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,* $P_X = P_Y$*) ;*

(ii) pour tout $B \in \mathcal{B}$, $\quad \mathbb{P}(X \in B) = \mathbb{P}(Y \in B)$ *;*

(iii) pour toute fonction mesurable positive $f : E \to \mathbb{R}$, $\quad \mathbb{E}[f(X)] = \mathbb{E}[f(Y)]$ *;*

(iv) pour toute fonction mesurable bornée $f : E \to \mathbb{R}$, $\quad \mathbb{E}[f(X)] = \mathbb{E}[f(Y)]$.

Démonstration : L'équivalence entre (i) et (ii) relève de la simple définition de 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. Voici un premier exemple de tel raffinement de la proposition précédente :

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,* $P_X = P_Y$*) ;*

(ii') pour tout pavé ouvert $P =]a_1, b_1[\times \cdots \times]a_d, b_d[\in \mathcal{B}$, $\quad \mathbb{P}(X \in P) = \mathbb{P}(Y \in P)$ *;*

(iii') pour toute fonction $f : \mathbb{R}^d \to \mathbb{R}$ *positive, de classe* $\mathcal{C}^\infty$ *et à support compact,* $\quad \mathbb{E}[f(X)] = \mathbb{E}[f(Y)]$.

Démonstration : Par la proposition précédente, (i) implique directement (ii') et (iii').
• Supposons (ii'). Notons d'abord que l'égalité $\mathbb{P}(X \in P) = \mathbb{P}(Y \in P)$ s'étend aux pavés semi-ouverts $P = [a_1,b_1[\times \cdots \times [a_d,b_d[$. En effet, on a

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

d'où, en utilisant (ii'),

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

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 P_X et 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 P_X et 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 $B \in \mathcal{B}(\mathbb{R}^d)$, $\mu(B) = \sup_{\substack{F \text{ fermé} \\ F \subset B}} \mu(F) = \inf_{\substack{G \text{ ouvert} \\ B \subset G}} \mu(G)$.

Or 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 cet ensemble est une tribu contenant les fermés, donc c'est l'ensemble des boréliens.
• 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 $\mathcal{C}^\infty$ 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*

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

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 $\{X < b\} = \{X \leq a\} \uplus \{X \in]a,b[\}$ et

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

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 $\big(t + \frac{1}{n}\big)_n$:

$$\lim_{s \to t^+} F_X(s) = \lim_n{\downarrow} F_X\Big(t + \frac{1}{n}\Big).$$

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

$$\lim_n{\downarrow} F_X\Big(t + \frac{1}{n}\Big) = \lim_n{\downarrow} \mathbb{P}\Big(X \leq t + \frac{1}{n}\Big) = \mathbb{P}(X \leq t) = F_X(t).$$

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

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

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 $t \in \mathbb{R}$,

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

d) La preuve résulte de l'exercice 11 : 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 $u \to -\infty$, $F_X(t) = \int_{-\infty}^{t} f_X$, 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 mais n'est pas dérivable en quelques points $a_1 < \cdots < a_n$, et F'_X (prolongée par une valeur quelconque en $a_1, \ldots, a_n$) 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.
Et inversement, 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 $\varphi : I \to J$ 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 $X_1, \ldots, X_n$ des variables aléatoires réelles indépendantes. On note

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

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

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

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 :

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

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

Méthode de simulation par inversion. Voir exercice 11.

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 $k \geq 0$,

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

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

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

ce qui converge vers $\mathbb{E}[X(X-1)\cdots(X-k+1)]$ quand $s \to 1^-$ 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,

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

■

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*

$$\begin{array}{rccl} \mathcal{L}_X : & \mathbb{R} & \to & [0,+\infty] \\ & t & \mapsto & \mathcal{L}_X(t) = \mathbb{E}[e^{tX}], \end{array}$$

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

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

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 $\lambda \in [0,1]$,

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

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 13. 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. 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),

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

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 :

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

(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 $\mathcal{C}^\infty$ 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). Ceci vaut pour tout $t \in \mathcal{I}_X$ et tout $\varepsilon > 0$, et est donc vrai sur tout $\mathcal{I}_X$.

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 $x \geq 0$, $x^{n+1} \leq Ce^{tx}$, 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,

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

et quand $t \to 0^+$, 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 $t \to 0$ dans $\mathcal{I}_X$,

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

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,

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

■

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*

$$\begin{array}{rccc} \Phi_X : & \mathbb{R} & \to & \mathbb{C} \\ & t & \mapsto & \Phi_X(t) = \mathbb{E}[e^{itX}], \end{array}$$

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*

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

— *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

$$\frac{\Phi_X^{(n-2)}(t) - 2\Phi_X^{(n-2)}(0) + \Phi_X^{(n-2)}(-t)}{t^2} \underset{t\to 0}{\longrightarrow} \Phi_X^{(n)}(0).$$

Or, on a une expression de la partie réelle

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

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

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

c'est-à-dire

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

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 $t \to 0$,

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

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 12 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,

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

■

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 $\mathbb{P}\Big(\bigcap_{i\in I} A_i\Big)=1$.

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

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

■

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*

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

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

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

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

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

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 propriété 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,

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

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

$$\mathbb{P}(\limsup_n A_n) = \mathbb{P}\Big(\bigcap_{n\geq 0}\searrow \bigcup_{k\geq n} A_k\Big) = \lim_n\downarrow \mathbb{P}\Big(\bigcup_{k\geq n} A_k\Big) = 1 - \lim_n\uparrow \mathbb{P}\Big(\bigcap_{k\geq n} (A_k)^c\Big).$$

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 $N \geq n$,

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

en utilisant l'inégalité $1-x \leq e^{-x}$, valable pour tout $x \in \mathbb{R}$. Quand $N \to \infty$, 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 $\{1,2,\ldots,N\}$. Soit une suite finie $(k_1,\ldots,k_q)$ d'éléments de $\{1,2,\ldots,N\}$. La suite d'événements $(A_n)_{n\geq 0}$ définie par

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

sont indépendants par la propriété d'indépendance par paquets (A_0 dépend de $X_1,\ldots,X_q$ puis A_1 de $X_{q+1},\ldots,X_{2q}$, 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 $(k_1,\ldots,k_q)$ apparaît infiniment souvent dans la suite $(X_n)_{n\geq 0}$. Ceci est vrai pour toute séquence finie $(k_1,\ldots,k_q)$. 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 $(X_n)_{n\geq 0}$.

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)

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

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 $\alpha > 0$. On s'intéresse aux événements

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

On a

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

donc (cf. exercice 6.1), comme $N = \lfloor (X_n)^{1/\alpha}\rfloor$ est à valeurs dans $\mathbb{N}$,

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

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

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

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

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

2 Loi d'une suite de variables aléatoires réelles

Si X_1, X_2,... 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 :

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

On peut vérifier alors que $(X_n)_{n\geq 0}$ 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 $X_0,\ldots,X_n$ sont indépendantes donc la loi de $(X_0,\ldots,X_n)$ est la loi produit $P_{X_0}\otimes\cdots\otimes P_{X_n}$. 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 en fait d'obtenir un cas plus général.

Proposition

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

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

(Autrement dit, $X_n=\lfloor 2^nU\rfloor-2\lfloor 2^{n-1}U\rfloor$). 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 $\varepsilon_1,\ldots,\varepsilon_N\in\{0,1\}$, on a

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

où $x=\sum_{n=1}^{N}\varepsilon_n2^{-n}$. 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 $\varepsilon_1,\ldots,\varepsilon_{N-1}$, 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 $X_1,\ldots,X_{N-1}$. On a donc, pour tous $\varepsilon_1,\ldots,\varepsilon_N\in\{0,1\}$,

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

ce qui montre que $X_1,\ldots,X_N$ 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 $\varphi:(\mathbb{N}^*)^2\to\mathbb{N}^*$, 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 11.2 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. Le cas de variables aléatoires dans $\mathbb{R}^d$ est plus délicat.

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).

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 $X_n \xrightarrow[n]{\text{(p)}} X$, si*

$$\text{pour tout } \delta > 0, \qquad \mathbb{P}(|X_n - X| > \delta) \underset{n}{\longrightarrow} 0.$$

Propriété

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

Démonstration : On suppose $X_n \xrightarrow[n]{\text{(p)}} X$ et $X_n \xrightarrow[n]{\text{(p)}} X'$. 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

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

Par les hypothèses, le membre de droite tend vers 0 quand $n \to \infty$. 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

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

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

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

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

et

$$\operatorname{Var}(\overline{X}_n) = \frac{1}{n^2}\operatorname{Var}(X_1 + \cdots + X_n) = \frac{1}{n^2}\big(\operatorname{Var}(X_1) + \cdots + \operatorname{Var}(X_n)\big) = \frac{\operatorname{Var}(X_1)}{n}$$

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

$$\mathbb{P}\Big(\big|\overline{X}_n - \mathbb{E}[X_1]\big| > \delta\Big) = \mathbb{P}\Big(\big|\overline{X}_n - \mathbb{E}[\overline{X}_n]\big| > \delta\Big) \leq \frac{\operatorname{Var}(\overline{X}_n)}{\delta^2} = \frac{\operatorname{Var}(X_1)}{n\delta^2} \underset{n}{\longrightarrow} 0.$$

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

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

Notons que les X_i valent 0 ou 1, donc $X_1 + \cdots + X_n$ est aussi le nombre d'indices entre 1 et n pour lesquels $X_i = 1$, de sorte que $\frac{X_1+\cdots+X_n}{n}$ est la proportion de tirages parmi $X_1, \ldots, X_n$ 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 redite ; on l'écrit ainsi pour la mettre en parallèle avec les autres modes de convergence.

Définition

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

$$\mathbb{P}\big(X_n \underset{n}{\longrightarrow} X\big) = 1.$$

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 $X_n \underset{n}{\longrightarrow} X$ p.s., alors $X_n \overset{(\mathrm{p})}{\underset{n}{\longrightarrow}} X$.
b) $X_n \underset{n}{\longrightarrow} X$ p.s. si, et seulement si, pour tout $\delta > 0$, $\mathbb{P}(\limsup_n\{|X_n - X| > \delta\}) = 0$. En particulier, si

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

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

Démonstration : a) On suppose $X_n \underset{n}{\longrightarrow} X$ p.s.. Par le lemme de Fatou, on a

$$0 = \mathbb{P}(\limsup_n\{|X_n - X| > \delta\}) = \mathbb{E}[\limsup_n \mathbf{1}_{\{|X_n - X|>\delta\}}] \geq \limsup_n \mathbb{E}[\mathbf{1}_{\{|X_n - X|>\delta\}}] = \limsup_n \mathbb{P}(|X_n - X| > \delta)$$

d'où $\mathbb{P}(|X_n - X| > \delta) \to 0$.
b) On a $X_n \not\to X$ 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,

$$1 - \mathbb{P}(X_n \to X) = \mathbb{P}\Big(\bigcup_N \limsup_n\{|X_n - X| > \frac{1}{N}\}\Big) = \lim_N{\uparrow} \mathbb{P}(\limsup_n\{|X_n - X| > \frac{1}{N}\}).$$

L'équivalence b) en résulte rapidement. Le cas particulier est une conséquence des lemmes de la section 1.
c) Si $X_n \to X$ en probabilité, on peut construire la suite $\varphi(k)$ par récurrence de telle sorte que, pour tout k, $\varphi(k+1) > \varphi(k)$ et

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

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

$$\frac{X_1 + \cdots + X_n}{n} \underset{n}{\longrightarrow} \mathbb{E}[X_1] \quad \text{p.s.}$$

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 $S_n = X_1 + \cdots + X_n$. Quitte à remplacer X_n par $X_n - \mathbb{E}[X_n]$, on peut supposer $\mathbb{E}[X_n] = 0$. En développant puis en regroupant les termes, avec $X = X_1$,

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

d'où (par le TCM),

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

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 $\frac{S_n}{n} \to 0$ 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 $\|X_n - X\|_p \underset{n}{\longrightarrow} 0$, c'est-à-dire si*

$$\mathbb{E}\big[|X_n - X|^p\big] \underset{n}{\longrightarrow} 0.$$

La proposition suivante donne quelques comparaisons entre modes de convergence :

Proposition

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

b) Si $X_n \underset{n}{\longrightarrow} X$ 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

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

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 $\overline{X}_n = \frac{1}{n}(X_1 + \cdots + X_n)$, on a, vu que $\mathbb{E}[\overline{X}_n] = \mathbb{E}[X_1]$,

$$\mathbb{E}\Big[\big|\overline{X}_n - \mathbb{E}[X_1]\big|^2\Big] = \mathbb{E}\Big[\Big(\overline{X}_n - \mathbb{E}\big[\overline{X}_n\big]\Big)^2\Big] = \mathrm{Var}\left(\overline{X}_n\right) = \frac{1}{n}\,\mathrm{Var}(X_1) \underset{n}{\longrightarrow} 0$$

(voir la preuve de la loi faible des grands nombres L^2), ce qui montre que $\overline{X}_n \xrightarrow[n]{L^2} \mathbb{E}[X_1]$. 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 $\varphi : \mathbb{R}^d \to \mathbb{R}$, $\mathbb{E}[\varphi(X_n)] \underset{n}{\longrightarrow} \mathbb{E}[\varphi(X)]$.*

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 $\varphi : \mathbb{R}^d \to \mathbb{R}$ *continue bornée,* $\mathbb{E}[\varphi(X_n)] \underset{n}{\longrightarrow} \mathbb{E}[\varphi(X)]$.

(iii) pour toute fonction $\varphi : \mathbb{R}^d \to \mathbb{R}$ *de classe* $\mathcal{C}^\infty$ *à support compact,* $\mathbb{E}[\varphi(X_n)] \underset{n}{\longrightarrow} \mathbb{E}[\varphi(X)]$.

(iv) ***si*** $d = 1$ ***:*** *en tout* $t \in \mathbb{R}$ *où la fonction de répartition* F_X *est continue (c'est-à-dire que* $\mathbb{P}(X = t) = 0$*),*

$$F_{X_n}(t) \underset{n}{\longrightarrow} F_X(t).$$

(v) pour tout $t \in \mathbb{R}^d$, $\Phi_{X_n}(t) \underset{n}{\longrightarrow} \Phi_X(t)$. *(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.

Démonstration : (i) équivaut à (ii) par définition, et (ii) implique (iii) et (v) comme cas particuliers (écrire $\Phi_X(t) = \mathbb{E}[\cos tX] + i\mathbb{E}[\sin tX]$ pour se ramener à des fonctions réelles continues bornées).
(ii) et (iv) : encadrements
(iii) implique (ii), et (v) implique (ii). On commence par montrer que la suite $(X_n)_n$ est tendue : pour tout $\varepsilon > 0$, il existe un compact K tel que, pour tout n, $\mathbb{P}(X_n \notin K) < \varepsilon$, et $\mathbb{P}(X \notin K) < \varepsilon$.
Depuis (iii) : On note ici $|x| = \max_i |x_i|$ la norme L^∞ de $x = (x_1, \ldots, x_d)$. Soit $\varepsilon > 0$. On a $\lim\downarrow_n \mathbb{P}(|X| > n) = 0$ (car les événements $A_n = \{|X| > n\}$ forment une suite décroissante dont l'intersection est $\{|X| = \infty\} = \emptyset$). Il existe donc M tel que $\mathbb{P}(|X| > M) \leq \varepsilon$.
On choisit une fonction ψ de classe $\mathcal{C}^\infty$ à support compact telle que $\psi(x) = 1$ si $|x| < M$, $\psi(x) = 1$ si $|x| > M + 1$ et $0 \leq \psi(x) \leq 1$ pour tout $x \in \mathbb{R}^d$. Comme $|\cdot|$ est la norme L^∞, il suffit de considérer une telle fonction $\psi_{\mathbb{R}} : \mathbb{R} \to \mathbb{R}$ (voir annexe), puis de poser $\psi(x) = \psi_{\mathbb{R}}(x_1) \cdots \psi_{\mathbb{R}}(x_d)$ pour tout $x = (x_1, \ldots, x_d)$.
On a alors, pour tout $x \in \mathbb{R}^d$, $\mathbf{1}_{\{|x|<M\}} \leq \psi(x) \leq \mathbf{1}_{\{|x|<M+1\}}$, d'où en intégrant par rapport à P_X,

$$\mathbb{P}(|X_n| < M + 1) \geq \mathbb{E}[\psi(X_n)] \underset{n}{\longrightarrow} \mathbb{E}[\psi(X)] \geq \mathbb{P}(|X| < M) > 1 - \varepsilon.$$

Ainsi il existe n_0 tel que, pour tout $n \geq n_0$,

$$\mathbb{P}(|X_n| < M + 1) \geq 1 - \varepsilon.$$

Transformer en exercice!!!
En supposant (v) : On voit facilement que la fonction caractéristique de chaque composante converge, et que si chaque composante est tendue alors la suite est tendue, donc on peut ne traiter que le cas réel. Supposons donc X à valeurs réelles. Pour tout $c > 0$, par le théorème de Fubini,

$$\int_{-c}^{c} (1 - \Phi_X(t))dt = \int_{-c}^{c} \mathbb{E}[1 - e^{itX}]dt = \mathbb{E}\Big[\int_{-c}^{c} (1 - e^{itX})dt\Big] = 2c\mathbb{E}\Big[1 - \frac{\sin cX}{cX}\Big] \geq 2c\mathbb{E}\Big[\Big(1 - \frac{\sin cX}{cX}\Big)\mathbf{1}_{\{|cX|\geq 2\}}\Big]$$
$$\geq c\mathbb{P}(|cX| > 2)$$

où la première inégalité repose sur le fait que $\sin x \leq x$ pour tout x, de sorte que le terme dans l'espérance de gauche est positif; et la deuxième vient du fait que, pour $x \geq 2$, $\sin x \leq 1 \leq x/2$. En particulier, pour tout $r > 0$,

$$\mathbb{P}(|X| > r) \leq \frac{r}{2} \int_{-2/r}^{2/r} \big(1 - \Phi_X(t)\big) dt.$$

Ainsi, en appliquant ceci à X_n, on a, pour tout $r > 0$, par théorème de convergence dominée (avec la domination $\Phi_{X_n}(t) \leq 1$)

$$\limsup_n \mathbb{P}(|X_n| > r) \leq \frac{r}{2} \int_{-2/r}^{2/r} \big(1 - \Phi_X(t)\big) dt.$$

Et, quand $r \to +\infty$, le terme de droite converge vers 0 par continuité de Φ_X en 0. On en déduit que la suite $(X_n)_n$ est tendue. ■

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) Dans le cas où X *est constante égale à* $c \in \mathbb{R}^d$, $X_n \xrightarrow[n]{\text{(p)}} c$ *équivaut à* $X_n \xrightarrow[n]{\text{(loi)}} c$.

On peut signaler deux cas particuliers utiles :

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* $x \in E$,

$$\mathbb{P}(X_n = x) \underset{n}{\longrightarrow} \mathbb{P}(X = x).$$

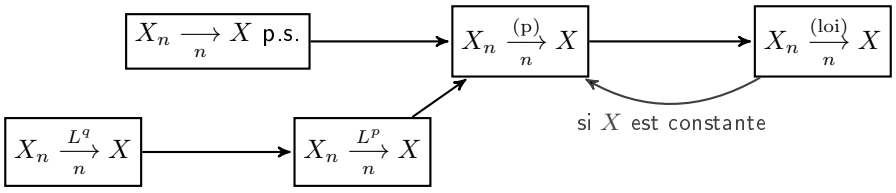
b) Si les v.a. X_n *ont chacune une densité* f_{X_n}, ***et*** *s'il existe une fonction* $q : \mathbb{R} \to \mathbb{R}$ *telle que* $f_{X_n}(x) \leq q(x)$ *pour tout* $x \in \mathbb{R}$ *et* $\int q(x)dx < \infty$, *alors*

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

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

Exemple (« loi des événements rares »). Si X_n suit la loi $\mathcal{B}(n,p_n)$ où $np_n \underset{n}{\longrightarrow} \lambda > 0$, 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* $\sigma^2 = \mathrm{Var}(X_1)$. *On a*

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

Signalons une autre écriture de cet énoncé :

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

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.

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

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

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

(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,

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

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

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

Rappelons-nous maintenant 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 $\frac{t}{\sqrt{n}}\to_n 0$, on a donc, pour tout réel t,

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

et par suite

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

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

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

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

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 $\mathrm{Var}(X_1)=1/4$. Par le théorème central limite,

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

où Z suit la loi $\mathcal{N}(0,1)$.
Ainsi, pour tous $a<b$ réels,

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

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.

Compléments au cours

— Intégrale et dérivée
— Subdivisions dyadiques, imbriquement et monotonie
— Fonctions de classe $\mathcal{C}^\infty$ à support compact

Intégrale et dérivée

On s'intéresse aux généralisations au cadre de l'intégrale de Lebesgue du théorème suivant, dont on donne deux formulations :

Théorème (*« Théorème fondamental du calcul intégral »*)

Intégrale de la dérivée. *Si $F : [a,b] \to \mathbb{C}$ est dérivable sur $[a,b]$ et sa dérivée F' est continue sur $[a,b]$, alors*

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

Dérivée de l'intégrale. *Si $f : [a,b] \to \mathbb{C}$ est continue, alors la fonction*

$$F : x \mapsto F(x) = \int_a^x f(t)dt$$

est dérivable sur $[a,b]$, de dérivée f.

• Dans quelle mesure peut-on affaiblir les hypothèses de la première formulation ?
- — De façon simple (découpage) : si F est continue, dérivable sauf un nombre fini de points, avec une dérivée continue sur chaque morceau (avec limites à gauche et à droite aux points de discontinuité), alors l'énoncé reste vrai.
- — Remarquons qu'en général, si F est dérivable, il se peut que F' ne soit pas intégrable au sens de Riemann (ou même de Lebesgue). Exemple : $F : x \mapsto x^2 \sin \frac{1}{x^2}$ sur [0,1] (F' n'est pas intégrable).
- — Par contre, si on *suppose* F' intégrable sur $[a,b]$, alors l'énoncé reste vrai. *(Théorème 7.21 de Rudin ; voir aussi le théorème 7.20)*
- — Et si on suppose seulement F continue sur $[a,b]$ et dérivable *sauf en un nombre fini de points* de $[a,b]$, avec F' intégrable sur $[a,b]$ (F' est définie presque partout), alors l'énoncé reste vrai aussi.
- — **Attention**, on ne peut cependant pas demander à F d'être seulement dérivable *presque partout*, même si F est continue : l'escalier de Cantor est une fonction continue sur [0,1], croissante, dérivable presque partout et dont la dérivée est nulle presque partout, de sorte que l'intégrale de F' est nulle alors que F n'est pas constante...

• Quant à la seconde formulation, une généralisation (à ne pas confondre avec les précédentes) serait : si f est intégrable sur $[a,b]$, alors F est dérivable presque partout, de dérivée égale presque partout à f. *(Théorème 7.11 de Rudin)*

NB. Les théorèmes du livre de Rudin (Analyse réelle et complexe) se situent au-delà du programme de l'agrégation. On les mentionne ici principalement à fin de souligner les difficultés que présentent les généralisations du lien entre dérivée et intégrale.

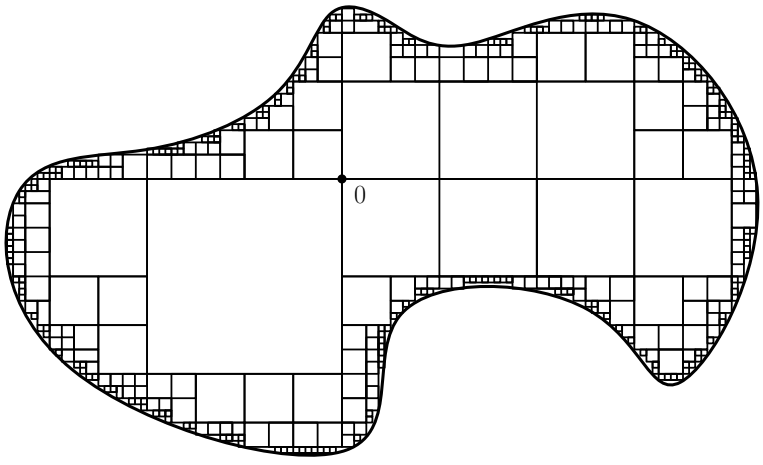


FIGURE 1 – Décomposition d'un ouvert en pavés dyadiques semi-ouverts disjoints

Subdivisions dyadiques

imbriquement et monotonie

Définition

*Les **nombres dyadiques** sont les réels de la forme $\frac{k}{2^n}$ où $k \in \mathbb{Z}$ et $n \in \mathbb{N}$. Ce sont, à la base 2, ce que les nombres décimaux sont à la base 10.*

Pour $n \in \mathbb{N}$, on appellera les réels de la forme $\frac{k}{2^n}$, où $k \in \mathbb{Z}$, les **nombres dyadiques d'ordre** n.
Pour tout $n \in \mathbb{N}$, les intervalles $\left[\frac{k}{2^n}, \frac{k+1}{2^n}\right[$, où k parcourt $\mathbb{Z}$, partitionnent $\mathbb{R}$. Appelons cette subdivision de $\mathbb{R}$ la subdivision dyadique d'ordre n. On constate que la subdivision d'ordre $n+1$ est un *raffinement* de la subdivision d'ordre n, ce qui n'aurait pas été le cas pour la suite de subdivisions par les intervalles $\left[\frac{k}{n}, \frac{k+1}{n}\right[$, $k \in \mathbb{Z}$.
Plus généralement, pour $n \in \mathbb{N}$, les pavés dyadiques d'ordre n semi-ouverts $\left[\frac{k_1}{2^n}, \frac{k_1+1}{2^n}\right[\times \cdots \times \left[\frac{k_d}{2^n}, \frac{k_d+1}{2^n}\right[$, où $k_1, \ldots, k_d$ parcourent $\mathbb{Z}$, partitionnent $\mathbb{R}^d$, et on a la même propriété de raffinement.

Proposition

Soit $f : E \to \mathbb{R}_+$ une fonction mesurable. Il existe une suite croissante $(\varphi_n)_{n\geq 0}$ de fonctions étagées positives qui converge vers f. De plus, on peut choisir la fonction φ_n à valeurs dans les nombres dyadiques d'ordre n.

Démonstration : On définit (en l'écrivant d'une façon ou d'une autre, de façon équivalente)

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

$$\varphi_n(x) = \max\left\{ \frac{k}{2^n} \middle| k \in \{0,1,\ldots,n2^n\} \text{ et } \frac{k}{2^n} \leq f(x) \right\}.$$

$$\varphi_n(x) = \min\left(n, \frac{\lfloor 2^n f(x) \rfloor}{2^n}\right)$$

Autrement dit, pour tout $x \in E$, $\varphi_n(x)$ est le plus grand nombre dyadique d'ordre n inférieur à $f(x)$, tronqué au-delà la valeur n (ceci, afin que φ_n ne prenne qu'un nombre fini de valeurs). Pour tout $x \in \mathbb{R}$, dès que $n > f(x)$, on a $\varphi_n(x) \leq f(x) < \varphi_n(x) + \frac{1}{2^n}$, d'où la convergence de φ_n vers f. Et comme tout nombre dyadique d'ordre $n+1$ est un nombre dyadique d'ordre n, on a, pour tout $x \in \mathbb{R}$, $\varphi_{n+1}(x) \geq \varphi_n(x)$. **Dessin ?** ■

La proposition précédente est importante pour l'étude (et la définition) de l'intégrale de Lebesgue ; en particulier, elle donne un moyen de l'obtenir comme une limite, en subdivisant l'*intervalle d'arrivée* de la fonction : par le théorème de convergence monotone, pour toute mesure μ sur E,

$$\int f d\mu = \lim_n{\uparrow} \int \varphi_n d\mu = \lim_n{\uparrow} \sum_{k=0}^{n2^n} \frac{k}{2^n} \mu\big(f^{-1}\left(\left[\frac{k}{2^n}, \frac{k+1}{2^n}\right[\right)\big) + n\mu\big(f^{-1}\left([n,+\infty[\right)\big).$$

(Si l'intégrale est finie, on peut voir que le dernier terme tend vers 0 et peut donc être enlevé)
Inversement, l'intégrale de Riemann procède par subdivision de l'*espace de départ*, et une subdivision dyadique permet d'avoir de la monotonie, ainsi que l'illustre la proposition suivante :

Proposition

Soit $f : \mathbb{R}^d \to \mathbb{R}$ une fonction continue. Il existe une suite croissante $(\varphi_n)_{n\geq 0}$ et une suite décroissante $(\psi_n)_{n\geq 0}$ de fonctions en escalier qui convergent vers f. De plus, on peut choisir les fonctions φ_n et ψ_n constantes sur chaque pavé dyadique semi-ouvert d'ordre n.

Démonstration : On considère

$$\varphi_n(x) = \inf_{C_n(x)} f \quad \text{et} \quad \psi_n(x) = \sup_{C_n(x)} f,$$

où $C_n(x)$ est le pavé dyadique semi-ouvert d'ordre n de $\mathbb{R}^d$ contenant x. La (dé)croissance vient à nouveau de l'imbriquement des pavés dyadiques, et la convergence résulte de la continuité de f. ■

Proposition

Soit U un ouvert de $\mathbb{R}^d$. Il existe une suite $(P_m)_{m\geq 0}$ de pavés semi-ouverts disjoints tels que $U = \bigcup_{m\geq 0} P_m$. De plus, on peut choisir ces pavés dyadiques.

Démonstration : On considère la famille $(P_m)_{m\geq 0}$ des pavés dyadiques semi-ouverts maximaux contenus dans U, et on vérifie qu'elle convient. Cela revient aussi à considérer la fonction φ_n de la démonstration précédente pour $f = \mathbf{1}_U$, car $\varphi_n = \mathbf{1}_{A_n}$, où $A_n \subset A_{n+1}$, $\bigcup_n A_n = U$, et $A_{n+1} \setminus A_n$ est réunion de pavés dyadiques d'ordre $n+1$. **Dessin ?** ■

Fonctions de classe $\mathcal{C}^\infty$ à support compact

Proposition

Il existe une fonction $f : \mathbb{R} \to \mathbb{R}_+$ de classe $\mathcal{C}^\infty$ telle que, pour $x \leq 0$, $f(x) = 0$.

Démonstration : On vérifie que la fonction $f : x \mapsto \begin{cases} e^{-1/x} & \text{si } x > 0 \\ 0 & \text{si } x \leq 0. \end{cases}$ est continue et que, pour tout $n \geq 1$, $f^{(n)}(x) \underset{x \to 0^+}{\longrightarrow} 0$ (on montre par récurrence que cette dérivée n-ième est de la forme $F(x)e^{-1/x}$ où F est une fraction rationnelle).
Il en résulte que f est de classe $\mathcal{C}^\infty$ sur $\mathbb{R}$, avec $f^{(n)}(0) = 0$ pour tout $n \geq 0$. ■

Corollaire

Il existe une fonction $\varphi : \mathbb{R} \to [0,1]$ de classe $\mathcal{C}^\infty$ telle que :
- *pour tout $t \leq 0$, $\varphi(t) = 0$;*
- *pour tout $t \geq 1$, $\varphi(t) = 1$;*
- *φ est strictement croissante sur $[0,1]$.*

Démonstration : Avec la notation de la proposition précédente, on observe que $g : t \mapsto f(t)f(1-t)$ est aussi une fonction $\mathcal{C}^\infty$ sur $\mathbb{R}$, positive, et que g vaut 0 en-dehors de $]0,1[$.
Si on définit alors

$$\varphi(x) = \frac{\int_0^x g(t)dt}{\int_0^1 g(t)dt}$$

on constate tout de suite que φ satisfait les conditions demandées. ■

Corollaire

Soit $a < u < v < b$. Il existe une fonction $\psi : \mathbb{R} \to [0,1]$ de classe $\mathcal{C}^\infty$ telle que
- *pour tout $t \notin]a,b[$, $\psi(t) = 0$;*
- *pour tout $t \in [u,v]$, $\psi(t) = 1$,*
- *ψ est strictement croissante sur $[a,u]$ et strictement décroissante sur $[v,b]$.*

Démonstration : Il suffit de jouer sur la fonction φ précédente : poser, pour $t \in \mathbb{R}$,

$$\psi(t) = \begin{cases} 0 & \text{si } t \leq a \\ \varphi\left(\frac{t-a}{u-a}\right) & \text{si } a < t < u \\ 1 & \text{si } u \leq t \leq v \\ \varphi\left(\frac{b-t}{b-v}\right) & \text{si } v < t < b \\ 0 & \text{si } t \geq b. \end{cases}$$

■

Proposition

Soit $a < b$. Il existe une suite croissante $(\psi_n)_{n \geq 0}$ de fonctions positives et de classe $\mathcal{C}^\infty$ qui converge ponctuellement vers $\mathbf{1}_{]a,b[}$.

Démonstration : Pour tout $n > \frac{2}{b-a}$, considérer la fonction ψ_n de la preuve précédente avec $u = a + \frac{1}{n}$ et $v = b - \frac{1}{n}$. ■