots \times [a_d, b_d[$. En effet, on a

$$P = \bigcap_n P_n, \quad \text{où } P_n =]a_1 - \frac{1}{n}, b_1[\times \dots \times]a_d - \frac{1}{n}, b_d[$$

d'où, en utilisant (ii'),

$$\mathbb{P}(X \in P) = \mathbb{P}_X(P) = \lim_n \downarrow \mathbb{P}_X(P_n) = \lim_n \downarrow \mathbb{P}_Y(P_n) = \mathbb{P}_Y(P) = \mathbb{P}(Y \in P).$$

Or tout ouvert de \mathbb{R}^d est réunion disjointe d'une famille dénombrable de pavés semi-ouverts, par exemple les pavés dyadiques semi-ouverts maximaux qu'il contient. On en déduit que l'égalité de \mathbb{P}_X et \mathbb{P}_Y sur les pavés semi-ouverts implique leur égalité sur les ouverts.

Pour en déduire leur égalité sur tous les boréliens, il reste alors à appliquer à \mathbb{P}_X et \mathbb{P}_Y le résultat suivant (partie 'inf') :

Lemme

Toute mesure finie μ sur $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ est régulière : pour tout $B \in \mathcal{B}(\mathbb{R}^d)$, $\mu(B) = \sup_{\substack{F \text{ fermé} \\ F \subset B}} \mu(F) = \inf_{\substack{G \text{ ouvert} \\ B \subset G}} \mu(G)$.

Ce lemme s'obtient en vérifiant que l'ensemble \mathcal{E} des boréliens pour lesquels le lemme est vrai est stable par passage au complémentaire (c'est direct), contient les fermés (le sup est évident, et pour l'inf on écrit B comme intersection décroissante des ouverts $G_n = B + B(0, \frac{1}{n})$), et est stable par union dénombrable (si $B_n \in \mathcal{E}$ et $B = \bigcup_n B_n$, alors pour tout $\varepsilon > 0$ il existe $F_n \subset B_n \subset G_n$ où F_n fermé, G_n ouvert et $\mu(G_n) - \mu(F_n) < \varepsilon 2^{-n}$ et N tel que $\mu(\bigcup_{n \geq N} F_n) \geq \mu(\bigcup_n F_n) - \varepsilon$, alors le fermé $F = \bigcup_{n < N} F_n$ et l'ouvert $G = \bigcup_n G_n$ sont tels que $F \subset B \subset G$ et $\mu(G) - \mu(F) \leq 2\varepsilon$, d'où $B \in \mathcal{E}$). De sorte que \mathcal{E} est une tribu incluse dans la tribu borélienne et contenant les fermés, donc c'est exactement la tribu borélienne.

• Supposons (iii'). Il suffit de montrer (ii'). Or, pour tout pavé ouvert P , la fonction indicatrice $\mathbf{1}_P$ est limite croissante d'une suite de fonctions positives, de classe \mathcal{C}^∞ et à support compact, de sorte que (ii') s'obtient par le théorème de convergence monotone. ■

1 Fonction de répartition (v.a. à valeurs dans \mathbb{R})

Pour les variables aléatoires à valeurs dans \mathbb{R} , il suffit de considérer les fonctions $\mathbf{1}_{]-\infty, t]}$, pour $t \in \mathbb{R}$.

Définition

La **fonction de répartition** d'une variable aléatoire X à valeurs dans \mathbb{R} est la fonction

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

Proposition

a) La loi de X est déterminée par sa fonction de répartition : si $F_X(t) = F_Y(t)$ pour tout $t \in \mathbb{R}$, alors X et Y ont même loi.

b) La fonction F_X est croissante, continue à droite, et admet pour limites 0 en $-\infty$ et 1 en $+\infty$.

c) Pour tout $t \in \mathbb{R}$, la limite de F_X en t à gauche est

$$F_X(t^-) = \mathbb{P}(X < t),$$

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

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

Démonstration : a) On suppose que $F_X = F_Y$. Pour tous $a < b$,

$$\mathbb{P}(X \in]a, b]) = \mathbb{P}(X < b) - \mathbb{P}(X \leq a) = \lim_{n \uparrow} \mathbb{P}(X \leq b - \frac{1}{n}) - \mathbb{P}(X \leq a)$$

car $\{X < b\} = \{X \leq a\} \cup \{X \in]a, b]\}$ et

$$\{X < b\} = \bigcup_{n \geq 1} \left\{ X \leq b - \frac{1}{n} \right\}.$$

Ainsi, les probabilités d'appartenance aux pavés (intervalles) ouverts sont données par F_X donc coïncident pour X et Y . On conclut à l'égalité entre les lois de X et de Y par la proposition précédente.

b) Pour tous $s < t$, on a $\{X \leq s\} \subset \{X \leq t\}$ donc

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

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

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

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

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

Ceci montre la continuité de F_X à droite en t .

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

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

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

$$\lim_{n \downarrow} F_X(-n) = \lim_{n \downarrow} \mathbb{P}(X \leq -n) = \mathbb{P}(\emptyset) = 0.$$

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

c) La preuve est très similaire à la précédente : pour tout $t \in \mathbb{R}$,

$$F_X(t^-) = \lim_{n \downarrow} \mathbb{P}\left(X \leq t - \frac{1}{n}\right) = \mathbb{P}\left(\bigcap \left\{t - \frac{1}{n}\right\}\right) = \mathbb{P}(X < t).$$

d) La preuve résulte de l'exercice 14 : soit F est une fonction vérifiant b), alors on définit $F^{-1} : u \mapsto \inf\{t \mid F(t) \geq u\}$ (inverse continue à gauche) et on vérifie que, si U suit la loi uniforme sur $[0,1]$, alors $X = F^{-1}(U)$ a pour fonction de répartition F . ■

La fonction de répartition a un lien simple avec la densité :

Proposition

- a) Si X a pour densité f_X alors, pour tout $t \in \mathbb{R}$, $F_X(t) = \int_{-\infty}^t f_X(u)du$. En particulier, F_X est continue.
→ Si f_X est continue, F_X est sa primitive « nulle en $-\infty$ »
- b) Si F_X est de classe \mathcal{C}^1 , alors X a pour densité $f_X = (F_X)'$.

Démonstration : a) La formule est conséquence directe de la définition d'une densité. La continuité résulte en fait du point c) de la proposition précédente : si X a une densité, alors $\mathbb{P}(X = t) = 0$ pour tout $t \in \mathbb{R}$.

b) La fonction $f_X = (F_X)'$ est bien une densité : elle est positive (F_X est croissante) et son intégrale sur \mathbb{R} vaut $\lim_{+\infty} F_X - \lim_{-\infty} F_X = 1 - 0 = 1$. De plus, pour tout $t \in \mathbb{R}$, Comme F_X est \mathcal{C}^1 , elle est égale à l'intégrale de sa dérivée : pour $u < t$,

$$F_X(t) - F_X(u) = \int_u^t f_X$$

d'où, pour $u \rightarrow -\infty$, $F_X(t) = \int_{-\infty}^t f_X$, ce qui montre, en comparant avec a), que X a même fonction de répartition qu'une variable aléatoire qui suit la loi de densité f_X . Il en résulte que c'est la loi de X : X a pour densité f_X . ■

NB. Souvent, F_X est continue sur \mathbb{R} , mais n'est pas dérivable en quelques points $a_1 < \dots < a_n$, et F_X' (prolongée par une valeur quelconque en a_1, \dots, a_n) est continue par morceaux (c'est-à-dire que F_X' est continue sur les intervalles ouverts délimités par les a_i , et avec limites à gauche et droite en a_i). La propriété s'étend alors.

Inversement (et plus simplement), si X admet une densité f_X continue par morceaux, alors F_X est continue sur \mathbb{R} , et est dérivable sur chacun des morceaux, de dérivée f_X .

Utilisations classiques

Loi image par une fonction croissante. Si X est à valeurs dans $I \subset \mathbb{R}$, et $\varphi : I \rightarrow J$ est une bijection strictement croissante, alors la fonction de répartition de $Y = \varphi(X)$ est donnée par :

$$\text{pour tout } y \in \mathbb{R}, \quad F_Y(y) = \mathbb{P}(Y \leq y) = \mathbb{P}(\varphi(X) \leq y) = \mathbb{P}(X \leq \varphi^{-1}(y)) = F_X(\varphi^{-1}(y)).$$

En particulier, si X a une densité f_X continue par morceaux, et si φ est dérivable, de dérivée non nulle, alors F_Y est continue, et dérivable (sauf aux images par φ des points de discontinuité de f_X), donc Y admet une densité f_Y donnée par :

$$\text{pour tout } y \in \mathbb{R}, \quad f_Y(y) = (F_Y)'(y) = \frac{1}{\varphi'(\varphi^{-1}(y))} f_X(\varphi^{-1}(y)).$$

(Un changement de variable fournirait le même résultat, mais l'utilisation de la fonction de répartition est parfois plus rapide)

Loi du maximum ou du minimum d'une famille de variables aléatoires indépendantes. Soit X_1, \dots, X_n des variables aléatoires réelles indépendantes. On note

$$m = \min(X_1, \dots, X_n) \quad \text{et} \quad M = \max(X_1, \dots, X_n).$$

La fonction de répartition de M a une expression simple : pour tout $t \in \mathbb{R}$,

$$F_M(t) = \mathbb{P}(M \leq t) = \mathbb{P}(\max(X_1, \dots, X_n) \leq t) = \mathbb{P}(X_1 \leq t, \dots, X_n \leq t) = F_{X_1}(t) \cdots F_{X_n}(t).$$

Si les X_i ont des densités continues par morceaux, on voit que F_M aussi, et on l'obtient en dérivant. Pour m , on se ramène à cette même méthode :

$$\begin{aligned} F_m(t) &= \mathbb{P}(m \leq t) = 1 - \mathbb{P}(m \geq t) = 1 - \mathbb{P}(\min(X_1, \dots, X_n) \geq t) = 1 - \mathbb{P}(X_1 \geq t, \dots, X_n \geq t) \\ &= 1 - (1 - F_{X_1}(t)) \cdots (1 - F_{X_n}(t)), \end{aligned}$$

d'où la même conclusion dans le cas à densité.

Méthode de simulation par inversion. Voir exercice 14.

Estimation d'une loi. Pour estimer la loi de X à partir de données, on utilisera la *fonction de répartition empirique* (cf. théorème de Glivenko-Cantelli). Elle interviendra aussi dans les tests statistiques de comparaison à une loi donnée.

Convergence en loi. Voir plus loin.

2 Fonction génératrice (v.a. à valeurs dans \mathbb{N})

Pour les variables aléatoires à valeurs dans \mathbb{N} , il suffit de considérer les fonctions $x \mapsto s^x$, pour $s \in [-1,1]$.

Définition

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

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

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

Comme la loi de X est caractérisée par les valeurs $\mathbb{P}(X = n)$ pour $n \in \mathbb{N}$, qui sont les coefficients de la série génératrice G_X , et que le rayon de convergence est strictement positif, on a :

Proposition

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

NB. Il suffirait d'avoir égalité sur un intervalle ouvert quelconque, par unicité du prolongement analytique.

Les moments de X se déduisent de G_X par dérivation :

Proposition

Pour tout $k \geq 0$,

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

Démonstration : Une série entière se dérive terme à terme dans son disque de convergence, donc pour tout $s \in]-1,1[$,

$$G_X^{(k)}(s) = \sum_{n=0}^{\infty} n(n-1)\cdots(n-k+1)s^{n-k}\mathbb{P}(X = n) = \mathbb{E}[X(X-1)\cdots(X-k+1)s^{X-k}]$$

ce qui converge vers $\mathbb{E}[X(X-1)\cdots(X-k+1)]$ quand $s \rightarrow 1^-$ par théorème de convergence monotone. ■

Ensuite, on obtient par exemple $\mathbb{E}[X^2] = \mathbb{E}[X(X-1)] + \mathbb{E}[X] = G_X''(1^-) + G_X'(1^-)$.

La propriété essentielle des fonctions génératrice est la suivante :

Proposition

Si X et Y sont des variables aléatoires indépendantes, à valeurs dans \mathbb{N} alors, pour tout $s \in [-1,1]$,

$$G_{X+Y}(s) = G_X(s)G_Y(s).$$

Démonstration : Il suffit d'écrire, pour $s \in [-1,1]$, grâce à l'indépendance entre X et Y ,

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

Utilisations classiques

Loi de la somme de v.a. indépendantes. Si X et Y sont à valeurs entières, et indépendantes, la proposition précédente permet de calculer G_{X+Y} à partir de G_X et G_Y . On peut alors reconnaître la fonction de répartition d'une loi connue, ou déduire de G_{X+Y} diverses propriétés de cette loi. Exemple : $X \sim \mathcal{P}(\lambda)$ et $Y \sim \mathcal{P}(\mu)$ (exercice).

Décomposition de loi. Inversement, on peut vouloir déterminer la loi de variables X et Y indépendantes, à valeurs entières, dont la somme suit une loi donnée, et vérifiant certaines hypothèses. Exemple :

- Montrer qu'il n'existe pas de façon de lester deux dés à 6 faces de sorte que, quand on additionne leurs résultats, les valeurs de 2 à 12 deviennent équiprobables.
- Montrer que l'on peut (et de façon unique) numéroter les faces de deux dés différemment (par des entiers ≥ 1) de telle sorte que la loi de la somme des résultats soit la même que pour deux dés usuels.

3 Transformée de Laplace (v.a. à valeurs dans \mathbb{R} , ou \mathbb{R}^d)

Définition

Soit X une variable aléatoire à valeurs dans \mathbb{R} . La **transformée de Laplace** de X est la fonction

$$\left| \begin{array}{l} \mathcal{L}_X : \mathbb{R} \rightarrow [0, +\infty] \\ t \mapsto \mathcal{L}_X(t) = \mathbb{E}[e^{tX}], \end{array} \right.$$

Pour X à valeurs dans \mathbb{R}^d , on peut aussi définir une fonction \mathcal{L}_X sur \mathbb{R}^d par : pour $t \in \mathbb{R}^d$, $\mathcal{L}(t) = \mathbb{E}[e^{\langle t, X \rangle}]$, où $\langle \cdot, \cdot \rangle$ est le produit scalaire usuel. (Dans la suite, on considère seulement le cas réel)

Attention, \mathcal{L}_X est définie sur \mathbb{R} (l'espérance d'une variable aléatoire positive est toujours définie) mais peut prendre la valeur $+\infty$. Il peut arriver que, pour tout $t \neq 0$, $\mathcal{L}_X(t) = \infty$. C'est pourquoi on ne peut pas toujours utiliser \mathcal{L}_X pour caractériser une loi. Notons

$$\mathcal{I}_X = \{t \in \mathbb{R} \mid \mathcal{L}_X(t) < \infty\}.$$

Proposition

- a) La fonction \mathcal{L}_X est convexe sur \mathbb{R} , et $\mathcal{L}_X(0) = 1$.
- b) L'ensemble \mathcal{I}_X est un intervalle contenant 0.
- c) Si X est à valeurs négatives, $\mathbb{R}_+ \subset \mathcal{I}_X$.
- d) Pour tout $c > 0$, $[-c, c] \subset \mathcal{I}_X$ si, et seulement si $\mathbb{E}[e^{c|X|}] < \infty$.

Démonstration : a) Pour tout $x \in \mathbb{R}$, $t \mapsto e^{tx}$ est convexe. Or \mathcal{L}_X est une intégrale de telles fonctions (par rapport à x), donc \mathcal{L}_X est convexe : pour tous $s, t \in \mathbb{R}$ et $\lambda \in [0, 1]$,

$$\begin{aligned} \mathcal{L}_X(\lambda s + (1 - \lambda)t) &= \mathbb{E}[e^{(\lambda s + (1 - \lambda)t)X}] = \int e^{(\lambda s + (1 - \lambda)t)x} d\mathbb{P}_X(x) \\ &\leq \int (\lambda e^{sx} + (1 - \lambda)e^{tx}) d\mathbb{P}_X(x) = \lambda \mathcal{L}_X(s) + (1 - \lambda)\mathcal{L}_X(t). \end{aligned}$$

- b) Notons que ceci vaut même si \mathcal{L}_X prend la valeur $+\infty$. En particulier, on note que si $\mathcal{L}_X(s) < \infty$ et $\mathcal{L}_X(t) < \infty$, alors $\mathcal{L}_X(\lambda s + (1 - \lambda)t) < \infty$. Ceci signifie que \mathcal{I}_X est convexe. C'est donc un intervalle.
- c) Si $X \leq 0$ et $t \geq 0$, alors $e^{tX} \leq 1$ donc $\mathcal{L}_X(t) \leq 1 < \infty$.
- d) On a évidemment, pour tout $c > 0$, $e^{c|X|} \leq e^{cX} + e^{-cX}$, donc $\mathbb{E}[e^{c|X|}] \leq \mathcal{L}_X(c) + \mathcal{L}_X(-c)$. De plus, $e^{cX} \leq e^{c|X|}$ et $e^{-cX} \leq e^{c|X|}$ d'où finalement $\max(\mathcal{L}_X(c), \mathcal{L}_X(-c)) \leq \mathbb{E}[e^{c|X|}] \leq \mathcal{L}_X(c) + \mathcal{L}_X(-c)$. Ceci donne d). ■

NB. On définit parfois $\mathcal{L}_X(t) = \mathbb{E}[e^{-tX}]$, par analogie avec la transformée de Laplace des fonctions, et de sorte que (par c)) la transformée de Laplace s'applique notamment aux variables aléatoires positives.

Proposition

Soit X et Y deux variables aléatoires réelles. On suppose qu'il existe un intervalle ouvert non vide I tel que, pour tout $t \in I$, $\mathcal{L}_X(t) = \mathcal{L}_Y(t) < \infty$. Alors X et Y ont même loi.

Démonstration : Pour un cas particulier plus simple, voir exercice 17. Pour la preuve générale, si $[a, b] \subset I$, on montre d'abord que \mathcal{L}_X et \mathcal{L}_Y s'étendent en des fonctions holomorphes sur la bande $[a, b] + i\mathbb{R} \subset \mathbb{C}$ (par la même formule) : utiliser le théorème d'holomorphie sous l'intégrale, avec la domination $|e^{zX}| = e^{\Re(z)X} \leq 1 + e^{bX}$. Elles s'étendent même sur $]0, b[+ i\mathbb{R}$, vu que $[0, b] \in \mathcal{I}_X$ par convexité de \mathcal{I}_X . Comme elles coïncident sur $[a, b]$, l'unicité du prolongement analytique montre qu'elles coïncident sur toute cette bande $]0, b[+ i\mathbb{R}$. Par ailleurs, le théorème de continuité sous l'intégrale montre que \mathcal{L}_X et \mathcal{L}_Y sont continues sur la bande fermée $[0, b] + i\mathbb{R}$ (même domination). En particulier, elles coïncident donc aussi sur l'axe $i\mathbb{R}$: pour tout $t \in \mathbb{R}$, $\mathbb{E}[e^{itX}] = \mathbb{E}[e^{itY}]$. Ceci nous ramène au cas des fonctions caractéristiques : voir partie suivante. ■

La preuve ci-dessus montre en particulier que \mathcal{L}_X est développable en série entière sur \mathcal{I}_X (si $\mathcal{I}_X \neq \{0\}$). Autour de 0, ce développement peut s'expliciter :

Proposition

On suppose qu'il existe $t > 0$ tel que $[-t, t] \subset \mathcal{I}_X$ (autrement dit, $\mathbb{E}[e^{t|X|}] < \infty$). Alors \mathcal{L}_X est développable en série entière sur $[-t, t]$, $\mathbb{E}[|X|^k] < \infty$ pour tout $k \in \mathbb{N}$, et

$$\mathcal{L}_X(t) = \sum_{k=0}^{\infty} \mathbb{E}[X^k] \frac{t^k}{k!}.$$

Démonstration : On a, par le théorème de convergence monotone (pour les séries à termes positifs),

$$\mathbb{E}[e^{t|X|}] = \mathbb{E}\left[\sum_{k=0}^{\infty} |X|^k \frac{t^k}{k!}\right] = \sum_{k=0}^{\infty} \mathbb{E}[|X|^k] \frac{t^k}{k!}$$

et le terme de gauche est supposé fini, d'où $\mathbb{E}[|X|^k] < \infty$ pour tout k , et la justification de l'utilisation du théorème de convergence dominée ici :

$$\mathcal{L}_X(t) = \mathbb{E}[e^{tX}] = \mathbb{E}\left[\sum_{k=0}^{\infty} X^k \frac{t^k}{k!}\right] = \mathbb{E}\left[\lim_n \sum_{k=0}^n X^k \frac{t^k}{k!}\right] = \lim_n \mathbb{E}\left[\sum_{k=0}^n X^k \frac{t^k}{k!}\right] = \lim_n \sum_{k=0}^n \mathbb{E}[X^k] \frac{t^k}{k!} = \sum_{k=0}^{\infty} \mathbb{E}[X^k] \frac{t^k}{k!}$$

(la domination est donnée, par inégalité triangulaire, par la série précédente). ■

En particulier, si 0 est un point intérieur de \mathcal{I}_X , on peut calculer les moments $\mathbb{E}[X^k]$ de X en dérivant \mathcal{L}_X et en évaluant les dérivées en 0. Inversement, si on connaît les moments *et si on sait que 0 est à l'intérieur de \mathcal{I}_X* , on peut retrouver \mathcal{L}_X et donc la loi de X (si $\mathcal{I}_X \neq \{0\}$...) :

Corollaire (Cas particulier du « Problème des moments »)

Soit X, Y des variables aléatoires réelles telles qu'il existe $t > 0$ pour lequel $\mathbb{E}[e^{t|X|}] < \infty$ et $\mathbb{E}[e^{t|Y|}] < \infty$ (par exemple X, Y bornées). On suppose que, pour tout $k \in \mathbb{N}$, $\mathbb{E}[X^k] = \mathbb{E}[Y^k]$. Alors X et Y ont même loi.

Démonstration : Cela résulte directement des deux propositions précédentes. ■

Si en revanche 0 est un point du bord de \mathcal{I}_X , alors \mathcal{L}_X peut ne pas être \mathcal{C}^∞ en 0 :

Proposition

On suppose $\mathcal{I}_X \neq \{0\}$.

a) \mathcal{L}_X est continue sur \mathcal{I}_X .

b) La fonction \mathcal{L}_X est n fois dérivable en 0 si, et seulement si $\mathbb{E}[|X|^n] < \infty$. Et alors

$$\mathcal{L}_X^{(n)}(0) = \mathbb{E}[X^n].$$

Démonstration : Sans perdre de généralité, on peut supposer que $\mathcal{I}_X \cap]0, +\infty[\neq \emptyset$. En effet la preuve s'adapte immédiatement au cas symétrique (ou on peut considérer $-X$).

a) Soit $t \in \mathcal{I}_X$. On a la domination suivante : pour $0 \leq s \leq t$, $e^{sX} \leq \mathbf{1}_{\{X < 0\}} + e^{tX} \mathbf{1}_{\{X \geq 0\}} \leq 1 + e^{tX}$, et $\mathbb{E}[1 + e^{tX}] < \infty$ car $t \in \mathcal{I}_X$. Vu que, pour tout x , $t \mapsto e^{tx}$ est continue, le théorème de continuité sous l'intégrale permet de conclure que \mathcal{L}_X est continue sur $[0, t]$. Ce qui implique a).

b) On montre par récurrence que, si $\mathbb{E}[|X|^n] < \infty$, alors la dérivée n -ième de \mathcal{L}_X sur \mathcal{I}_X existe et vaut

$$\mathcal{L}_X^{(n)}(t) = \mathbb{E}[X^n e^{tX}],$$

d'où en particulier $\mathcal{L}_X^{(n)}(0) = \mathbb{E}[X^n]$, et qu'inversement si $\mathbb{E}[|X|^n] = \infty$ alors \mathcal{L}_X n'est pas n fois dérivable en 0.

C'est évident si $n = 0$. On suppose que c'est vrai pour n .

Supposons $\mathbb{E}[|X|^{n+1}] < \infty$. Soit $t \in \mathcal{I}_X$ avec $t > 0$: $\mathbb{E}[e^{tX}] < \infty$. Soit $\varepsilon > 0$. Il existe $C_{n,t,\varepsilon} > 0$ telle que, pour $0 \leq s \leq t(1 - \varepsilon)$, pour tout $x \geq 0$ on a $x^{n+1} e^{sx} \leq C_{n,t,\varepsilon} e^{tx}$, d'où la domination

$$|X^{n+1} e^{sX}| \leq |X|^{n+1} \mathbf{1}_{\{X \leq 0\}} + C_{n,t,\varepsilon} e^{tX} \mathbf{1}_{\{X > 0\}} \leq |X|^{n+1} + C_{n,t,\varepsilon} e^{tX}$$

qui montre que $\mathcal{L}_X^{(n)}$ est dérivable sur $[0, t(1 - \varepsilon)]$, et sa dérivée $\mathcal{L}_X^{(n+1)}$ est donnée par la formule annoncée (théorème de dérivation sous l'intégrale). Ainsi \mathcal{L}_X est dérivable sur tout $\mathcal{I}_X \cap]0, \infty[$ (en 0 c'est une dérivée à droite). Pour le cas $t < 0$ (si $\mathcal{I}_X \cap]-\infty, 0[\neq \emptyset$), on applique la preuve à $-X$, et les dérivées en 0 à droite et à gauche coïncident, elles valent $\mathbb{E}[X^n]$.

Supposons inversement $\mathbb{E}[|X|^{n+1}] = \infty$. On peut supposer $\mathbb{E}[|X|^n] < \infty$ (sans quoi \mathcal{L}_X n'est même pas n fois dérivable). Pour $t \in \mathcal{I}_X$ avec $t > 0$, il existe $C > 0$ tel que, pour tout $x \geq 0$, $x^{n+1} \leq C e^{tx}$, donc

$$\mathbb{E}[|X|^{n+1} \mathbf{1}_{\{X \geq 0\}}] \leq C \mathbb{E}[e^{tX}] < \infty,$$

de sorte que, plus précisément, l'hypothèse s'écrit $\mathbb{E}[|X|^{n+1} \mathbf{1}_{\{X < 0\}}] = \infty$. Or,

$$\frac{\mathcal{L}_X^{(n)}(t) - \mathcal{L}_X^{(n)}(0)}{t} = \mathbb{E}\left[X^n \frac{e^{tX} - 1}{t}\right] = \mathbb{E}\left[X^n \frac{e^{tX} - 1}{t} \mathbf{1}_{\{X < 0\}}\right] + \mathbb{E}\left[X^n \frac{e^{tX} - 1}{t} \mathbf{1}_{\{X > 0\}}\right],$$

et quand $t \rightarrow 0^+$, par théorème de convergence monotone ($\forall x, t \mapsto \frac{e^{tx} - 1}{t}$ est croissante), les espérances de droites convergent vers $\mathbb{E}[X^{n+1} \mathbf{1}_{\{X < 0\}}]$ et $\mathbb{E}[X^{n+1} \mathbf{1}_{\{X > 0\}}]$. Or la seconde limite est finie, et la première vaut $\pm\infty$ (selon la parité de n), donc \mathcal{L}_X n'est pas $n + 1$ fois dérivable en 0. ■

En particulier, par la formule de Taylor-Young, si $\mathbb{E}[|X|^n] < \infty$ alors, quand $t \rightarrow 0$ dans \mathcal{I}_X ,

$$\mathcal{L}_X(t) = 1 + t\mathbb{E}[X] + t^2 \frac{\mathbb{E}[X^2]}{2} + \dots + t^n \frac{\mathbb{E}[X^n]}{n!} + o(t^n).$$

Une propriété essentielle de la transformée de Laplace est la suivante :

Proposition

Si X et Y sont des variables aléatoires indépendantes, alors, pour tout $t \in \mathbb{R}$,

$$\mathcal{L}_{X+Y}(t) = \mathcal{L}_X(t)\mathcal{L}_Y(t).$$

En particulier, $\mathcal{I}_{X+Y} = \mathcal{I}_X \cap \mathcal{I}_Y$.

Démonstration : Il suffit d'écrire, grâce à l'indépendance de X et Y ,

$$\mathcal{L}_{X+Y}(t) = \mathbb{E}[e^{t(X+Y)}] = \mathbb{E}[e^{tX} e^{tY}] = \mathbb{E}[e^{tX}] \mathbb{E}[e^{tY}] = \mathcal{L}_X(t)\mathcal{L}_Y(t). \quad \blacksquare$$

Utilisations classiques

Loi de la somme de variables indépendantes. Par la proposition précédente, on peut déduire \mathcal{L}_{X+Y} de \mathcal{L}_X et \mathcal{L}_Y , et identifier la transformée de Laplace d'une loi connue, ou en déduire diverses propriétés de $X + Y$. Exemples : loi normale, loi gamma, loi de Poisson,...

Inégalités de concentration. L'inégalité de Markov donne, pour tout $\lambda > 0$ et $t \in \mathbb{R}$,

$$\mathbb{P}(X > t) = \mathbb{P}(e^{\lambda X} > e^{\lambda t}) \leq \mathbb{E}[e^{\lambda X}] e^{-\lambda t} = \exp(-\lambda t + \ln \mathcal{L}_X(\lambda)).$$

C'est l'**inégalité de Chernoff**. On note qu'elle dépend d'un paramètre λ , que l'on peut choisir pour minimiser le terme de droite. On l'utilise en général dans le cas où X est une somme de variables indépendantes : cf. inégalité d'Hoeffding. C'est un outil important de la théorie des grandes déviations.

Calcul de fonctions caractéristiques. Dans certains cas, la transformée de Laplace est plus simple à calculer que la fonction caractéristique (définie ci-après). On peut néanmoins en déduire la fonction caractéristique directement par prolongement analytique et continuité (voir preuve de l'injectivité plus haut). Formellement, cela se ramène à remplacer « t » par « $-it$ ». Exemples : loi normale, loi gamma.

4 Fonction caractéristique (v.a. à valeurs dans \mathbb{R}^d)

Définition

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

$$\left\{ \begin{array}{l} \Phi_X : \mathbb{R} \rightarrow \mathbb{C} \\ t \mapsto \Phi_X(t) = \mathbb{E}[e^{itX}], \end{array} \right.$$

Pour X à valeurs dans \mathbb{R}^d , on définit aussi la fonction caractéristique Φ_X sur \mathbb{R}^d par : pour $t \in \mathbb{R}^d$,

$$\Phi_X(t) = \mathbb{E}[e^{i\langle t, X \rangle}],$$

où $\langle \cdot, \cdot \rangle$ est le produit scalaire usuel.

Il est important de noter que l'espérance $\mathbb{E}[e^{itX}]$ est bien définie, quel que soit t : on a $\mathbb{E}[|e^{itX}|] = \mathbb{E}[1] = 1 < \infty$. D'ailleurs, par l'inégalité triangulaire ceci montre aussi que $|\Phi_X(t)| \leq 1$.

Si X a pour densité f_X sur \mathbb{R}^d , on a, pour tout $t \in \mathbb{R}^d$,

$$\Phi_X(t) = \int e^{i\langle t, x \rangle} f_X(x) dx,$$

donc Φ_X est la transformée de Fourier de f_X (à changement de variable $t \mapsto -t$ près).

Proposition

a) La fonction Φ_X est continue sur \mathbb{R}^d , et $\Phi_X(0) = 1$.

b) On suppose X à valeurs réelles.

— Pour tout $n \in \mathbb{N}$, si $\mathbb{E}[|X|^n] < \infty$, alors Φ_X est n fois dérivable sur \mathbb{R} , et dans ce cas

$$\Phi_X^{(n)}(0) = i^n \mathbb{E}[X^n].$$

— Inversement, pour tout $n \in \mathbb{N}$ pair, si $\mathbb{E}[X^n] = \infty$, alors Φ_X n'est pas n fois dérivable en 0.

Remarques. • Par contre, on peut avoir $\mathbb{E}[|X|] = \infty$ et Φ_X dérivable : l'équivalence n'a lieu que pour n pair. • À la différence de la transformée de Fourier d'une fonction L^1 , Φ_X ne tend pas nécessairement vers 0 l'infini : par exemple, si $X = 0$ p.s., alors $\Phi_X(t) = 1$ pour tout t .

Démonstration : La continuité et la dérivabilité se déduisent facilement des théorèmes de continuité et dérivation sous l'intégrale, grâce aux dominations (qui sont même des égalités) du type : pour tout $t \in \mathbb{R}$, $|(iX)^n e^{itX}| = |X|^n$. On obtient donc, si $\mathbb{E}[|X|^n] < \infty$: pour tout $t \in \mathbb{R}$,

$$\Phi_X^{(n)}(t) = i^n \mathbb{E}[X^n e^{itX}].$$

Dernier point) Soit n pair, ≥ 2 , et tel que $\mathbb{E}[|X|^{n-2}] < \infty$ et $\mathbb{E}[|X|^n] = \infty$. Si Φ_X était n fois dérivable, alors en particulier on aurait

$$\frac{\Phi_X^{(n-2)}(t) - 2\Phi_X^{(n-2)}(0) + \Phi_X^{(n-2)}(-t)}{t^2} \xrightarrow[t \rightarrow 0]{} \Phi_X^{(n)}(0).$$

Or, on a une expression de la partie réelle

$$\Re \left(\frac{\Phi_X^{(n-2)}(t) - 2\Phi_X^{(n-2)}(0) + \Phi_X^{(n-2)}(-t)}{i^{n-2} t^2} \right) = -2\mathbb{E} \left[X^{n-2} \frac{1 - \cos(tX)}{t^2} \right]$$

et le lemme de Fatou (on a n pair donc $X^{n-2}(1 - \cos(tX)) \geq 0$) donne

$$2\mathbb{E} \left[\liminf_{t \rightarrow 0} X^{n-2} \frac{1 - \cos(tX)}{t^2} \right] \leq \liminf_{t \rightarrow 0} 2\mathbb{E} \left[X^{n-2} \frac{1 - \cos(tX)}{t^2} \right],$$

c'est-à-dire

$$+\infty = \mathbb{E}[X^n] \leq \Re \left(-\frac{1}{i^{n-2}} \Phi_X^{(n)}(0) \right) < +\infty,$$

ce qui est contradictoire, et montre donc que Φ_X n'est pas n fois dérivable en 0. ■

En particulier, par la formule de Taylor-Young, si $\mathbb{E}[|X|^n] < \infty$ alors, quand $t \rightarrow 0$,

$$\Phi_X(t) = 1 + it\mathbb{E}[X] - t^2 \frac{\mathbb{E}[X^2]}{2} + \dots + i^n t^n \frac{\mathbb{E}[X^n]}{n!} + o(t^n).$$

Comme son nom l'indique, la fonction caractéristique caractérise la loi :

Proposition

- a) Soit X et Y deux variables aléatoires à valeurs dans \mathbb{R}^d . On suppose que, pour tout $t \in \mathbb{R}^d$, $\Phi_X(t) = \Phi_Y(t)$. Alors X et Y ont même loi.
- b) De plus, si la fonction Φ_X est intégrable sur \mathbb{R} , alors la loi de X admet une densité f_X continue bornée, donnée par la **formule d'inversion de Fourier** : pour $x \in \mathbb{R}$,

$$f_X(x) = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} e^{-i\langle t, x \rangle} \Phi_X(t) dt.$$

Démonstration : a) Voir exercice 17 pour une preuve utilisant le théorème de Weierstrass (densité des polynômes trigonométriques dans les fonctions continues sur un segment).

b) Voir une preuve de la formule d'inversion de Fourier : dans Rudin, Queffelec-Zuily (à adapter),... ou un livre de probabilités. ■

Une propriété essentielle de la fonction caractéristique est la suivante :

Proposition

Si X et Y sont des variables aléatoires indépendantes, alors, pour tout $t \in \mathbb{R}^d$,

$$\Phi_{X+Y}(t) = \Phi_X(t)\Phi_Y(t).$$

Démonstration : Il suffit d'écrire, grâce à l'indépendance de X et Y ,

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

Signalons une autre propriété, à ne pas confondre avec la précédente. Celle-ci concerne la loi du couple.

Proposition

Soit X et Y sont des variables aléatoires réelles. X et Y sont indépendantes si, et seulement si, pour tout $(s, t) \in \mathbb{R}^2$,

$$\Phi_{(X,Y)}(s, t) = \Phi_X(s)\Phi_Y(t).$$

Démonstration : On suppose X, Y indépendantes. Il suffit alors d'écrire

$$\Phi_{(X,Y)}(s, t) = \mathbb{E}[e^{i(sX+tY)}] = \mathbb{E}[e^{isX} e^{itY}] = \mathbb{E}[e^{isX}] \mathbb{E}[e^{itY}] = \Phi_X(s)\Phi_Y(t).$$

Inversement, supposons que, pour tous $s, t \in \mathbb{R}$, $\Phi_{(X,Y)}(s, t) = \Phi_X(s)\Phi_Y(t)$. Par la première implication, cette égalité exprime le fait que la fonction caractéristique de (X, Y) est la même que celle d'un couple de deux variables aléatoires indépendantes dont les lois sont les mêmes que celles de X et de Y . Comme la fonction caractéristique caractérise la loi (sur \mathbb{R}^2 , du couple (X, Y)), et que l'indépendance est une propriété de la loi du couple, on en déduit que X et Y sont indépendantes. ■

Utilisations classiques

Loi de la somme de variables indépendantes. Par une proposition précédente, on peut déduire Φ_{X+Y} de Φ_X et Φ_Y , et identifier la fonction caractéristique d'une loi connue, ou en déduire diverses propriétés de $X + Y$. Exemples : loi normale, loi gamma, loi de Cauchy. Inversement, on peut aussi chercher à « diviser » une variable aléatoire X en somme de variables aléatoires indépendantes (et de même loi), et écrivant Φ_X comme produit de fonctions caractéristiques ; c'est à la fois un outil pratique (décomposition d'une loi donnée) et théorique (détermination de toutes les lois vérifiant tel ou tel type de décomposition).

Loi d'un couple de variables aléatoires. Dans certains cas, on peut calculer $\Phi_{(X,Y)}$ et, grâce à la proposition précédente, conclure à l'indépendance (ou non) de X et Y . Exemple : pour X, Y indépendantes et de loi $\mathcal{N}(0,1)$, montrer que $U = X + Y$ et $V = X - Y$ sont indépendantes.

Convergence en loi. On verra que la convergence en loi équivaut à la convergence des fonctions caractéristiques (théorème de Lévy). Cela en fait un outil particulièrement utile. Voir par exemple la preuve du théorème central limite.

III. 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.

Considérons une expérience qui consiste en des tirages successifs d'une pièce de monnaie équilibrée. On observe en pratique que, si on effectue un grand nombre de tirages, la proportion de fois où on obtient « pile » s'approche de 0,5. La loi des grands nombres justifie cette observation dans notre modèle mathématique.

La proportion de « piles » est rarement exactement égale à 0,5. Si on effectue 1000 tirages, on peut obtenir des proportions telles que 0,521 ou 0,485, voire 0,001 dans des cas exceptionnels... Comment se comportent ces déviations par rapport à la valeur limite 0,5 ? Représentées sur un histogramme (où on note par exemple combien de fois on obtient une proportion entre 0 et 0,01, puis entre 0,01 et 0,02, etc. quand on répète les 1000 tirages un certain nombre de fois), on observe qu'elles se distribuent selon une « courbe en cloche » centrée en 0,5. Le théorème central limite précise et justifie cette observation dans notre modèle.

Ces deux questions 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 deux théorèmes qui sont le but principal de ce chapitre.

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

1 Suites d'événements : deux outils

1.1 Intersection dénombrable d'événements presque sûrs

On commence par rappeler une propriété simple qui servira à plusieurs reprises :

Lemme

« Une intersection dénombrable d'événements presque sûrs est presque sûre » : Si $(A_i)_{i \in I}$ est une famille d'événements telle que $\mathbb{P}(A_i) = 1$ pour tout $i \in I$, et si I est dénombrable, alors $\mathbb{P}\left(\bigcap_{i \in I} A_i\right) = 1$.

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

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

Ce lemme peut se voir comme un échange de quantificateurs : à condition que I soit dénombrable,

si : pour tout $i \in I$, p.s. A_i est réalisé, alors : p.s., pour tout $i \in I$, A_i est réalisé.

1.2 Lemme de Borel-Cantelli

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

Définition

On définit deux nouveaux événements : la **limite supérieure** de la suite $(A_n)_{n \geq 0}$ est

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

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

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

NB. On aurait pu souligner la décroissance ou croissance de certaines suites d'événements :

$$\limsup_n A_n = \bigcap_{n \geq 0} \bigcup_{k \geq n} A_k \quad \text{et} \quad \liminf_n A_n = \bigcup_{n \geq 0} \bigcap_{k \geq n} A_k.$$

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

Le résultat suivant est d'usage très fréquent pour obtenir des propriétés presque sûres concernant des suites de variables aléatoires.

Théorème (Lemme de Borel-Cantelli)

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

$$\mathbb{P}(\limsup_n A_n) = 0.$$

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

$$\mathbb{P}(\limsup_n A_n) = 1.$$

Démonstration : a) On note N le nombre (aléatoire) d'événements de la suite $(A_n)_{n \geq 0}$ qui se réalisent. On a bien sûr

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

C'est une série de fonctions mesurables positives donc, par le TCM,

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

Ainsi, si $\mathbb{P}(A_n) < \infty$, alors $\mathbb{E}[N] < \infty$, donc $N < \infty$ p.s., ce que l'on voulait démontrer, vu la définition de $\limsup_n A_n$.

b) On suppose $\sum_n \mathbb{P}(A_n) = \infty$, et que la famille $(A_n)_n$ est indépendante. On a

$$\mathbb{P}(\limsup_n A_n) = \mathbb{P}\left(\bigcap_{n \geq 0} \bigcup_{k \geq n} A_k\right) = \liminf_n \mathbb{P}\left(\bigcup_{k \geq n} A_k\right) = 1 - \limsup_n \mathbb{P}\left(\bigcap_{k \geq n} (A_k)^c\right).$$

On souhaite donc montrer que, pour tout n , la dernière probabilité vaut 0. Si les événements $(A_k)_k$ sont indépendants, alors leurs complémentaires aussi. On a donc, pour tout $N \geq n$,

$$\mathbb{P}\left(\bigcap_{n \leq k \leq N} (A_k)^c\right) = \prod_{n \leq k \leq N} \mathbb{P}((A_k)^c) = \prod_{n \leq k \leq N} (1 - \mathbb{P}(A_k)) \leq \prod_{n \leq k \leq N} e^{-\mathbb{P}(A_k)} = \exp\left(-\sum_{n \leq k \leq N} \mathbb{P}(A_k)\right),$$

en utilisant l'inégalité $1 - x \leq e^{-x}$, valable pour tout $x \in \mathbb{R}$. Quand $N \rightarrow \infty$, le terme de droite tend vers 0 vu l'hypothèse, et le terme de gauche converge vers $\mathbb{P}(\bigcap_{k \geq n} A_k)$ car $\bigcap_{k \geq n} A_k$ est l'intersection de la suite décroissante d'événements $B_N = \bigcap_{n \leq k \leq N} A_k$. On obtient ainsi le résultat annoncé. ■

Applications classiques

Singes dactylographes. Soit $N \geq 2$. Supposons que $(X_n)_{n \geq 1}$ est une suite de variables aléatoires indépendantes, chacune de loi uniforme dans $\{1, 2, \dots, N\}$. Soit une suite finie (k_1, \dots, k_q) d'éléments de $\{1, 2, \dots, N\}$. La suite d'événements $(A_n)_{n \geq 0}$ définie par

$$A_n = \{X_{qn+1} = k_1, X_{qn+2} = k_2, \dots, X_{qn+q} = k_q\}$$

sont indépendants par la propriété d'indépendance par paquets (A_0 dépend de X_1, \dots, X_q puis A_1 de X_{q+1}, \dots, X_{2q} , etc.), et ont tous pour probabilité $\frac{1}{N^q}$, d'où $\sum_n \mathbb{P}(A_n) = \infty$. Par le point b) du lemme de Borel-Cantelli, presque sûrement, l'événement A_n se produit pour une infinité de n . En particulier, p.s., la séquence (k_1, \dots, k_q) apparaît infiniment souvent dans la suite $(X_n)_{n \geq 1}$. Ceci est vrai pour toute séquence finie (k_1, \dots, k_q) . Or l'ensemble de ces séquences est dénombrable (union dénombrable d'ensembles finis), donc par le premier lemme on a finalement démontré :

Presque sûrement, toute séquence finie apparaît infiniment souvent dans la suite $(X_n)_{n \geq 1}$.

Plus grande séquence de piles consécutifs. Soit $(X_n)_{n \geq 1}$ une suite de variables aléatoires indépendantes, de loi $\mathcal{B}(1/2)$. On les interprète comme des tirages à pile (1) ou face (0). On considère l'événement

$$A_n = \{\text{parmi les } n \text{ premiers tirages, il y a eu au moins } \lfloor 3 \log_2 n \rfloor \text{ piles consécutifs.}\}$$

On note $\ell_n = \lfloor 3 \log_2 n \rfloor$ (où \log_2 est le logarithme en base 2 et $\lfloor \cdot \rfloor$ la partie entière). Remarquons que les événements A_n ne sont pas du tout indépendants. On a (par sous-additivité, puis indépendance à la fin)

$$\mathbb{P}(A_n) = \mathbb{P}\left(\bigcup_{1 \leq k < n - \ell_n} \{X_k = 1, \dots, X_{k+\ell_n-1} = 1\}\right) \leq \sum_{1 \leq k \leq n} \mathbb{P}(X_k = 1, \dots, X_{k+\ell_n-1} = 1) = \frac{n}{2^{\ell_n}}$$

et $n2^{-\ell_n} \leq n2^{-3 \log_2 n + 1} = n2n^{-3} = 2n^{-2}$, d'où $\sum_n \mathbb{P}(A_n) < \infty$. Ainsi, par le point a) du lemme de Borel-Cantelli, presque sûrement, un nombre fini d'événements A_n se réalisent. Autrement dit, on a montré :

Presque sûrement, pour tout n assez grand, il y a au plus $3 \log_2 n$ piles consécutifs parmi les n premiers tirages.

NB. On pourrait remplacer 3 par $2 + \varepsilon$ pour tout $\varepsilon > 0$. On pourrait aussi utiliser le point b), en découpant les n tirages en blocs disjoints de longueur ℓ_n , pour obtenir une minoration par $(1 - \varepsilon) \log_2 n$ quel que soit $\varepsilon > 0$.

Moments et majoration presque sûre. Soit $(X_n)_{n \geq 1}$ une suite de variables aléatoires indépendantes et de même loi, réelles positives. Soit $\alpha > 0$. Soit $\varepsilon > 0$. On s'intéresse aux événements

$$A_n = \{X_n \geq \varepsilon n^\alpha\}.$$

On a

$$\mathbb{P}(A_n) = \mathbb{P}(X_n \geq \varepsilon n^\alpha) = \mathbb{P}\left(\left(\frac{1}{\varepsilon} X_n\right)^{1/\alpha} \geq n\right) = \mathbb{P}\left(\left\lfloor \left(\frac{1}{\varepsilon} X_n\right)^{1/\alpha} \right\rfloor \geq n\right),$$

donc (cf. exercice 3.1), comme $N = \left\lfloor \left(\frac{1}{\varepsilon} X_n\right)^{1/\alpha} \right\rfloor$ est à valeurs dans \mathbb{N} ,

$$\sum_{n \geq 1} \mathbb{P}(A_n) = \sum_{n \geq 1} \mathbb{P}(N \geq n) = \mathbb{E}[N] = \mathbb{E}\left[\left\lfloor \left(\frac{1}{\varepsilon} X_n\right)^{1/\alpha} \right\rfloor\right]$$

ce qui donne l'encadrement (vu que $x \leq \lfloor x \rfloor < x + 1$ pour tout $x \in \mathbb{R}$)

$$\mathbb{E}[(X_n)^{1/\alpha}] \leq \varepsilon^{1/\alpha} \sum_{n \geq 1} \mathbb{P}(A_n) \leq \mathbb{E}[(X_n)^{1/\alpha}] + \varepsilon^{1/\alpha}.$$

Par les points a) et b) du lemme de Borel-Cantelli, on obtient :

$$\mathbb{E}[(X_n)^{1/\alpha}] < \infty \text{ si, et seulement si, presque sûrement, pour tout } n \text{ assez grand, } X_n < \varepsilon n^\alpha.$$

Vu que ceci vaut pour tout $\varepsilon > 0$, en prenant $\varepsilon = \frac{1}{i}$, où $i \in I = \mathbb{N}^*$, le premier lemme de cette partie donne :

$$\mathbb{E}[(X_n)^{1/\alpha}] < \infty \text{ si, et seulement si, presque sûrement, pour tout } i \in \mathbb{N}^*, \text{ pour tout } n \text{ assez grand, } X_n < \varepsilon n^\alpha.$$

Autrement dit,

$$\mathbb{E}[(X_n)^{1/\alpha}] < \infty \text{ si, et seulement si, presque sûrement } \frac{X_n}{n^\alpha} \xrightarrow[n]{n} 0.$$

2 Loi d'une suite de variables aléatoires réelles

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

Définition

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

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

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

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

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

En conséquence de la définition de \mathcal{B} , pour décrire la loi de la suite $(X_n)_{n \geq 0}$, il suffit de se donner la loi de (X_0, \dots, X_n) pour tout $n \geq 0$.

En particulier, si la suite $(X_n)_{n \geq 0}$ est indépendante, alors (par définition) pour tout n , les variables aléatoires X_0, \dots, X_n sont indépendantes donc la loi de (X_0, \dots, X_n) est la loi produit $\mathbb{P}_{X_0} \otimes \dots \otimes \mathbb{P}_{X_n}$. La loi de $(X_n)_{n \geq 0}$ ne dépend alors que des lois de chacun des X_n ; 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 évoqué en introduction : celui d'une suite de tirages à pile ou face équilibrés. Ce cas permet ensuite d'obtenir un cas plus général (voir plus bas).

Proposition

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

$$U = \overline{0, X_1 X_2 \dots} = \sum_{n=1}^{\infty} \frac{X_n}{2^n} \quad (*)$$

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

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

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

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

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 $\varepsilon_N \in \{0,1\}$,

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

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 $\varepsilon_1, \dots, \varepsilon_N \in \{0,1\}$,

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

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. ■

NB. Si on dispose d'une variable aléatoire U de loi uniforme sur $[0,1]$, alors avec une bijection $\varphi : (\mathbb{N}^*)^2 \rightarrow \mathbb{N}^*$, on peut définir, avec les notations ci-dessus, pour tous $k, n \in \mathbb{N}^*$, $X_n^{(k)} = X_{\varphi(n,k)}$, puis pour tout k ,

$$U_k = \sum_{n=1}^{\infty} \frac{X_n^{(k)}}{2^n}.$$

Par la proposition (2 fois) et l'indépendance par paquets, on obtient que $(U_k)_{k \geq 1}$ est une suite de variables aléatoires indépendantes, de loi uniforme sur $[0,1]$.

Ayant construit cette suite $(U_k)_k$, si on se donne maintenant une suite de lois $(\mu_k)_{k \geq 1}$ sur \mathbb{R} , ayant pour fonctions de répartitions respectives $(F_k)_{k \geq 1}$, alors par l'exercice 14.2 la suite $(Y_k)_{k \geq 1} = (F_k^{-1}(U_k))_{k \geq 1}$ est une suite de variables aléatoires indépendantes, telle que Y_k suit la loi μ_k .

Comme on sait définir U (cela revient à l'existence de la mesure de Lebesgue sur \mathbb{R}), on sait ainsi définir, donc montrer l'existence, d'une suite de variables aléatoires réelles indépendantes suivant des lois prescrites.

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).

Mentionnons enfin le résultat suivant, figurant au programme :

Proposition (Loi du 0-1 de Kolmogorov)

Soit $(X_n)_n$ une suite de variables aléatoires indépendantes. On définit, pour tout n , la tribu $\mathcal{F}^n = \sigma(X_n, X_{n+1}, \dots)$, et la tribu asymptotique $\mathcal{F}^\infty = \bigcap_n \mathcal{F}^n$. Pour tout $A \in \mathcal{F}^\infty$, on a $\mathbb{P}(A) \in \{0,1\}$.

Pour des variables réelles $(X_n)_n$, voici quelques exemples d'événements appartenant à \mathcal{F}^∞ , c'est-à-dire ne dépendant de la suite $(X_n)_n$ qu'à partir d'un rang arbitrairement grand : $\{\text{la suite } (X_n)_n \text{ est bornée}\}$, $\limsup_n \{X_n = X_{n+1}\}$, $\{X_n \xrightarrow[n]{\rightarrow} 0\}$.

3 Convergence d'une suite de variables aléatoires

Soit $(X_n)_{n \geq 0}$ une suite de variables aléatoires à valeurs dans \mathbb{R}^d (ou dans un espace métrique quelconque), et X une variable aléatoire à valeurs dans le même espace.

3.1 Convergence en probabilité

Définition

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

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

Propriété

La limite est unique (à égalité presque sûre près) : si $X_n \xrightarrow[n]{(p)} X$ et $X_n \xrightarrow[n]{(p)} X'$, alors $X = X'$ p.s..

Démonstration : On suppose $X_n \xrightarrow[n]{(p)} X$ et $X_n \xrightarrow[n]{(p)} X'$. Soit $N \in \mathbb{N}^*$. On a $|X - X'| \leq |X - X_n| + |X' - X_n|$ donc, si $|X - X'| > \frac{2}{N}$, alors $|X - X_n| + |X' - X_n| > \frac{2}{N}$ et l'un des termes (au moins) doit alors être plus grand que $\frac{1}{N}$, d'où

$$P(|X - X'| > \frac{2}{N}) \leq P(\{|X - X_n| > \frac{1}{N}\} \cup \{|X' - X_n| > \frac{1}{N}\}) \leq P(|X - X_n| > \frac{1}{N}) + P(|X' - X_n| > \frac{1}{N}).$$

Par les hypothèses, le membre de droite tend vers 0 quand $n \rightarrow \infty$. On a donc : $\forall N \in \mathbb{N}^*$, p.s. $|X - X'| \leq \frac{2}{N}$. Comme \mathbb{N}^* est dénombrable, le premier lemme de ce chapitre donne : p.s., $\forall N \in \mathbb{N}^*$, $|X - X'| \leq \frac{2}{N}$, c'est-à-dire p.s., $X = X'$. ■

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

Proposition (Loi faible des grands nombres L^2)

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

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

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

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

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

et

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

car X_1, \dots, X_n sont indépendantes (avant-dernière égalité), et de même loi (dernière égalité). Soit $\delta > 0$. L'inégalité de Tcheychev donne donc ici

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

C'est le résultat annoncé. ■

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

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

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

3.2 Convergence presque sûre

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

Définition

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

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

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

Proposition

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

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

$$\text{pour tout } \delta > 0, \quad \sum_{n \geq 0} \mathbb{P}(|X_n - X| > \delta) < \infty$$

alors $X_n \xrightarrow[n]{p.s.} X$.

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

Démonstration : a) On suppose $X_n \xrightarrow[n]{p.s.} X$. Soit $\delta > 0$. Vu la convergence, on a : p.s., pour tout n grand, $|X_n - X| \leq \delta$. Autrement dit, p.s., il y a un nombre fini de n pour lesquels $|X_n - X| > \delta$, ou encore :

$$\mathbb{P}\left(\limsup_n \{|X_n - X| > \delta\}\right) = 0.$$

Or

$$\mathbb{P}\left(\limsup_n \{|X_n - X| > \delta\}\right) = \mathbb{P}\left(\bigcap_n \bigcup_{k \geq n} \{|X_k - X| > \delta\}\right) = \lim_n \mathbb{P}\left(\bigcup_{k \geq n} \{|X_k - X| > \delta\}\right)$$

D'où

$$\mathbb{P}(|X_n - X| > \delta) \leq \mathbb{P}\left(\bigcup_{k \geq n} \{|X_k - X| > \delta\}\right) \searrow 0.$$

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

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

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

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

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

Alors, pour tout $\delta > 0$,

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

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

Théorème (Loi forte des grands nombres)

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

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

Démonstration : Dans le cas où les variables sont dans L^4 (c'est-à-dire $\mathbb{E}[(X_1)^4] < \infty$), qui couvre en particulier le cas borné (tirages à pile-ou-face, sondages...), on peut donner une preuve très courte. Notons $S_n = X_1 + \dots + X_n$. Quitte à remplacer X_n par $X_n - \mathbb{E}[X_n]$, on peut supposer $\mathbb{E}[X_n] = 0$. En développant et en regroupant les termes, en notant $X = X_1$,

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

d'où (par le TCM),

$$\mathbb{E}\left[\sum_{n \geq 1} \left(\frac{S_n}{n}\right)^4\right] = \sum_{n \geq 1} \frac{\mathbb{E}[(S_n)^4]}{n^4} < \infty.$$

et par conséquent $\sum_n \left(\frac{S_n}{n}\right)^4 < \infty$ p.s., donc le terme général $\frac{(S_n)^4}{n^4}$ tend vers 0 p.s., et en particulier $\frac{S_n}{n} \rightarrow 0$ p.s., ce que l'on voulait démontrer. ■

NB. Le théorème s'étend immédiatement au cas de v.a. à valeurs dans \mathbb{R}^d , en l'appliquant à chaque composante. Notons que ce résultat (difficile dans le cas général L^1) implique la loi faible des grands nombres donnée plus haut. De plus, dans le cas de tirages à pile ou face, il traduit tout à fait l'observation : pour (presque) toute suite de tirages, la proportion de piles converge vers 1/2. De même, si la probabilité de pile est p (en prenant X_i de loi $\mathcal{B}(p)$), la proportion de piles converge vers p . Ceci justifie l'interprétation de la probabilité $\mathbb{P}(A)$ *a posteriori* comme fréquence d'occurrence de l'événement A si on répète l'expérience *de façon indépendante* un grand nombre de fois.

3.3 Convergence L^p

En tant que fonctions mesurables à valeurs dans \mathbb{R}^d , on peut définir pour les variables aléatoires les normes L^p , où $1 \leq p < \infty$, et donc les espaces $L^p = L^p(\Omega, \mathcal{A}, \mathbb{P})$ et les convergences associées (en probabilités, on utilise peu la convergence L^∞) :

Définition

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

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

La proposition suivante donne quelques comparaisons entre modes de convergence :

Proposition

- a) Si $X_n \xrightarrow[n]{L^p} X$, alors $X_n \xrightarrow[n]{(p)} X$.
- b) Si $X_n \xrightarrow[n]{\rightarrow} X$ p.s., et si $|X_n| \leq Z$ pour tout n , où $Z \in L^p$, alors $X_n \xrightarrow[n]{L^p} X$.
- c) Si $1 \leq p < q$, $X_n \xrightarrow[n]{L^q} X$, implique $X_n \xrightarrow[n]{L^p} X$.

Démonstration : a) résulte de l'inégalité de Markov. b) résulte du théorème de convergence dominée. c) se déduit par exemple de l'inégalité de Jensen : $\varphi : x \mapsto x^{q/p}$ est convexe, donc

$$\mathbb{E}[|X_n - X|^{q/p}]^{p/q} = \varphi(\mathbb{E}[|X_n - X|]) \leq \mathbb{E}[\varphi(|X_n - X|)] = \mathbb{E}[|X_n - X|^q],$$

d'où $\|X_n - X\|_p \leq \|X_n - X\|_q$. ■

En revanche, en général, la convergence p.s. n'implique pas la convergence dans L^p .

NB. On peut démontrer que, dans la loi des grands nombres, la convergence a lieu dans L^1 , et même dans L^2 si on ajoute l'hypothèse que les variables aléatoires sont dans L^2 . Cette dernière assertion est élémentaire : si $\mathbb{E}[(X_1)^2] < \infty$, alors en notant $\bar{X}_n = \frac{1}{n}(X_1 + \dots + X_n)$, on a, vu que $\mathbb{E}[\bar{X}_n] = \mathbb{E}[X_1]$,

$$\mathbb{E}\left[|\bar{X}_n - \mathbb{E}[X_1]|^2\right] = \mathbb{E}\left[\left(\bar{X}_n - \mathbb{E}[\bar{X}_n]\right)^2\right] = \text{Var}(\bar{X}_n) = \frac{1}{n} \text{Var}(X_1) \xrightarrow[n]{\rightarrow} 0$$

(voir la preuve de la loi faible des grands nombres L^2), ce qui montre que $\bar{X}_n \xrightarrow[n]{L^2} \mathbb{E}[X_1]$. On retrouve d'ailleurs la loi faible des grands nombres L^2 , vu que la convergence L^2 implique la convergence en probabilité.

4 Convergence en loi

Soit $(X_n)_{n \geq 0}$ une suite de variables aléatoires, à valeurs dans \mathbb{R}^d (ou dans un espace topologique quelconque), et X une variable aléatoire à valeurs dans le même espace.

4.1 Définition et propriétés

Définition

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

Une remarque essentielle est que la variable aléatoire X pourrait être remplacée par n'importe quelle variable aléatoire ayant la même loi. Pour cette raison, on peut aussi dire que X_n **converge en loi vers la loi μ** , noté $X_n \xrightarrow[n]{\text{loi}} \mu$ si $X_n \xrightarrow[n]{\text{loi}} X$ où X suit la loi μ .

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

On dispose de diverses manières équivalentes et souvent plus pratiques de démontrer une convergence en loi :

Théorème

Les assertions suivantes sont équivalentes :

- (i) $(X_n)_n$ converge vers X en loi.
- (ii) pour toute fonction $\varphi : \mathbb{R}^d \rightarrow \mathbb{R}$ continue bornée, $\mathbb{E}[\varphi(X_n)] \xrightarrow[n]{} \mathbb{E}[\varphi(X)]$.
- (iii) pour toute fonction $\varphi : \mathbb{R}^d \rightarrow \mathbb{R}$ de classe \mathcal{C}^∞ à support compact, $\mathbb{E}[\varphi(X_n)] \xrightarrow[n]{} \mathbb{E}[\varphi(X)]$.
- (iv) **si $d = 1$** : en tout $t \in \mathbb{R}$ où la fonction de répartition F_X est continue (c'est-à-dire que $\mathbb{P}(X = t) = 0$),

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

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

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

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

Proposition

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

Démonstration : a) On suppose que $(X_n)_n$ converge vers X en probabilité. Soit $\varphi : \mathbb{R}^d \rightarrow \mathbb{R}$ de classe \mathcal{C}^∞ à support compact. Soit $\varepsilon > 0$. Comme φ est continue sur un compact, elle est uniformément continue (thm de Heine) donc il existe δ tel que $|x - y| < \delta$ implique $|\varphi(x) - \varphi(y)| < \varepsilon$. Alors, pour tout n ,

$$\begin{aligned} |\mathbb{E}[\varphi(X_n)] - \mathbb{E}[\varphi(X)]| &\leq \mathbb{E}[|\varphi(X_n) - \varphi(X)|] = \mathbb{E}[|\varphi(X_n) - \varphi(X)| \mathbf{1}_{\{|X_n - X| < \delta\}}] + \mathbb{E}[|\varphi(X_n) - \varphi(X)| \mathbf{1}_{\{|X_n - X| \geq \delta\}}] \\ &\leq \varepsilon + 2\|\varphi\|_\infty \mathbb{P}(|X_n - X| \geq \delta) \end{aligned}$$

(en majorant $|\varphi(X_n) - \varphi(X)|$ par ε dans la première espérance et par $2\|\varphi\|_\infty$ dans la seconde). Par la convergence en probabilité, on a $\mathbb{P}(|X_n - X| \geq \delta) \leq \varepsilon$ pour n assez grand, d'où la conclusion.

b) On suppose $X_n \xrightarrow[n]{\text{loi}} c$. Soit $\delta > 0$. On a $\mathbb{P}(|X_n - c| \geq \delta) = \mathbb{E}[\mathbf{1}_{B(c, \delta)^c}(X_n)]$ et on peut facilement trouver φ continue bornée telle que $\mathbf{1}_{B(c, \delta)^c} \leq \varphi$ et $\varphi(c) = 0$ (par exemple, $\varphi(x) = \min(1, \frac{\|x - c\|}{\delta})$). D'où $\mathbb{P}(|X_n - c| \geq \delta) \leq \mathbb{E}[\varphi(X_n)] \rightarrow \mathbb{E}[\varphi(c)] = 0$ par la convergence en loi, ce qui donne la convergence en probabilité. ■

On peut signaler deux cas particuliers utiles et simples : (le deuxième est connu comme le Lemme de Scheffé)

Proposition

a) Si les v.a. X_n sont discrètes, à valeurs dans un **même** sous-ensemble discret fermé $E \subset \mathbb{R}^d$ (c'est-à-dire sans point d'accumulation), alors $X_n \xrightarrow[n]{\text{(loi)}} X$ si, et seulement si, pour tout $x \in E$,

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

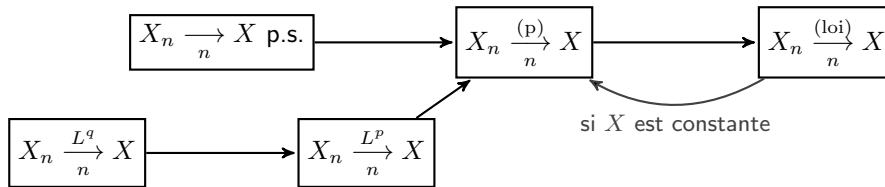
b) Si les v.a. X_n ont chacune une densité f_n , et si $(f_n)_n$ converge p.p. vers une fonction f **qui est une densité**, alors

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

NB. Des variables discrètes peuvent converger vers une loi à densité et vice-versa : X_n de loi uniforme sur $\{\frac{1}{n}, \frac{2}{n}, \dots, \frac{n}{n}\}$ converge vers la loi uniforme sur $[0,1]$, et X_n de loi uniforme sur $[0, \frac{1}{n}]$ converge vers la loi δ_0 .

Exemple (« loi des événements rares »). Si X_n suit la loi $\mathcal{B}(n, p_n)$ où $np_n \xrightarrow[n]{} \lambda > 0$, alors $X_n \xrightarrow[n]{\text{(loi)}} \mathcal{P}(\lambda)$.

Le schéma suivant résume les implications entre les modes de convergence : si $1 \leq p < q$,



4.2 Théorème central limite

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

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

En particulier, la loi $\mathcal{N}(0,1)$ a pour fonction caractéristique $t \mapsto e^{-t^2/2}$. Et la formule précédente reste vraie avec la convention $\mathcal{N}(m,0) = \delta_m$ (une variable aléatoire de variance nulle est p.s. égale à son espérance).

Théorème (Théorème central limite)

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

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

Signalons une autre écriture de cet énoncé :

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

Ceci montre que l'erreur dans la loi des grands nombres est de l'ordre de $\frac{1}{\sqrt{n}}$ et qu'elle est approximativement distribuée selon une loi normale. Dans la suite on suppose $\sigma \neq 0$ (sinon, les v.a. sont constantes).

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.

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

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

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

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

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

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

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

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

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

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

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

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

et par suite

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

(attention, il s'agit d'un logarithme complexe, mais il vérifie bien les propriétés voulues au voisinage de 1 dans \mathbb{C}) autrement dit

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

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

Reprenons notre exemple classique : Soit $(X_n)_{n \geq 1}$ une suite de variables aléatoires indépendantes, de loi $\mathcal{B}(1/2)$. On a $\mathbb{E}[X_1] = 1/2$ et $\operatorname{Var}(X_1) = 1/4$. Par le théorème central limite,

$$2\sqrt{n} \left(\frac{X_1 + \dots + X_n}{n} - \frac{1}{2} \right) \xrightarrow[n]{(\text{loi})} Z$$

où Z suit la loi $\mathcal{N}(0, 1)$.

Ainsi, pour tous $a < b$ réels,

$$\begin{aligned}\mathbb{P} \left(\frac{a}{2\sqrt{n}} < \frac{X_1 + \dots + X_n}{n} - \frac{1}{2} < \frac{b}{2\sqrt{n}} \right) &= \mathbb{P} \left(a < 2\sqrt{n} \left(\frac{X_1 + \dots + X_n}{n} - \frac{1}{2} \right) < b \right) \\ &\xrightarrow[n]{} \mathbb{P}(a < Z < b) = \int_a^b e^{-t^2/2} \frac{dt}{\sqrt{2\pi}}.\end{aligned}$$

Par exemple pour $n = 1000$, la probabilité que la proportion de piles s'écarte de moins de 3% de 50% est (approximativement)

$$\mathbb{P}(-2 \cdot 0,03 \cdot \sqrt{1000} < Z < 2 \cdot 0,03 \cdot \sqrt{1000}) = \int_{-1,9}^{1,9} e^{-t^2/2} \frac{dt}{\sqrt{2\pi}} \simeq 0,94.$$

Autrement dit, dans 94% des cas, après 1000 tirages, la proportion de piles s'écarte de moins de 3% de la moitié.