

Intégration et probabilités

GRÉGORY MIERMONT

L3 2014–2017
ENS de Lyon

Avant-propos

Ces notes correspondent au cours « Intégration et probabilités » donné au second semestre de 2014 à 2017 à l'Ecole Normale Supérieure de Lyon. Les prérequis de ce cours sont les fondamentaux de la théorie de la mesure : mesures positives, intégrales par rapport à une mesure, théorèmes limites usuels, mesure de Lebesgue, espaces L^p .

Le cours contient deux parties. Outre quelques compléments d'intégration sur la convolution et le changement de variables, la première partie donne les bases de l'analyse de Fourier : séries de Fourier pour les fonctions périodiques sur \mathbb{R} , et la transformation de Fourier des fonctions intégrables et des mesures de probabilités sur \mathbb{R}^d . La seconde partie est une introduction à la théorie moderne des probabilités, en se focalisant sur les notions fondamentales suivantes :

- espaces de probabilités, variables aléatoires
- indépendance
- théorèmes limites : lois des grands nombres et théorème central limite.

Ces points sont illustrés par des exemples concrets, ponctués par deux chapitres de compléments.

Table des matières

| | |
|---|-----------|
| Avant-propos | 5 |
| I Introduction à l'analyse de Fourier | 11 |
| 1 Quelques compléments d'intégration | 15 |
| Quelques notations | 15 |
| 1.1 Compléments sur les espaces L^p | 15 |
| 1.2 Lemme de Riemann-Lebesgue | 16 |
| 1.3 Convolution | 17 |
| 1.4 Approximations de l'unité | 19 |
| 2 Séries de Fourier | 23 |
| 2.1 Polynômes et séries trigonométriques | 24 |
| 2.2 Série de Fourier d'une fonction | 25 |
| 2.3 Convergence des séries de Fourier dans L^2 | 27 |
| 2.4 Convergence ponctuelle des séries de Fourier | 28 |
| 2.4.1 Le cas \mathcal{C}^1 par morceaux | 28 |
| 2.4.2 Convergence de Cesaro | 30 |
| 2.5 *Preuve du théorème de Stone-Weierstrass | 31 |
| 3 La transformée de Fourier dans \mathbb{R}^d | 33 |
| 3.1 Transformée de Fourier d'une fonction intégrable | 33 |
| b. Continuité, lemme de Riemann-Lebesgue | 34 |
| c. Régularité | 35 |
| d. Lien avec la convolution | 37 |
| 3.2 L'exemple de la densité gaussienne. | 37 |
| 3.3 La formule d'inversion | 38 |
| 3.4 La transformée de Fourier L^2 | 40 |
| 3.5 Transformée de Fourier d'une mesure signée | 42 |
| 3.6 Une application à l'analyse de l'équation de la chaleur | 44 |
| 4 Changement de variables | 47 |
| 4.1 Mesure image | 47 |
| 4.2 Coordonnées polaires dans \mathbb{R}^d | 47 |
| 4.3 Changement de variables linéaire | 48 |
| 4.4 Changement de variables \mathcal{C}^1 | 50 |

| | |
|---|----|
| II Bases des probabilités | 51 |
| 5 Bases de la théorie des probabilités | 55 |
| 5.1 Espaces de probabilités, variables aléatoires | 55 |
| Premiers exemples d'espaces de probabilités. | 55 |
| Une infinité de lancers de pièces ? | 56 |
| Variables aléatoires. | 57 |
| Variables aléatoires discrètes. | 58 |
| Variables aléatoires à densité. | 59 |
| 5.2 Espérance d'une variable aléatoire | 59 |
| a. Définition et formule de transfert | 59 |
| b. Caractérisation de la loi à l'aide de l'espérance | 60 |
| c. Moments d'une variable aléatoire | 61 |
| d. Variance et covariance | 62 |
| e. Médiane et quantiles | 64 |
| 5.3 Fonctions associées à une variable aléatoire | 65 |
| a. Fonction de répartition | 65 |
| b. Fonction génératrice | 66 |
| c. Fonction caractéristique | 67 |
| d. Transformée de Laplace | 68 |
| 5.4 Exemples fondamentaux de lois de variables aléatoires | 69 |
| a. Lois discrètes | 69 |
| Loi uniforme sur un ensemble fini | 69 |
| Loi de Bernoulli | 70 |
| Loi binomiale | 70 |
| Loi géométrique | 70 |
| Loi de Poisson | 70 |
| b. Lois à densité | 71 |
| Loi uniforme sur un sous-ensemble mesurable de \mathbb{R}^d | 71 |
| Lois exponentielles | 71 |
| Lois gaussiennes sur \mathbb{R} | 72 |
| 6 Indépendance | 73 |
| 6.1 Probabilités conditionnelles élémentaires | 73 |
| 6.2 Indépendance d'événements | 74 |
| 6.3 Indépendance de σ -algèbres | 75 |
| 6.4 Indépendance de variables aléatoires | 77 |
| a. σ -algèbre associée à une variable aléatoire | 77 |
| b. Indépendance de variables aléatoires | 77 |
| c. Critères d'indépendance de variables aléatoires | 78 |
| 6.5 Sommes de variables aléatoires indépendantes | 80 |
| 6.6 Lemme de Borel-Cantelli | 82 |
| 6.6.1 L'énoncé, et un exemple | 82 |

| | |
|---|------------|
| Exemple. Nombre de « pile » consécutifs | 82 |
| 6.6.2 Lemme « réciproque » | 84 |
| Une mesure « uniforme » sur \mathbb{N} ? | 84 |
| Motifs dans une suite de pile ou face | 85 |
| 6.7 Loi du 0-1 de Kolmogorov | 86 |
| 6.8 Complément : existence d'une suite de variables aléatoires indépendantes | 87 |
| 7 Lois des grands nombres | 91 |
| 7.1 Différentes notions de convergence pour des variables aléatoires | 91 |
| a. Convergence presque sûre | 91 |
| b. Convergence L^p | 91 |
| c. Convergence en probabilité | 92 |
| 7.2 La loi forte des grands nombres | 94 |
| 7.2.1 Le cas L^4 | 95 |
| 7.2.2 Le cas L^2 | 96 |
| 7.2.3 Le cas L^1 par la méthode d'écrêtement | 97 |
| 7.2.4 Le cas L^1 : une seconde preuve | 99 |
| 7.2.5 Quelques ramifications de la loi des grands nombres | 100 |
| Cas d'une espérance bien définie, mais infinie | 100 |
| Cas où l'espérance n'existe plus nécessairement | 100 |
| 7.3 Quelques applications | 100 |
| 7.3.1 Marches aléatoires non centrées | 100 |
| 7.3.2 Approximation d'intégrales par la méthode de Monte-Carlo | 101 |
| 8 Convergence en loi et théorème central limite | 103 |
| 8.1 Convergence étroite, convergence en loi | 103 |
| 8.1.1 Exemples élémentaires | 104 |
| Lois sur \mathbb{N} | 104 |
| Lemme de Scheffé et convergence ponctuelle de densités | 105 |
| Exemple d'approximation de la mesure de Lebesgue | 106 |
| 8.1.2 Liens avec les autres notions de convergence. | 106 |
| Convergence en probabilité | 106 |
| Convergence en variation totale | 107 |
| 8.1.3 Caractérisations de la convergence en loi | 107 |
| 8.2 Le théorème central limite | 111 |
| Application aux statistiques : estimation paramétrique et intervalles de confiance | 112 |
| 8.3 Vecteurs aléatoires gaussiens et théorème central limite multidimensionnel | 114 |
| 8.3.1 Vecteurs aléatoires gaussiens | 114 |
| 8.3.2 Théorème central limite : le cas de \mathbb{R}^d | 117 |
| 8.3.3 Une application : le test d'adéquation du χ^2 | 118 |
| 8.4 L'inégalité de Hoeffding | 119 |
| 9 Récurrence et transience pour la marche aléatoire simple sur \mathbb{Z}^d | 123 |

| | |
|--|-----|
| 10 Processus de branchement | 129 |
| Phase sous-critique : $m < 1$ | 131 |
| Phase critique : $m = 1$ | 131 |
| Phase sur-critique : $m > 1$ | 132 |
| Références | 135 |

Partie I

Introduction à l'analyse de Fourier

Chapitre 1

Quelques compléments d'intégration

Quelques notations

Si $d \geq 1$ est un entier, notons $\langle x, y \rangle = \sum_{i=1}^d x_i y_i$ le produit scalaire usuel de \mathbb{R}^d , et $|x| = \sqrt{\langle x, x \rangle}$ la norme euclidienne.

On considérera des fonctions mesurables définies sur l'espace mesuré $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d), \lambda_d)$, où $\mathcal{B}(\mathbb{R}^d)$ est la tribu borélienne de \mathbb{R}^d , et λ_d est la mesure de Lebesgue. On notera en général

$$\int_{\mathbb{R}^d} f(x) dx = \int_{\mathbb{R}^d} f(x) \lambda_d(dx).$$

Sauf mention contraire, les fonctions considérées seront à valeurs dans le corps \mathbb{C} des nombres complexes, lui-même muni de la tribu borélienne.

1.1 Compléments sur les espaces L^p

Soit $p \in [1, \infty[$. On note $\mathcal{L}^p = \mathcal{L}^p(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d), \lambda_d)$ l'ensemble des telles fonctions f mesurables telles que $|f|^p$ est intégrable, et on note \mathcal{L}^∞ l'ensemble des fonctions mesurables essentiellement bornées, c'est-à-dire telles qu'il existe $M > 0$ tel que $\lambda_d(\{|f| > M\}) = 0$.

On note également L^p l'ensemble quotient \mathcal{L}^p / \equiv , où l'on a noté $f \equiv g$ si $\lambda_d(\{f \neq g\}) = 0$. On le munit de la norme L^p usuelle notée $\|\cdot\|_p$.

Nous aurons recours au résultat suivant. Si $f: \mathbb{R}^d \rightarrow \mathbb{C}$ est une fonction mesurable, et $y \in \mathbb{R}^d$, on note

$$\tau_y f(x) = f(x - y), \quad x \in \mathbb{R}^d$$

Lemme 1.1. *Pour tout $p \in [1, \infty]$ et tout $y \in \mathbb{R}^d$, l'application $f \mapsto \tau_y f$ définit une isométrie linéaire de L^p sur lui-même. De plus, si $1 \leq p < \infty$ et si $f \in L^p$, l'application $y \mapsto \tau_y f$ est uniformément continue de \mathbb{R}^d dans L^p .*

Démonstration. Fixons d'abord $y \in \mathbb{R}^d$. Il est évident que deux fonctions f et g sont égales presque partout au sens de Lebesgue si et seulement si il en est de même de $\tau_y f$ et $\tau_y g$, et donc τ_y induit bien une transformation de L^p , qui est clairement linéaire et préserve la norme (y compris pour $p = \infty$).

On se donne alors $p \neq \infty$. Soit $f \in L^p$, et $\varepsilon > 0$. Soit g une fonction continue à support compact telle que $\|f - g\|_p < \varepsilon/3$. Alors pour tout $x, y \in \mathbb{R}^d$, on a

$$\|\tau_y f - \tau_x f\|_p \leq 2\|f - g\|_p + \|\tau_y g - \tau_x g\|_p \leq \frac{2\varepsilon}{3} + \int_{\mathbb{R}^d} |g(z - (x - y)) - g(z)| dz,$$

où l'on a utilisé l'inégalité triangulaire et la propriété d'isométrie de τ_y et τ_x , et un changement de variable affine simple. Comme g est à support compact, si l'on suppose que $|x - y| \leq 1$, on voit que la dernière intégrale est égale à la même intégrale restreinte au compact $K = V_1(\text{supp}(g))$, où par définition $V_r(A) = \{x \in \mathbb{R}^d : \inf_{y \in A} |x - y| \leq r\}$ est le r -voisinage fermé de A . Le compact K ne dépend plus de x et y , et par conséquent on conclut que la dernière intégrale ci-dessus (une fois restreinte à K) converge vers 0 lorsque $|x - y| \rightarrow 0$, par convergence dominée. On a bien montré qu'il existe un $\alpha > 0$ tel que $|x - y| < \alpha$ implique que $\|\tau_y f - \tau_x f\|_p \leq \varepsilon$, comme voulu. \square

Question: où a-t-on utilisé le fait que $p < \infty$?

Exemple 1.2. Si $A \subset \mathbb{R}$ est un ensemble mesurable avec $\lambda(A) > 0$, alors l'ensemble $A - A = \{x - y : x, y \in A\}$ contient un voisinage de 0.

En effet, supposons sans perte de généralité que $\lambda(A) \in]0, \infty[$, quitte à prendre l'intersection avec un intervalle compact assez grand. Alors $\mathbf{1}_A$ est dans L^1 , et par conséquent $\tau_h \mathbf{1}_A$ converge dans L^1 vers $\mathbf{1}_A$ lorsque $h \rightarrow 0$. Mais $\tau_h \mathbf{1}_A(x) = \mathbf{1}_{A+h}(x)$ et donc $\|\tau_h \mathbf{1}_A - \mathbf{1}_A\|_1 = \lambda(A \Delta (A + h))$ converge vers 0 lorsque $h \rightarrow 0$, où Δ désigne la différence symétrique.

Or on a $\lambda(A \cap (A + h)) = \lambda(A \cup (A + h)) - \lambda(A \Delta (A + h))$, qui est supérieur à $\lambda(A) - \lambda(A \Delta (A + h))$, et on conclut que $A \cap (A + h)$ est non vide car de mesure strictement positive pour tout $|h|$ assez petit. cela revient à dire que h appartient à $A - A$ dès que $|h|$ est assez petit.

1.2 Lemme de Riemann-Lebesgue

Le théorème ci-dessous traite du comportement à l'infini de certaines intégrales « oscillantes ». Nous verrons très vite que l'intégrale définie dans l'énoncé est, à quelques détails près, la transformée de Fourier de f en ξ .

Théorème 1.3. (Lemme de Riemann-Lebesgue) *Pour tout élément $f \in L^1$, l'intégrale*

$$\int_{\mathbb{R}^d} f(x) \exp(i\langle \xi, x \rangle) dx$$

est bien définie pour tout $\xi \in \mathbb{R}^d$, et converge vers 0 lorsque $|\xi| \rightarrow \infty$.

Démonstration. Tout d'abord, il est clair que pour tout ξ , l'intégrale ci-dessus est bien définie puisque $|f(x) \exp(i\langle \xi, x \rangle)| = |f(x)|$ est intégrable en x .

On démontre d'abord le résultat pour f de la forme $f(x) = \mathbf{1}_Q(x)$, où $Q = \prod_{j=1}^d [a_j, b_j]$ est un pavé. Dans ce cas,

$$\int_{\mathbb{R}^d} f(x) \exp(i\langle \xi, x \rangle) dx = \prod_{j=1}^d \left(\frac{e^{i\xi_j b_j} - e^{i\xi_j a_j}}{i\xi_j} \right)$$

où le j -ème terme du produit s'interprète comme $(b_j - a_j)$ si $\xi_j = 0$. Clairement, ce produit tend vers 0 lorsque $|\xi| \rightarrow \infty$. Par linéarité, on obtient le même résultat pour les fonctions f qui sont combinaisons linéaires de telles indicatrices. En se restreignant aux pavés Q dyadiques, c'est-à-dire pour lesquels il existe des entiers n, k_1, \dots, k_d tels que $a_i = k_i 2^{-n}$ et $b_i = (k_i + 1) 2^{-n}$ avec les notations ci-dessus, on constate par un argument aisé de compacité que les telles combinaisons linéaires sont denses dans l'ensemble $\mathcal{C}_c(\mathbb{R}^d)$ des fonctions continues à support compact, pour la norme L^1 . En utilisant la densité de $\mathcal{C}_c(\mathbb{R}^d)$ dans L^1 , on conclut que pour tout $f \in L^1$, et pour tout $\varepsilon > 0$, il existe une fonction g qui est une combinaison linéaire d'indicatrices de pavés telle que $\|f - g\|_1 < \varepsilon$. On a alors

$$\left| \int_{\mathbb{R}^d} f(x) \exp(i\langle \xi, x \rangle) dx \right| \leq \|f - g\|_1 + \left| \int_{\mathbb{R}^d} g(x) \exp(i\langle \xi, x \rangle) dx \right|,$$

et on déduit que la limite supérieure du membre de gauche lorsque $|\xi| \rightarrow \infty$ est majorée par ε . Comme ε est arbitraire, on conclut.

On peut avoir recours à une autre méthode, également instructive. Tirant parti de la formule $e^{i\pi} = -1$, on peut réécrire

$$\begin{aligned} \int_{\mathbb{R}^d} f(x) \exp(i\langle \xi, x \rangle) dx &= - \int_{\mathbb{R}^d} \exp\left(i\left\langle \xi, x + \frac{\pi \xi}{|\xi|^2} \right\rangle\right) f(x) dx \\ &= - \int_{\mathbb{R}^d} \exp(i\langle \xi, x \rangle) f\left(x - \frac{\pi \xi}{|\xi|^2}\right) dx. \end{aligned}$$

De ce fait, on a

$$2 \int_{\mathbb{R}^d} f(x) \exp(i\langle \xi, x \rangle) dx = \int_{\mathbb{R}^d} \exp(i\langle \xi, x \rangle) \left(f(x) - f\left(x - \frac{\pi \xi}{|\xi|^2}\right) \right) dx,$$

et on déduit par l'inégalité triangulaire que le module est majoré par

$$\|\tau_{\pi \xi / |\xi|^2} f - f\|_1.$$

Lorsque $|\xi| \rightarrow \infty$, on a $\xi / |\xi|^2 \rightarrow 0$, et par conséquent la preuve découle immédiatement du Lemme 1.1. \square

1.3 Convolution

Soit $f, g \in \mathcal{L}^1$. Le produit de convolution de f par g , noté $f * g$, est défini par la formule

$$f * g(x) = \int_{\mathbb{R}^d} f(x - y) g(y) dy = \int_{\mathbb{R}^d} f(y) g(x - y) dy = g * f(x),$$

ce qui a bien un sens à condition que $\int_{\mathbb{R}^d} |f(x - y) g(y)| dy < \infty$.

Proposition 1.4. *La formule ci-dessus est bien définie pour λ_d -presque tout x , et définit un élément de L^1 pour lequel on a $\|f * g\|_1 \leq \|f\|_1 \|g\|_1$.*

Démonstration. La fonction $(x, y) \mapsto |f(x-y)g(y)|$ est mesurable et positive sur l'espace produit $\mathbb{R}^d \times \mathbb{R}^d$ muni de la tribu produit, et de plus, par le théorème de Fubini, son intégrale est

$$\int_{\mathbb{R}^d} dy |g(y)| \int_{\mathbb{R}^d} |f(x-y)| dx = \|f\|_1 \|g\|_1 < \infty.$$

Par conséquent, on déduit des résultats généraux sur les espaces produit que la fonction $(x, y) \mapsto f(x-y)g(y)$ est dans $L^1(\mathbb{R}^d \times \mathbb{R}^d, \lambda_d \otimes \lambda_d)$ et que son intégrale par rapport à la variable y est finie pour λ_x -presque tout x , et intégrable en la variable x . La majoration de la norme provient alors de l'inégalité triangulaire. \square

Il existe de nombreuses autres situations où la formule définissant la convolution est bien définie. On donne deux tels exemples.

Proposition 1.5. *Supposons que $f \in L^p$ et $g \in L^q$, où $p, q \in [1, \infty]$ et $(1/p) + (1/q) = 1$. Alors $f * g(x)$ est bien défini pour tout x , et définit une fonction uniformément continue et bornée sur \mathbb{R}^d .*

Démonstration. Pour le fait que $f * g(x)$ est bien défini et est borné en x , il suffit de constater que par l'inégalité de Hölder,

$$\int_{\mathbb{R}^d} |f(x-y)g(y)| dy \leq \|\tau_{-x}f\|_p \|g\|_q = \|f\|_p \|g\|_q,$$

ce qui fait que $y \mapsto f(x-y)g(y)$ est bien intégrable pour tout $x \in \mathbb{R}^d$. Ensuite, on écrit, toujours par l'inégalité de Hölder,

$$|f * g(x) - f * g(y)| \leq \|\tau_{-x}f - \tau_{-y}f\|_p \|g\|_q$$

et on conclut par le lemme 1.1 si $p < \infty$, dans le cas contraire on échange les rôles de f et g . \square

On voit dans le résultat précédent la première expression d'un fait général : la convolution a tendance à *régulariser* les fonctions. Si par exemple f est une fonction de classe $\mathcal{C}_c^\infty(\mathbb{R}^d)$, on pourra par exemple montrer aisément que pour tout $g \in L^p(\mathbb{R}^d)$ pour un $p \in [1, \infty]$, $f * g$ est de classe $\mathcal{C}^\infty(\mathbb{R}^d)$, avec toutes ses dérivées bornées.

Dans la suite, nous aurons besoin d'une troisième situation où le produit de convolution est bien défini.

Proposition 1.6. *Soit $f \in L^1$ et $g \in L^p$ pour un $p \in [1, \infty]$. Alors le produit de convolution $f * g(x)$ est bien défini pour λ_d -presque tout x , et définit un élément de L^p . De plus, on a $\|f * g\|_p \leq \|f\|_1 \|g\|_p$.*

Démonstration. Le résultat pour $p = \infty$ est traité par la proposition précédente. On suppose donc $p < \infty$, et que f n'est pas nulle presque partout (le résultat est trivial dans le cas contraire). On utilise alors le fait que pour tout x , la mesure $|f(x-y)|dy / \|f\|_1$ est une mesure de probabilités, ce qu'il permet d'utiliser l'inégalité de Jensen :

$$\int_{\mathbb{R}^d} dx \left(\int_{\mathbb{R}^d} |f(x-y)g(y)| dy \right)^p \leq \|f\|_1^p \int_{\mathbb{R}^d} dx \int_{\mathbb{R}^d} dy \frac{|f(x-y)|}{\|f\|_1} |g(y)|^p$$

et le majorant vaut $\|f\|_1^p \|g\|_p^p$, qui est fini par hypothèse. Cela montre que $f * g(x)$ est bien défini pour presque tout x , et la conclusion suit aisément par inégalité triangulaire. \square

Enfin, notons que le produit de convolution s'étend aux mesures de la façon suivante.

Définition 1.7. Soit μ, ν deux mesures positives finies, ou signées, sur \mathbb{R}^d . Le produit de convolution de μ par ν , noté $\mu * \nu$, est la mesure sur \mathbb{R}^d définie comme mesure image de la mesure produit $\mu \otimes \nu$ par l'application $(x, y) \mapsto x + y$ de $\mathbb{R}^d \times \mathbb{R}^d$ dans \mathbb{R}^d . Autrement dit, pour toute fonction f mesurable bornée, on a

$$\int_{\mathbb{R}^d} f(z) \mu * \nu(dz) = \int_{\mathbb{R}^d \times \mathbb{R}^d} f(x + y) \mu(dx) \nu(dy).$$

Notons que si $\mu(dx) = f(x)dx$ est absolument continue, avec (nécessairement) f dans L^1 , le produit de convolution $\mu * \nu$ est la mesure absolument continue dont la densité est donnée par $f * \nu$ définie par :

$$f * \nu(x) = \int_{\mathbb{R}^d} f(x - y) \nu(dy).$$

La preuve est aisée, et laissée en exercice. Si à son tour $\nu(dx) = g(x)dx$ est à densité, on a $f * \nu = f * g$.

1.4 Approximations de l'unité

Avec les notations de la fin de la section précédente, notons que $f * \delta_0 = f$ pour toute fonction f dans L^1 . On peut montrer (cela sera facile avec la transformée de Fourier) qu'il n'existe pas de fonction g qui puisse remplacer la mesure δ_0 dans ce rôle, c'est-à-dire telle que $f * g = f$ pour toute fonction f dans L^1 . Néanmoins, on peut trouver des fonctions qui remplissent *presque* ce rôle. Il s'agit de fonctions d'intégrale 1 (comme δ_0) qui sont « très concentrées » autour de 0, au sens suivant.

Définition 1.8. On dit que la suite de fonctions mesurables $(\gamma_n, n \geq 0)$ est une approximation de l'unité si

- $\sup_{n \geq 0} \|\gamma_n\|_1 < \infty$,
- $\int_{\mathbb{R}^d} \gamma_n(x) dx = 1$ pour tout $n \geq 0$, et
- pour tout $\delta > 0$ on a $\int_{\{|x| > \delta\}} |\gamma_n(x)| dx \rightarrow 0$.

Une classe importante de partitions de l'unité s'obtient en se donnant une fonction $\gamma \in L^1$ d'intégrale 1, et en posant $\gamma_n(x) = n^d \gamma(nx)$, ou plus généralement $\gamma_n(x) = a_n^d \gamma(a_n x)$ pour une suite $(a_n, n \geq 0)$ de limite $+\infty$. Remarquons que si $\gamma_n \geq 0$ pour tout n , le premier point est impliqué par le second.

Proposition 1.9. *Soit $(\gamma_n, n \geq 0)$ une approximation de l'unité et f une fonction continue bornée sur \mathbb{R}^d . Alors $\gamma_n * f$ converge vers f uniformément sur les compacts. Si de plus f est à support compact, alors $\gamma_n * f$ converge uniformément sur \mathbb{R}^d .*

Démonstration. Comme γ_n est positive d'intégrale 1, on a pour tout $x \in \mathbb{R}^d$, et tout $\delta \in]0, 1[$,

$$\begin{aligned} |\gamma_n * f(x) - f(x)| &= \left| \int_{\mathbb{R}^d} \gamma_n(y) dy (f(x-y) - f(x)) \right| \\ &\leq \int_{\mathbb{R}^d} |\gamma_n(y)| dy |f(x-y) - f(x)| \\ &\leq 2\|f\|_\infty \int_{\{|y|>\delta\}} |\gamma_n(y)| dy \\ &\quad + C \sup \{|f(x-y) - f(x)| : |y| \leq \delta\}. \end{aligned}$$

où C est un majorant uniforme des normes $\|\gamma_n\|_1$.

Si x prend ses valeurs dans un compact K donné, on peut utiliser l'uniforme continuité de f sur le 1-voisinage fermé $V_1(K) = \{x \in \mathbb{R}^d : \exists y \in K, |x-y| \leq 1\}$ de K pour obtenir le résultat : pour un $\varepsilon > 0$ donné, on choisit $\delta \in]0, 1[$ tel que le deuxième terme du majorant soit plus petit que $\varepsilon/2$, et on a alors que pour tout n assez grand, le premier terme est majoré par $\varepsilon/2$. Si de plus f est à support compact, on a automatiquement l'uniforme continuité de f partout, et il n'est pas nécessaire de restreindre x à un compact dans l'argument précédent. \square

Proposition 1.10. *Fixons $p \in [1, \infty[$. Soit $(\gamma_n, n \geq 0)$ une approximation de l'unité, et $f \in L^p$. Alors $\|\gamma_n * f - f\|_p \rightarrow 0$ lorsque $n \rightarrow \infty$.*

Démonstration. On écrit, en utilisant que γ_n est d'intégrale 1,

$$\|\gamma_n * f - f\|_p^p = \int_{\mathbb{R}^d} dx \left| \int_{\mathbb{R}^d} \gamma_n(y) (f(x-y) - f(x)) dy \right|^p.$$

Ensuite, on utilise l'inégalité triangulaire, et on divise et remultiplie γ_n par sa norme 1 pour obtenir que ceci est majoré par

$$\|\gamma_n\|_1^p \int_{\mathbb{R}^d} dx \left(\int_{\mathbb{R}^d} \frac{|\gamma_n(y)|}{\|\gamma_n\|_1} |f(x-y) - f(x)| dy \right)^p$$

comme la mesure $|\gamma_n(y)| dy / \|\gamma_n\|_1$ est une mesure de probabilités, on peut majorer par l'inégalité de Jensen, et on trouve le majorant

$$\|\gamma_n\|_1^p \int_{\mathbb{R}^d} dx \int_{\mathbb{R}^d} \frac{|\gamma_n(y)|}{\|\gamma_n\|_1} |f(x-y) - f(x)|^p dy$$

À ce stade, on applique le théorème de Tonelli pour changer l'ordre d'intégration, et on majore $\|\gamma_n\|_1$ uniformément par une constante C , ce qui donne le majorant

$$\begin{aligned} C^{p-1} \int_{\mathbb{R}^d} |\gamma_n(y)| dy \|\tau_y f - f\|_p &\leq 2C^{p-1} \|f\|_p \int_{\{|x|>\delta\}} |\gamma_n(y)| dy \\ &\quad + C^{p-1} \sup \{\|\tau_y f - f\|_p : |y| \leq \delta\} \end{aligned}$$

pour tout $\delta > 0$. Si l'on se donne $\varepsilon > 0$, on peut choisir $\delta > 0$ tel que le second terme de droite soit borné supérieurement par ε , par le lemme 1.1. En faisant alors tendre $n \rightarrow \infty$ pour ce choix de δ , le premier terme de droite converge vers 0 par définition d'une approximation de l'unité, ce qui donne le résultat. \square

Comme exemple d'application, citons le théorème d'approximation de Weierstrass pour les polynômes.

Théorème 1.11. *Soit f une fonction continue sur un intervalle compact $[a, b]$. Alors pour tout $\varepsilon > 0$, il existe un polynôme P réel tel que $\sup \{|f(x) - P(x)| : x \in [a, b]\} < \varepsilon$.*

Démonstration. Quitte à changer f en $f(2(b-a)x + (3a-b)/2)$, on peut supposer que $a = 1/4$ et $b = 3/4$ sans perte de généralité. On prolonge la fonction f à \mathbb{R} tout entier par la valeur 0 en dehors de $[0, 1]$, et par des fonctions affines sur $[0, 1/4]$ et $[3/4, 1]$ de sorte que la fonction prolongée, encore appelée f , soit continue à support dans $[0, 1]$. Posons $\gamma_n(x) = c_n(1-x^2)^n \mathbf{1}_{\{|x| \leq 1\}}$, où

$$c_n = \frac{1}{\int_{-1}^1 (1-x^2)^n dx}$$

de sorte que $\int_{\mathbb{R}} \gamma_n = 1$. Comme $\gamma_n \geq 0$, on aura montré que $(\gamma_n, n \geq 0)$ est une approximation de l'unité si $\int_{|y| > \delta} \gamma_n(y) dy \rightarrow 0$ pour tout $\delta > 0$. Pour cela, on montre aisément (c_n^{-1} est une intégrale de Wallis d'ordre impair) que

$$c_n = \frac{(2n+1)!}{2 \cdot 4^n (n!)^2} \underset{n \rightarrow \infty}{\sim} \sqrt{\frac{n}{\pi}}$$

où l'on a utilisé la formule de Stirling pour trouver l'équivalent. On voit donc que pour tout $\delta \in]0, 1[$,

$$c_n \int_{|y| > \delta} (1-y^2)^n dy \leq 2c_n (1-\delta^2)^n \xrightarrow{n \rightarrow \infty} 0,$$

ce qui montre bien que $(\gamma_n, n \geq 0)$ est une approximation de l'unité. Or

$$\gamma_n * f(x) = c_n \int_0^1 f(y) (1-(x-y)^2)^n \mathbf{1}_{\{|x-y| \leq 1\}} dy.$$

Pour $x, y \in [0, 1]$ on a que $|x-y| \leq 1$, et par conséquent on peut enlever l'indicatrice dans l'intégrale précédente. En développant le produit, on voit qu'en restriction à $[0, 1]$, la fonction $\gamma_n * f(x)$ est un polynôme (en x) de degré au plus $2n$. Par la proposition 1.9, on a convergence uniforme vers la fonction f . D'où le résultat. \square

Chapitre 2

Séries de Fourier

Dans ce chapitre, on étudie la décomposition d'une fonction périodique de \mathbb{R} dans \mathbb{C} en termes de « signaux » élémentaires, les fonctions trigonométriques. Nous allons nous concentrer sur les fonctions 2π -périodiques, sachant que toute la discussion de ce chapitre peut être faite dans le cas d'une période quelconque. À l'origine de cette théorie, Fourier s'intéresse à l'équation décrivant la propagation de la chaleur dans \mathbb{R} , donnée par

$$\frac{\partial u}{\partial t} = \frac{1}{2} \frac{\partial^2 u}{\partial x^2}$$

et dont l'inconnue est une fonction $u(t, x)$ de deux variables, décrivant la température d'un milieu donné au point x et au temps t . On s'intéresse de plus à des solutions définies sur un domaine $[0, \infty[\times [-\pi, \pi]$. Fourier note que pour tout $n \in \mathbb{N}$ et $a_n, b_n \in \mathbb{R}$, les fonctions

$$(t, x) \mapsto \exp(-n^2 t / 2)(a_n \cos(nx) + b_n \sin(nx))$$

sont des solutions, ainsi que toute somme d'un nombre fini d'entre elles. Il stipule alors que toute solution est une superposition, éventuellement infinie, de telles solutions élémentaires. Cela pose une double question

- sous quelles conditions une série infinie de fonctions de la forme ci-dessus converge-t-elle ?
- sous quelles condition une fonction peut-elle se représenter sous la forme d'une telle série ?

Dans la suite, on note $\mathbb{T} = \mathbb{R} / 2\pi\mathbb{Z}$, que l'on identifie à l'intervalle $]-\pi, \pi]$, muni de la restriction de la mesure de Lebesgue $\lambda(dx) = dx \mathbf{1}_{\{-\pi < x \leq \pi\}} / 2\pi$. La renormalisation par 2π de la mesure de Lebesgue est utile en de nombreuses occasions, et elle sera systématique. En particulier, si f, g sont deux fonctions intégrables sur \mathbb{T} , on adoptera la notation renormalisée

$$f * g(x) = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(x-y)g(y)dy.$$

Une fonction f sur \mathbb{T} est naturellement associée à une fonction $\tilde{f}: \mathbb{R} \rightarrow \mathbb{C}$ périodique de période 2π , et vice-versa. Pour $k \geq 0$, on notera $\mathcal{C}^k(\mathbb{T}, \mathbb{C})$ l'espace des fonctions sur \mathbb{T} dont l'extension à \mathbb{R} tout entier est de classe $\mathcal{C}^k(\mathbb{R}, \mathbb{C})$. On notera $L^p(\mathbb{T}) = L^p(\mathbb{T}, \mathcal{B}(\mathbb{T}), \lambda)$.

L'espace $L^2(\mathbb{T})$ est muni du produit scalaire hermitien usuel, qui en fait un espace de Hilbert

$$(f, g) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \overline{f(x)} g(x) dx.$$

2.1 Polynômes et séries trigonométriques

Pour tout $n \in \mathbb{Z}$, notons e_n la fonction $e_n(x) = \exp(inx)$, qui est continue et 2π -périodique.

Lemme 2.1. *La famille $(e_n, n \in \mathbb{Z})$ est orthonormale dans $L^2(\mathbb{T})$.*

Démonstration. Il suffit de constater que si $n \neq m$,

$$(e_n, e_m) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \exp(i(m-n)x) = \left[\frac{\exp(i(n-m)x)}{i(n-m)} \right]_{-\pi}^{\pi} = 0,$$

et que cette même intégrale vaut 1 si $m = n$. □

Définition 2.2. *Une combinaison linéaire des fonctions $(e_n, n \in \mathbb{Z})$ est appelée un polynôme trigonométrique. Le degré d'un polynôme trigonométrique est la plus grande valeur de $|n|$ pour laquelle le coefficient de e_n est non nul.*

Notons que l'écriture $\sum_{n \in \mathbb{Z}} c_n e_n$ d'un polynôme trigonométrique, où $(c_n, n \in \mathbb{Z})$ est une suite à support fini, est unique, puisque $(e_n, n \in \mathbb{Z})$ est une famille libre.

Théorème 2.3. (Théorème d'approximation de Weierstrass) *L'espace $\text{Vect}(e_n, n \in \mathbb{Z})$ est dense dans $\mathcal{C}(\mathbb{T}, \mathbb{C})$: toute fonction continue sur \mathbb{T} est limite uniforme d'une suite de polynômes trigonométriques.*

Ce théorème est la conséquence d'un résultat très général.

Théorème 2.4. (Stone-Weierstrass) *Soit X un espace topologique compact, et \mathcal{A} une algèbre de fonctions continues $X \rightarrow \mathbb{C}$ contenant au moins une fonction constante, stable par conjugaison complexe $f \mapsto \bar{f}$, et qui sépare les points, au sens où pour tout $x, y \in X$ avec $x \neq y$, il existe $f \in \mathcal{A}$ telle que $f(x) \neq f(y)$.*

Alors \mathcal{A} est dense dans $\mathcal{C}(X, \mathbb{C})$ pour la norme uniforme.

On laisse en exercice le soin de vérifier que le théorème s'applique dans le cas où $X = \mathbb{T}$ et où \mathcal{A} est l'algèbre des polynômes trigonométriques. Nous donnerons un peu plus loin deux autres preuves, plus directes et *ad hoc*, du Théorème 2.3. Le théorème de Stone-Weierstrass sera démontré à la fin du chapitre.

Une série trigonométrique est une somme infinie de la forme

$$\sum_{n \in \mathbb{Z}} c_n e_n.$$

Bien sûr une telle série n'est pas définie pour tout choix de $(c_n, n \in \mathbb{Z})$. On a néanmoins le résultat suivant.

Proposition 2.5. Soit $(c_n, n \in \mathbb{Z})$ une suite sommable de nombres complexes. Alors la série trigonométrique $\sum_{n \in \mathbb{Z}} c_n e_n$ converge normalement vers une fonction f continue sur \mathbb{T} . De plus, on a que $c_n = c_n(f)$ pour tout $n \in \mathbb{Z}$.

Démonstration. La convergence normale est immédiate puisque $\|e_n\|_\infty = 1$. On déduit le résultat du théorème classique d'interversion entre somme et intégrale, conséquence de la convergence dominée. \square

Un exemple important de séries trigonométriques est donné à partir de séries entières. En effet, si $S(z) = \sum_{n \geq 0} s_n z^n$ est une série entière de rayon de convergence $R > 0$, alors pour tout $r \in [0, R[$, la série $S(re^{ix}) = \sum_{n \geq 0} r^n s_n e^{inx}$ converge normalement.

Exemple 2.6. La série trigonométrique suivante converge normalement pour tout $r \in [0, 1[$.

$$\sum_{n \in \mathbb{Z}} r^{|n|} e^{inx} = \frac{1 - r^2}{1 - 2r \cos(x) + r^2},$$

on l'appelle le noyau de Poisson, il joue un rôle important en analyse.

Si $f = \sum_{n \in \mathbb{Z}} c_n e_n$ est la somme d'une série entière normalement convergente, on peut retrouver le coefficient c_n par la formule

$$c_n = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(x) e^{-inx} dx = (e_n, f).$$

Il suffit pour le voir d'intervertir la somme et l'intégrale (ce qui est valide car la série converge uniformément, et l'intégrale est sur un compact), et utiliser le lemme 2.1. Ceci motive la définition ci-dessous.

2.2 Série de Fourier d'une fonction

Soit $f \in L^1(\mathbb{T})$. Le n -ème coefficient de Fourier de f , où $n \in \mathbb{Z}$, est par définition le nombre

$$c_n(f) = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(x) \exp(-inx) dx.$$

Si $f \in L^2(\mathbb{T})$, ce nombre est bien sûr égal au produit scalaire (e_n, f) , mais la quantité ci-dessus est bien définie dès que f est intégrable. La définition et le lemme de Riemann-Lebesgue montré au chapitre précédent donnent immédiatement le résultat suivant.

Proposition 2.7. Soit $f \in L^1(\mathbb{T})$. Alors on a que $|c_n(f)| \leq \|f\|_1$ pour tout $n \in \mathbb{Z}$. De plus, $c_n(f) \rightarrow 0$ lorsque $|n| \rightarrow \infty$.

Remarque 2.8. En revanche, il n'est pas vrai que toute suite de nombre complexes de limite nulle à l'infinie est la suite des coefficients de Fourier d'une fonction intégrable.

Pour $N \geq 0$, la N -ème somme de Fourier de f est par définition le polynôme trigonométrique

$$S_N f = \sum_{n=-N}^N c_n(f) e_n.$$

Soit $f \in L^1(\mathbb{T})$. Notons que l'on a une autre écriture de cette somme, en regroupant les termes deux par deux. En effet, pour tout $n > 0$,

$$\begin{aligned} c_n(f)e^{inx} + c_{-n}(f)e^{-inx} &= \frac{1}{\pi} \int_{-\pi}^{\pi} f(y) \cos(n(x-y)) dy \\ &= a_n(f) \cos(nx) + b_n(f) \sin(nx) \end{aligned}$$

où

$$a_n(f) = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) \cos(nx) dx, \quad b_n(f) = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) \sin(nx) dx.$$

Avec la convention $a_0(f) = (1/\pi) \int_{-\pi}^{\pi} f(x) dx = 2c_0(f)$, on obtient ainsi que

$$S_N f(x) = \frac{a_0(f)}{2} + \sum_{n=1}^N (a_n(f) \cos(nx) + b_n(f) \sin(nx)).$$

On appelle cette expression l'écriture réelle de la somme de Fourier de f . Noter que les coefficients $a_n(f), b_n(f)$, sont des nombres complexes en général.

Proposition 2.9. *Si f est à valeurs réelles, on a*

$$a_n(f) = 2\Re(c_n(f)), \quad \text{et} \quad b_n(f) = -2\Im(c_n(f)).$$

La question que l'on se pose alors est celle de la convergence de $S_N f$, lorsque $N \rightarrow \infty$. Un cas particulier relativement simple est quand la suite de coefficients de Fourier est sommable.

Proposition 2.10. *Soit $f \in L^1(\mathbb{T})$ telle que la famille $(c_n(f), n \in \mathbb{Z})$ est sommable, c'est-à-dire dans $\ell^1(\mathbb{Z})$. Alors sa série de Fourier $\sum_{n \in \mathbb{Z}} c_n(f) e_n$ converge normalement, et est de plus égale à f presque partout.*

Lemme 2.11. *Soit $f, g \in L^1(\mathbb{T})$ telles que $c_n(f) = c_n(g)$ pour tout $n \in \mathbb{Z}$. Alors $f = g$.*

Démonstration. Posons $h = f - g \in L^1(\mathbb{T})$, de sorte que $c_n(h) = 0$ pour tout $n \in \mathbb{Z}$. Alors on a que pour tout polynôme trigonométrique P ,

$$\frac{1}{2\pi} \int_{\mathbb{T}} P(x) h(x) dx = 0.$$

Par le théorème de Weierstrass, on en déduit que $\int_{\mathbb{T}} \psi(x) h(x) dx = 0$ pour toute fonction ψ continue sur \mathbb{T} . En utilisant la densité des fonctions continues dans l'espace $L^1(\mathbb{T}, h(x) dx)$, on obtient la même identité pour tout ψ dans cet espace. En appliquant le résultat à $|h(x)| \mathbf{1}_{\{h(x) \neq 0\}} / h(x)$, qui est bornée par 1 en module et donc dans cet espace, on obtient que $\int_{\mathbb{T}} |h(x)| dx = 0$. Donc $h = 0$. \square

Démonstration de la proposition 2.10. Le fait que la série trigonométrique $\sum_{n \in \mathbb{Z}} c_n(f) e_n$ converge normalement vers une fonction continue g telle que $c_n(g) = c_n(f)$ pour tout $n \in \mathbb{Z}$ est une conséquence de la proposition 2.5. On en déduit que $f = g$ presque partout par le lemme 2.11. \square

2.3 Convergence des séries de Fourier dans L^2

La réponse la plus simple à la question précédente est que dans le cadre L^2 hilbertien, cette convergence a toujours lieu.

Une conséquence du théorème d'approximation de Weierstrass est que la famille $\{e_n, n \in \mathbb{Z}\}$ forme une base hilbertienne de $L^2(\mathbb{T})$, c'est-à-dire que tout élément de $L^2(\mathbb{T})$ se décompose comme série sur cette base. Rappelons la preuve de ce fait, qui est un résultat général sur les espaces de Hilbert.

Théorème 2.12. *Pour toute fonction $f \in L^2(\mathbb{T})$, la famille $(c_n(f), n \in \mathbb{Z})$ est de carré sommable, et de plus, l'application $f \mapsto (c_n(f), n \in \mathbb{Z})$ réalise une isométrie de $L^2(\mathbb{T})$ sur $\ell^2(\mathbb{Z})$, muni de la structure hilbertienne usuelle :*

$$\|f\|_{L^2(\mathbb{T})}^2 = \frac{1}{2\pi} \int_{-\pi}^{\pi} |f(x)|^2 dx = \sum_{n \in \mathbb{Z}} |c_n(f)|^2 = \|(c_n(f), n \in \mathbb{Z})\|_{\ell^2(\mathbb{Z})}^2.$$

Cette identité s'appelle *égalité de Parseval*.

Démonstration. Ce résultat est un théorème général sur les espaces de Hilbert munis d'une base hilbertienne, c'est-à-dire une famille orthonormale qui engendre un sous-espace dense. Nous la redonnons dans le cas particulier qui nous intéresse.

Pour tout $N \geq 0$, soit $\mathcal{T}_N = \text{Vect}(e_n, -N \leq n \leq N)$ l'espace des polynômes trigonométriques de degré au plus N . Par définition, la somme de Fourier $S_N f$ est la projection orthogonale de f sur \mathcal{T}_N . On a alors par le théorème de Pythagore

$$\|f\|_2^2 = \|f - S_N f\|_2^2 + \|S_N f\|_2^2 = \|f - S_N f\|_2^2 + \sum_{n=-N}^N |c_n(f)|^2.$$

Comme tous les termes sont positifs, on a que

$$\|f\|_2^2 \geq \sum_{n=-N}^N |c_n(f)|^2$$

pour tout N , et donc

$$\|f\|_2^2 \geq \sum_{n \in \mathbb{Z}} |c_n(f)|^2,$$

ce que l'on appelle l'inégalité de Bessel. On utilise alors le théorème d'approximation de Weierstrass : comme $\text{Vect}(e_n, n \in \mathbb{Z})$ est dense dans $\mathcal{C}(\mathbb{T}, \mathbb{C})$ pour la norme uniforme, la même chose est vraie pour la norme L^2 , et par densité des fonctions continues dans les fonction L^2 , on déduit que $\text{Vect}(e_n, n \in \mathbb{Z})$ est dense pour la norme L^2 dans $L^2(\mathbb{T})$. On en déduit que pour tout $\varepsilon > 0$, il existe un polynôme trigonométrique P tel que $\|f - P\|_2 < \varepsilon$. Mais si le degré de P est N_0 , on voit que

$$\|f - S_{N_0} f\|_2 \leq \|f - P\|_2 < \varepsilon,$$

par la propriété de la projection orthogonale : $S_{N_0}f$ est le point de \mathcal{T}_{N_0} le plus proche de f en norme L^2 . On en déduit immédiatement que $S_N f \rightarrow f$ dans $L^2(\mathbb{T})$. Finalement, on a bien que l'inégalité de Bessel est une égalité.

Il reste à montrer que l'application $f \mapsto (c_n(f), n \in \mathbb{Z})$ est surjective sur $\ell^2(\mathbb{Z})$. Mais si $(c_n, n \in \mathbb{Z})$ est de carré sommable, alors $\sum_{n \in \mathbb{Z}} c_n e_n$ converge dans $L^2(\mathbb{T})$ et définit un élément f , tel que $c_n(f) = (e_n, f) = c_n$, d'où le résultat. \square

Corollaire 2.13. *Pour tout $f \in L^2(\mathbb{T})$, on a que la somme $\sum_{n \in \mathbb{Z}} c_n(f) e_n$ converge dans $L^2(\mathbb{T})$ et est égale à f . C'est également la limite de $S_N f$ dans $L^2(\mathbb{T})$ lorsque $N \rightarrow \infty$.*

Il convient cependant de ne pas se laisser abuser par l'énoncé précédent. En effet, il ne stipule absolument pas que les deux fonctions

$$f(x) \quad \text{et} \quad \sum_{n \in \mathbb{Z}} c_n(f) \exp(inx)$$

sont égales en tout x , ni même en un seul x : en fait, la convergence de la série à droite en un point donné n'est pas garantie, puisque $c_n(f)$ est seulement supposée de carré sommable.

2.4 Convergence ponctuelle des séries de Fourier

La question de savoir si l'on a convergence en un point x de la série de Fourier d'une fonction est un problème en général très difficile. Nous allons donner quelques résultats très partiels en ce sens. De façon évidente à partir de nos résultats sur les séries trigonométriques, on a que $S_N f$ converge uniformément vers f dès lors que $\sum_{n \in \mathbb{Z}} |c_n(f)| < \infty$. Nous allons montrer que cela est impliqué par une condition de régularité de f .

2.4.1 Le cas \mathcal{C}^1 par morceaux

Une remarque importante est que la série de Fourier peut être représentée par un produit de convolution.

Définition 2.14. *Le noyau de Dirichlet d'ordre N est la fonction*

$$D_N(x) = \sum_{n=-N}^N e^{inx} = \frac{\sin((N+1/2)x)}{\sin(x/2)}, \quad x \in \mathbb{T}.$$

Pour vérifier la formule annoncée, il suffit de constater que la somme est géométrique, et vaut $(e^{i(N+1)x} - e^{-iNx}) / (e^{ix} - 1)$ et factoriser haut et bas par $e^{ix/2}$.

Lemme 2.15. *Soit $f \in L^1(\mathbb{T})$. La N -ème somme de Fourier de f est donnée par*

$$S_N f(x) = D_N * f(x), \quad x \in \mathbb{T}.$$

La preuve est immédiate : par définition

$$S_N f(x) = \sum_{n=-N}^N \frac{1}{2\pi} \int_{-\pi}^{\pi} f(y) e^{in(x-y)} dy = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(y) D_N(x-y) dy.$$

Il est assez tentant d'appliquer les résultats sur les approximations de l'unité du chapitre précédent. Malheureusement, la famille $(D_N, n \geq 0)$ n'est pas une approximation de l'unité, même si l'on a la propriété que

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} D_N(x) dx = 1,$$

ce qui est clair à partir de la définition de D_N comme somme de fonctions trigonométriques.

Théorème 2.16. *Soit $f: \mathbb{T} \rightarrow \mathbb{C}$ une fonction de classe $\mathcal{C}^1(\mathbb{T}, \mathbb{C})$ par morceaux. Pour tout $x \in \mathbb{T}$, on a la convergence*

$$S_N f(x) \xrightarrow{N \rightarrow \infty} \frac{f(x+) + f(x-)}{2} = \tilde{f}(x),$$

où $f(x-), f(x+)$ désignent les limites à gauche et à droite de f en x .

Démonstration. Fixons $x \in \mathbb{T}$. On écrit, en utilisant le fait que D_N est une fonction paire

$$S_N f(x) = D_N * f(x) = \frac{1}{2\pi} \int_0^{\pi} f(x-y) D_N(y) dy + \frac{1}{2\pi} \int_0^{\pi} f(x+y) D_N(y) dy.$$

Comme D_N est d'intégrale (normalisée) 1, on en déduit que

$$S_N f(x) - \tilde{f}(x) = \frac{1}{2\pi} \int_0^{\pi} (f(x-y) + f(x+y) - 2\tilde{f}(x)) D_N(y) dy.$$

Notons $g(y) = (f(x-y) + f(x+y) - 2\tilde{f}(x)) / 2\sin(y/2)$ pour y dans $]0, \pi]$, et $g(0) = f'(x+) - f'(x-)$. Alors la fonction g est continue sur $[0, \pi]$, et

$$S_N f(x) - \tilde{f}(x) = \frac{1}{\pi} \int_0^{\pi} g(y) \sin\left(\left(N + \frac{1}{2}\right)y\right) dy \xrightarrow{N \rightarrow \infty} 0$$

par le lemme de Riemann-Lebesgue. □

De même, on montrerait le résultat suivant par la même méthode.

Proposition 2.17. *Soit $f: \mathbb{T} \rightarrow \mathbb{C}$ une fonction Hölder-continue d'exposant $\alpha \in]0, 1]$, c'est-à-dire telle qu'il existe $C \in]0, \infty[$ avec*

$$|f(x) - f(y)| \leq C|x - y|^\alpha, \quad x, y \in \mathbb{T}.$$

Alors $S_N f(x)$ converge en tout point vers f .

On peut se demander si le résultat précédent peut se renforcer en une convergence uniforme. Clairement, si f n'est pas continue, il n'est pas possible de l'approcher uniformément par une suite de fonctions continues, donc par un polynôme trigonométrique. Nous allons

Proposition 2.18. *Soit $f: \mathbb{T} \rightarrow \mathbb{C}$ une fonction de classe $\mathcal{C}^1(\mathbb{T}, \mathbb{C})$, ou plus généralement continue, et de classe \mathcal{C}^1 par morceaux. Alors*

$$c_n(f') = in c_n(f).$$

De plus, on a que $\sum_{n \in \mathbb{Z}} |c_n(f)| < \infty$. En particulier, la série de Fourier converge normalement vers f , et les sommes de Fourier $S_N f$ convergent uniformément vers f .

Démonstration. L'identité sur les coefficients de Fourier est immédiate par intégration par parties

$$\int_{-\pi}^{\pi} f'(x) e^{-inx} dx = [f(x) e^{-inx}]_{-\pi}^{\pi} + in \int_{-\pi}^{\pi} f(x) e^{-inx} dx,$$

en constatant que le terme de crochet est nul par périodicité. Comme on a supposé que f' est une fonction continue sur \mathbb{T} , elle est en particulier dans L^2 et donc ses coefficients de Fourier forment une famille de carré sommable. Ainsi

$$\sum_{n \in \mathbb{Z}} |n c_n(f)|^2 < \infty.$$

Ensuite, on utilise l'inégalité de Cauchy-Schwarz en écrivant $c_n(f) = n c_n(f) / n$:

$$\sum_{n \in \mathbb{Z}} |c_n(f)| \leq \sqrt{\sum_{n \in \mathbb{Z}} |n c_n(f)|^2} \cdot \sum_{n \in \mathbb{Z}} \frac{1}{n^2} < \infty$$

et on en déduit que $(c_n(f), n \in \mathbb{Z})$ est sommable. De ce fait, les sommes de Fourier $S_N f$ convergent uniformément, et la limite est f par le théorème 2.16. \square

Notons que cette proposition n'utilise pas les résultats que nous avons énoncés sur le cas hilbertien, à l'exception de l'inégalité de Bessel stipulant que

$$\sum_{n \in \mathbb{Z}} |c_n(g)|^2 \leq \|g\|_2^2$$

pour toute fonction g de carré intégrable, ce qui est simplement une conséquence du théorème de Pythagore. Or, en constatant que $S_N f$ est un polynôme trigonométrique pour tout N , ceci montre par un argument différent la densité des polynômes trigonométriques dans $\mathcal{C}^1(\mathbb{T}, \mathbb{C})$ pour la norme uniforme, et donc dans $\mathcal{C}^0(\mathbb{T}, \mathbb{C})$ (théorème de Weierstrass).

2.4.2 Convergence de Cesaro

Enfin, une manière d'obtenir une convergence uniforme pour des fonctions continues est de remplacer la convergence des sommes de Fourier par leur moyenne de Cesaro

$$C_N f(x) = \frac{S_0 f(x) + \cdots + S_{N-1}(x)}{N}.$$

Théorème 2.19. *Soit $f \in \mathcal{C}^0(\mathbb{T}, \mathbb{C})$. Alors la suite $(C_N f, N \geq 1)$ converge uniformément vers f .*

Démonstration. On constate d'abord que

$$C_N f = K_N * f$$

où $K_N(x) = N^{-1} \sum_{k=0}^{N-1} D_k(x)$ est le *noyau de Féjer*. Ceci est une conséquence directe du fait que $S_N f = D_N * f$ et de la définition de C_N . On constate alors que

$$\begin{aligned} K_N(x) &= \frac{1}{N} \sum_{k=0}^{N-1} \sum_{n=-k}^k e^{inx} = \frac{1}{N} \sum_{k=0}^{N-1} \frac{e^{i(k+1)x} - e^{-ikx}}{e^{ix} - 1} \\ &= \frac{1}{N(e^{ix} - 1)} \left(\frac{e^{iNx} - 1}{1 - e^{-ix}} - \frac{1 - e^{-iNx}}{1 - e^{-ix}} \right) \\ &= \frac{2 - 2 \cos(Nx)}{N |e^{ix} - 1|^2} = \frac{1}{N} \frac{\sin^2(Nx/2)}{\sin^2(x/2)} \end{aligned}$$

À partir de la définition de K_N , on voit que $\int_{-\pi}^{\pi} K_N(y) dy = 2\pi$, et par la formule précédente, on a $K_N \geq 0$. Enfin, on a clairement que pour tout $\delta \in]0, \pi[$,

$$\int_{-\pi}^{\pi} K_N(y) \mathbf{1}_{\{|y| > \delta\}} dy \leq \frac{1}{N} \frac{2\pi}{\sin^2(\delta/2)} \xrightarrow{N \rightarrow \infty} 0.$$

Donc $(K_N, N \geq 1)$ est une approximation de l'unité, et le résultat découle de la proposition 1.9. \square

Remarquons que $C_N f$ est un polynôme trigonométrique pour tout N , et donc ce résultat nous donne une troisième preuve, encore différente des deux autres, de la densité des polynômes trigonométriques dans $\mathcal{C}^0(\mathbb{T}, \mathbb{C})$.

2.5 *Preuve du théorème de Stone-Weierstrass

On montre d'abord que la fonction $x \mapsto |x|$ peut être approximée uniformément sur $[-1, 1]$ par une suite de polynômes réels. Pour cela, on peut utiliser le théorème d'approximation de Weierstrass pour les polynômes. Ou bien, on peut utiliser le fait que $|x| = \sqrt{1 - (1 - x^2)} = \sum_{n \geq 0} \binom{1/2}{n} (-1 - x^2)^n$, du fait du développement en série entière de la fonction $z \mapsto \sqrt{1 - z}$, où la convergence des sommes partielles de la série a lieu uniformément sur $[-1, 1]$.

Supposons d'abord que l'algèbre \mathcal{A} soit formée de fonctions à valeurs réelles. Notre but est de montrer que l'adhérence de \mathcal{A} est égale à $\mathcal{C}^0(X, \mathbb{R})$, et sans perte de généralité on peut supposer que \mathcal{A} est fermée. Dans ce cas, pour toute fonction $f \in \mathcal{A}$, on a que $P(f) \in \mathcal{A}$ pour tout polynôme réel P , puisque \mathcal{A} contient les fonctions constantes. Comme f est continue sur X compact, elle est bornée, et $f / \|f\|_{\infty}$ prend ses valeurs dans $[-1, 1]$. Par le résultat d'approximation de la valeur absolue rappelé plus haut, et comme \mathcal{A} est fermée, on en déduit que $|f| / \|f\|_{\infty} \in \mathcal{A}$, et donc $|f| \in \mathcal{A}$.

On en déduit alors que pour tout $f, g \in \mathcal{A}$, les fonctions

$$f \wedge g = \frac{f + g - |f - g|}{2}, \quad f \vee g = \frac{f + g + |f - g|}{2}$$

sont aussi dans \mathcal{A} .

Fixons maintenant une fonction $f \in \mathcal{C}(X, \mathbb{R})$, et $x \in X$. Pour tout $y \in X$, comme \mathcal{A} sépare les points, on peut trouver une fonction $g_{x,y} \in \mathcal{A}$ telle que $g_{x,y}(x) = f(x)$ et $g_{x,y}(y) = f(y)$. Pour tout $\varepsilon > 0$, il existe alors un voisinage $V_{x,y}$ de y tel que $g_{x,y}(z) > f(z) - \varepsilon$ pour tout $z \in V_{x,y}$. Par compacité, on peut recouvrir X par un nombre fini de tels voisinages, disons $V_{x,y_1}, \dots, V_{x,y_k}$. Notons $g_x = \max(g_{x,y_1}, \dots, g_{x,y_k})$, de sorte qu'on a $g_x \in \mathcal{A}$ par ce que l'on a montré ci-dessus. La construction étant valide pour tout $x \in X$, on obtient une famille de fonctions $(g_x, x \in X)$ telles que $g_x(x) = f(x)$ et $g_x(z) > f(z) - \varepsilon$ pour tout $z \in X$.

Pour tout x , on peut alors trouver un voisinage V_x de x tel que $g_x(z) < f(z) + \varepsilon$ pour tout $z \in V_x$. Comme précédemment, on peut trouver un sous-recouvrement fini par V_{x_1}, \dots, V_{x_l} disons. Si l'on pose $g = \min(g_{x_1}, \dots, g_{x_l}) \in \mathcal{A}$, on obtient que pour tout $z \in X$, on a $f(z) - \varepsilon < g(z) < f(z) + \varepsilon$, c'est-à-dire que $\|f - g\|_\infty < \varepsilon$. D'où le résultat dans le cas où \mathcal{A} est constitué de fonctions réelles.

Dans le cas complexe, on utilise le fait que \mathcal{A} est stable par conjugaison pour obtenir que si $f \in \mathcal{A}$, alors $\Re f$ et $\Im f$ sont aussi dans \mathcal{A} . Donc \mathcal{A} contient une algèbre de fonctions réelles qui séparent les points, et qui contient les fonctions constantes, et donc son adhérence contient $\mathcal{C}(X, \mathbb{R})$. Donc l'adhérence de \mathcal{A} contient $\mathcal{C}(X, \mathbb{C})$ en approchant partie réelle et partie imaginaire de la fonction que l'on essaie d'approcher.

Chapitre 3

La transformée de Fourier dans \mathbb{R}^d

Dans tout ce chapitre, nous travaillerons avec la mesure de Lebesgue sur \mathbb{R}^d renormalisée par $(2\pi)^{d/2}$, que nous noterons $\lambda_d(dx) = dx / (2\pi)^{d/2}$, plutôt qu'avec la mesure de Lebesgue standard. La raison de ce choix apparaîtra un peu plus tard.

3.1 Transformée de Fourier d'une fonction intégrable

Définitions

Soit $f \in \mathcal{L}^1 = \mathcal{L}^1(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d), \lambda_d)$. Pour tout $\xi \in \mathbb{R}^d$, on note

$$\begin{aligned}\hat{f}(\xi) &= \frac{1}{(2\pi)^{d/2}} \int_{\mathbb{R}^d} \exp(-i\langle \xi, x \rangle) f(x) dx \\ &= \int_{\mathbb{R}^d} \exp(-i\langle \xi, x \rangle) f(x) \lambda_d(dx)\end{aligned}$$

Comme l'intégrande a pour module $|f(x)|$, qui est intégrable, cette intégrale est bien définie pour tout $\xi \in \mathbb{R}^d$. On dit que la fonction \hat{f} est la *transformée de Fourier de f* . Plus généralement, si $f \in L^1$, alors la formule ci-dessous détermine également une fonction \hat{f} indépendante du choix du représentant de f dans \mathcal{L}^1 . Par la suite, nous ne précisons pas toujours si l'on travaille avec une fonction mesurable ou avec une classe de fonctions égales presque partout.

Il est légitime de se demander pourquoi la normalisation ci-dessus a été choisie. Notons que formellement, on peut noter

$$\hat{f}(\xi) = \langle e_\xi, f \rangle_{L^2}$$

où $\langle \cdot, \cdot \rangle_{L^2}$ est le produit scalaire hermitien usuel sur $L^2 = L^2(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d), \lambda_d)$ défini par $\langle f, g \rangle_{L^2} = \int_{\mathbb{R}^d} \overline{f(x)} g(x) dx / (2\pi)^{d/2}$, et $e_\xi(x) = \exp(i\langle \xi, x \rangle)$. Bien sûr, e_ξ n'est pas un élément de L^2 , donc cette écriture est seulement formelle. La raison pour la renormalisation par $(2\pi)^{d/2}$ apparaîtra plus clairement plus loin, lorsque l'on verra que l'application $f \mapsto \hat{f}$ est une isométrie sur une partie dense de L^2 .

a. Propriétés élémentaires.

La transformée de Fourier est clairement \mathbb{C} -linéaire: si f, g sont intégrables et $a \in \mathbb{C}$, on a $\widehat{af + g} = a\hat{f} + \hat{g}$.

Si f est une fonction mesurable, et $y \in \mathbb{R}^d$, on note

$$\tau_y f(x) = f(x - y) \quad \text{et} \quad e_y f(x) = e^{i\langle y, x \rangle} f(x), \quad x \in \mathbb{R}^d$$

Soit $f \in \mathcal{L}^1$ et $y \in \mathbb{R}^d$. Alors

$$\widehat{\tau_y f} = e_{-y} \widehat{f}, \quad \text{et} \quad \widehat{e_y f} = \tau_y \widehat{f}. \quad (3.1)$$

La première formule s'obtient par un simple changement de variable, et la seconde est une conséquence immédiate de la définition.

Si maintenant M est une matrice de $\text{GL}_d(\mathbb{R})$, et si $g(x) = f(M^{-1}x)$, où f est toujours supposée intégrable, on a

$$\widehat{g}(\xi) = |\det M| \widehat{f}(M^* \xi) \quad (3.2)$$

où M^* est la matrice transposée de M . À nouveau, ceci s'obtient facilement par un changement de variables^{3.1} linéaire (poser $u = M^{-1}x$) dans l'intégrale

$$\widehat{g}(\xi) = \frac{1}{(2\pi)^{d/2}} \int_{\mathbb{R}^d} e^{-i\langle \xi, x \rangle} f(M^{-1}x) dx,$$

en notant que $\langle \xi, Mu \rangle = \langle M^* \xi, u \rangle$. Par exemple, pour M la matrice diagonale dont tous les coefficients diagonaux sont égaux à un réel non nul a , on obtient $g(x) = f(x/a)$ et

$$\widehat{g}(\xi) = |a|^d \widehat{f}(a\xi). \quad (3.3)$$

Pour $a = -1$, ceci donne, si l'on note $Rf(x) = f(-x)$, la formule

$$\widehat{Rf}(\xi) = \widehat{f}(-\xi) = R\widehat{f}(\xi).$$

Par ailleurs, notons que

$$\widehat{f}(-\xi) = \frac{1}{(2\pi)^{d/2}} \int_{\mathbb{R}^d} e^{-i\langle \xi, x \rangle} \overline{f(x)} dx = \overline{\widehat{f}(\xi)},$$

où \bar{z} est le complexe conjugué de z . On peut réécrire cela sous la forme concise

$$\widehat{\bar{f}} = R\widehat{f}. \quad (3.4)$$

Noter que si f est réelle, alors cela se simplifie en $\overline{\widehat{f}(\xi)} = \widehat{f}(-\xi)$, et si on suppose que f est réelle et paire, alors clairement $\widehat{f}(\xi) = \widehat{f}(-\xi)$ et on en déduit que \widehat{f} est une fonction paire à valeurs réelles.

b. Continuité, lemme de Riemann-Lebesgue

Proposition 3.1. *Soit f une fonction de \mathcal{L}^1 . Alors la fonction \widehat{f} est continue, et vérifie*

$$\|\widehat{f}\|_\infty \leq \|f\|_1.$$

En particulier, l'application linéaire $f \mapsto \widehat{f}$ de L^1 dans L^∞ est continue. Par ailleurs, on a

$$\lim_{|\xi| \rightarrow \infty} \widehat{f}(\xi) = 0.$$

^{3.1.} Nous verrons au chapitre suivant quelques compléments concernant le changement de variables, dont une justification de cette formule

Noter que la proposition implique en particulier que \hat{f} est uniformément continue (exercice). Le fait qu'une transformée de Fourier (d'une fonction intégrable) soit nulle à l'infini est appelé *Lemme de Riemann-Lebesgue*. Noter également le fait suivant: même si l'on suppose que f est un élément de L^1 , la même conclusion est vraie: rappelons qu'une transformée de Fourier de fonction L^1 est bien définie partout, et définit bien une vraie fonction et non une classe de fonctions. En particulier, parler de continuité de \hat{f} a bien un sens.

Démonstration. La continuité est une conséquence triviale de (3.1) et de la continuité sous le signe intégrale. \square

Le fait qu'une transformée de Fourier soit une fonction mesurable bornée implique que, si f, ϕ sont toutes deux dans L^1 , alors $f\hat{\phi}$ et $\hat{f}\phi$ sont dans L^1 . L'énoncé suivant est appelé « formule de réciprocity ».

Proposition 3.2. *Soit $f, \phi \in L^1$. Alors on a*

$$\int_{\mathbb{R}^d} f(x)\hat{\phi}(x)dx = \int_{\mathbb{R}^d} \hat{f}(\xi)\phi(\xi)d\xi.$$

Démonstration. La fonction $(x, \xi) \mapsto e^{i\langle \xi, x \rangle} f(x)\phi(\xi)$ est intégrable par rapport à $dx d\xi$, ce qui est exactement ce dont on a besoin pour appliquer le théorème de Fubini permettant l'interversion suivante:

$$\int_{\mathbb{R}^d} f(x)dx \int_{\mathbb{R}^d} e^{-i\langle x, \xi \rangle} \phi(\xi) \frac{d\xi}{(2\pi)^{d/2}} = \int_{\mathbb{R}^d} \phi(\xi) d\xi \int_{\mathbb{R}^d} e^{-i\langle \xi, x \rangle} f(x) \frac{dx}{(2\pi)^{d/2}},$$

ce qui est exactement ce qu'on voulait. \square

c. Régularité

Proposition 3.3. *Soit $f \in L^1$ une fonction telle que $x \mapsto |x|f(x)$ est intégrable. Alors la fonction \hat{f} est de classe $\mathcal{C}^1(\mathbb{R}^d, \mathbb{C})$ et de plus, pour tout $j \in \{1, 2, \dots, d\}$, on a*

$$\frac{\partial \hat{f}}{\partial \xi_j}(\xi) = - \int_{\mathbb{R}^d} e^{-i\langle \xi, x \rangle} i x_j f(x) \frac{dx}{(2\pi)^{d/2}},$$

qui est la transformée de Fourier de $x \mapsto i x_j f(x)$.

Plus généralement, si $x \mapsto |x|^k f(x)$ est intégrable pour un entier $k \geq 1$, alors \hat{f} est de classe $\mathcal{C}^k(\mathbb{R}^d, \mathbb{C})$, et pour tout $\alpha = (\alpha_1, \dots, \alpha_d) \in \mathbb{N}^d$ tel que $|\alpha| = \sum_{j=1}^d \alpha_j \leq k$, on a

$$\frac{\partial^{|\alpha|} \hat{f}}{\partial \xi_1^{\alpha_1} \dots \partial \xi_d^{\alpha_d}}(\xi) = \int_{\mathbb{R}^d} e^{-i\langle \xi, x \rangle} (-ix)^\alpha f(x) \frac{dx}{(2\pi)^{d/2}},$$

où l'on note par convention $x^\alpha = x_1^{\alpha_1} \dots x_d^{\alpha_d}$.

Pour simplifier, on notera par la suite $\partial_x^\alpha h = \partial^{|\alpha|} h / \partial x_1^{\alpha_1} \dots \partial x_d^{\alpha_d}$ les dérivées partielles d'une fonction $x \mapsto h(x)$. Par exemple, si h est une fonction de deux variables x et y , on pose $\partial_x h = \partial h / \partial x$, ou $\partial_{xy} h = \partial^2 h / \partial x \partial y$.

Démonstration. C'est une conséquence immédiate du théorème de dérivation sous le signe intégrale, en remarquant que $|e^{i\langle \xi, x \rangle} (ix)^\alpha f(x)| \leq |x|^k |f(x)|$. \square

On voit donc qu'une propriété de décroissance à l'infini de f implique une propriété de régularité de \widehat{f} . Nous allons montrer une sorte de propriété duale de celle-ci : la régularité de f implique une propriété de décroissance à l'infini de \widehat{f} .

Proposition 3.4. *Soit $f \in \mathcal{L}^1 \cap \mathcal{C}^1(\mathbb{R}^d, \mathbb{C})$ une fonction de classe \mathcal{C}^1 intégrable, telle que $\partial f / \partial x_j$ est intégrable pour tout $j \in \{1, 2, \dots, d\}$. Alors on a*

$$\frac{\widehat{\partial f}}{\partial x_j}(\xi) = i\xi_j \widehat{f}(\xi).$$

Plus généralement, si l'on suppose qu'il existe un entier $k \geq 1$ tel que $f \in \mathcal{C}^k(\mathbb{R}^d, \mathbb{C})$ et $\partial_{x^\alpha} f \in \mathcal{L}^1$ pour tout multi-indice α vérifiant $|\alpha| \leq k$, alors on a, pour ces mêmes multi-indices,

$$\widehat{\partial_{x^\alpha} f}(\xi) = (i\xi)^\alpha \widehat{f}(\xi).$$

Remarque. En réalité l'hypothèse que f est de classe $\mathcal{C}^1(\mathbb{R}^d, \mathbb{C})$ est superflue, il suffit de supposer que les dérivées partielles considérées existent en *tout* point (mais pas seulement en presque tout point !). En utilisant le Théorème 7.21 du livre de Rudin, stipulant que si f est dérivable en tout point d'un intervalle $[a, b]$ et a sa dérivée f' dans \mathcal{L}^1 , alors on a $\int_x^y f'(z) dz = f(y) - f(x)$, la preuve ci-dessous s'adapte *verbatim*.

Démonstration. Supposons sans perte de généralité que $j = 1$. Pour un $(x_2, x_3, \dots, x_d) \in \mathbb{R}^{d-1}$ fixé, on a

$$f(y, x_2, x_3, \dots, x_d) - f(x, x_2, x_3, \dots, x_d) = \int_x^y \partial_{x_1} f(z, x_2, x_3, \dots, x_d) dz.$$

Comme f et $\partial_{x_1} f$ sont supposées intégrables, le théorème de Fubini montre que pour λ_{d-1} -presque tout (x_2, \dots, x_d) , les fonctions $x \mapsto f(x, x_2, \dots, x_d)$ et $x \mapsto \partial_{x_1} f(x, x_2, \dots, x_d)$ sont intégrables. On déduit de la formule précédente que pour λ_{d-1} -presque tout (x_2, \dots, x_d) , la fonction $x \mapsto f(x, x_2, \dots, x_d)$ admet une limite en $\pm\infty$. Comme f est intégrable, cette limite est nulle pour λ_{d-1} -presque tout (x_2, \dots, x_d) , par une nouvelle application du théorème de Fubini. On écrit alors, en notant $x' = (x_2, \dots, x_d)$ et $\xi' = (\xi_2, \dots, \xi_d)$,

$$\begin{aligned} (2\pi)^{d/2} \widehat{\partial_{x_1} f}(\xi) &= \int_{\mathbb{R}^{d-1}} e^{-i\langle \xi', x' \rangle} dx' \lim_{K \rightarrow \infty} \int_{-K}^K e^{-i\xi_1 x_1} \partial_{x_1} f(x_1, x') dx_1 \\ &= \int_{\mathbb{R}^{d-1}} e^{-i\langle \xi', x' \rangle} dx' \\ &\quad \times \lim_{K \rightarrow \infty} \left([e^{-i\xi_1 x_1} f(x_1, x')]_{x_1=-K}^{x_1=K} + i\xi_1 \int_{-K}^K e^{-i\xi_1 x_1} f(x_1, x') dx_1 \right) \\ &= i\xi_1 \int_{\mathbb{R}^{d-1}} e^{-i\langle \xi', x' \rangle} dx' \int_{-\infty}^{\infty} e^{-i\xi_1 x_1} f(x_1, x') dx_1 = (2\pi)^{d/2} i\xi_1 \widehat{f}(\xi). \end{aligned}$$

Ici, on a utilisé le théorème de Fubini à la première et dernière étapes, et les observations précédentes conjointement à une intégration par parties pour les autres étapes.

Le résultat plus général s'obtient par une récurrence aisée, qui est laissée au lecteur. \square

Le lemme de Riemann-Lebesgue appliqué à $\partial_{x^\alpha} f$, conjointement au précédent résultat, donne le résultat suivant.

Corollaire 3.5. *Si l'on suppose qu'il existe un entier $k \geq 1$ tel que $f \in \mathcal{C}^k(\mathbb{R}^d, \mathbb{C})$ et $\partial_{x^\alpha} f \in \mathcal{L}^1$ pour tout multi-indice α vérifiant $|\alpha| \leq k$, alors on a*

$$\hat{f}(\xi) = o\left(\frac{1}{|\xi|^k}\right).$$

d. Lien avec la convolution

La transformée de Fourier est un morphisme multiplicatif pour la convolution dans L^1 . Nous noterons, lorsque cela a un sens,

$$\begin{aligned} f * g(x) &= \frac{1}{(2\pi)^{d/2}} \int_{\mathbb{R}^d} f(x-y)g(y)dy, \\ &= \int_{\mathbb{R}^d} f(x-y)g(y)\lambda_d(dy) \end{aligned}$$

avec la normalisation par $(2\pi)^{d/2}$, contrairement à la convention adoptée au premier chapitre.

Proposition 3.6. *Soit $f, g \in L^1$, alors on a $\widehat{f * g}(\xi) = \hat{f}(\xi)\hat{g}(\xi)$ pour tout $\xi \in \mathbb{R}^d$.*

Démonstration. On écrit simplement

$$\begin{aligned} \widehat{f * g}(\xi) &= \int_{\mathbb{R}^d} e^{-i\langle \xi, x \rangle} \lambda_d(dx) \int_{\mathbb{R}^d} f(x-y)g(y)\lambda_d(dy) \\ &= \int_{\mathbb{R}^d} g(y) \lambda_d(dy) \tau_y \hat{f}(\xi) = \hat{f}(\xi) \int_{\mathbb{R}^d} e^{-i\langle \xi, y \rangle} g(y) \lambda_d(dy), \end{aligned}$$

où l'on a appliqué le théorème de Fubini dans la deuxième inégalité, $(x, y) \mapsto e^{-i\langle \xi, x \rangle} f(x-y)g(y)$ étant clairement intégrable par rapport à $dx dy$. \square

Cette propriété élémentaire est l'une des plus importantes de la transformée de Fourier. On verra en particulier le rôle qu'elle joue lorsqu'on somme des variables aléatoires indépendantes.

3.2 L'exemple de la densité gaussienne.

Pour $\sigma > 0$, on note

$$g_\sigma(x) = \frac{1}{\sigma^d} \exp\left(-\frac{|x|^2}{2\sigma^2}\right), \quad x \in \mathbb{R}^d,$$

que l'on appelle densité gaussienne isotrope dans \mathbb{R}^d . Notons que l'on a la relation élémentaire de changement d'échelle suivante : pour tout $\sigma > 0$,

$$g_\sigma(x) = \frac{1}{\sigma^d} g_1\left(\frac{x}{\sigma}\right), \quad x \in \mathbb{R}^d. \quad (3.5)$$

Proposition 3.7. *Pour tout $\sigma > 0$, on a que $\int_{\mathbb{R}^d} g_\sigma(x) \lambda_d(dx) = 1$.*

Démonstration. Du fait de la relation (3.5), il suffit de montrer ce résultat pour $\sigma = 1$. Par ailleurs, comme $g_1(x) = \prod_{i=1}^d e^{x_i^2/2}$, le théorème de Tonelli montre qu'il suffit de traiter le cas où $d = 1$.

On utilise alors le théorème de Tonelli et un changement de variables en coordonnées polaires pour obtenir :

$$\left(\int_{\mathbb{R}} e^{-\frac{x^2}{2}} dx \right)^2 = \int_{\mathbb{R}^2} e^{-\frac{x^2+y^2}{2}} dx dy = \int_{\mathbb{R}_+} r dr e^{-\frac{r^2}{2}} \int_{-\pi}^{\pi} d\theta = 2\pi \left[-e^{-\frac{r^2}{2}} \right]_{r=0}^{r=\infty} = 2\pi,$$

d'où le résultat. \square

Comme g_1 est une fonction positive, noter que pour la relation (3.5) implique que pour toute suite $(\sigma_n, n \geq 0)$ strictement positive de limite nulle, la suite $(g_{\sigma_n}, n \geq 0)$ est une approximation de l'unité.

Il est évident que la fonction g_σ est une fonction de classe $\mathcal{C}^\infty(\mathbb{R}^d, \mathbb{C})$, et de surcroît que $|x|^k \partial_{x^\alpha} g_\sigma(x)$ est de limite nulle lorsque $|x| \rightarrow \infty$ pour tout $k \geq 0$ et tout multi-indice α . On dit que g_σ est un élément de la *classe de Schwartz* \mathcal{S} . En particulier, g_σ et toutes ses dérivées partielles sont dans \mathcal{L}^1 , et on peut bien parler de leur transformée de Fourier.

Proposition 3.8. *On a, pour tout $\xi \in \mathbb{R}^d$,*

$$\hat{g}_\sigma(\xi) = \exp\left(-\frac{\sigma^2 |\xi|^2}{2}\right) = \frac{1}{\sigma^d} g_{1/\sigma}(\xi).$$

Démonstration. Du fait de la relation (3.5), les propriétés usuelles de la transformée de Fourier donnent que $\hat{g}_\sigma(\xi) = \hat{g}_1(\sigma\xi)$, et il suffit donc de traiter le cas où $\sigma = 1$. Par ailleurs, de façon similaire à la preuve de la proposition 3.7, il suffit, par une application du théorème de Fubini, de montrer le résultat pour $d = 1$. On suppose donc maintenant que $\sigma = 1, d = 1$, et on pose $g = g_1$. Les remarques précédant l'énoncé de la proposition, jointes aux propositions 3.3 et 3.4, montrent que

$$\hat{g}'(\xi) = -i \int_{\mathbb{R}^d} e^{-i\xi x} x e^{-\frac{x^2}{2}} \frac{dx}{\sqrt{2\pi}} = i \int_{\mathbb{R}^d} e^{-i\xi x} g'(x) \frac{dx}{\sqrt{2\pi}} = i \hat{g}'(\xi) = -\xi \hat{g}(\xi).$$

La fonction \hat{g} est donc solution de l'équation différentielle $h'(\xi) = -\xi h(\xi)$, et donc on a

$$\hat{g}(\xi) = \hat{g}(0) e^{-\xi^2/2}, \quad \xi \in \mathbb{R},$$

et comme $\hat{g}(0) = \int_{\mathbb{R}^d} g(x) dx / \sqrt{2\pi} = 1$ par la proposition 3.7, on conclut. \square

3.3 La formule d'inversion

Le résultat principal de ce chapitre dit que, lorsque l'on peut prendre la transformée de Fourier de \hat{f} , cette transformée égale Rf presque partout.

Théorème 3.9. Soit $f \in L^1$ une fonction telle que $\hat{f} \in L^1$. Alors si l'on pose

$$g(x) = \frac{1}{(2\pi)^{d/2}} \int_{\mathbb{R}^d} e^{i\langle x, \xi \rangle} \hat{f}(\xi) d\xi,$$

on a que $f(x) = g(x)$ pour λ_d -presque tout x , c'est-à-dire que $f = g$ dans L^1 . De façon concise, si $Rf(x) = f(-x)$,

$$R\hat{f} = f$$

Remarque. Noter que ce théorème admet la conséquence suivante, du fait de la proposition 3.1 : si $f \in L^1$ est telle que $\hat{f} \in L^1$, alors f (et \hat{f}) est égale presque partout à une fonction continue de limite nulle à l'infini. Cela restreint donc sensiblement l'ensemble des fonctions auxquelles le théorème précédent est susceptible de s'appliquer !

Avant de donner la preuve, notons qu'une approche naïve consisterait à appliquer la formule de réciprocity (Proposition 3.2) dans la définition de $g(x)$, et à écrire

$$g(x) = \frac{1}{(2\pi)^{d/2}} \int_{\mathbb{R}^d} \widehat{e^{i\langle x, \cdot \rangle}}(y) f(y) dy.$$

Le problème est que cette expression n'a pas de sens bien défini, car $e_x = e^{i\langle x, \cdot \rangle}$ n'est pas un élément de L^1 . Néanmoins, on peut se convaincre que la seule valeur « naturelle » à donner à $\widehat{e_x}(y)$ est 0 si $y \neq x$ et $+\infty$ si $y = x$, ce qui semble indiquer que la « fonction » $\widehat{e_x}$ est la masse de Dirac en x . On peut donner un sens à cela dans le cadre de la *théorie des distributions* qui sera étudiée en M1 : en fait, le théorème 3.9 permet de définir $\widehat{e_x} = \delta_x$.

Démonstration du théorème 3.9. Comme l'approche naïve décrite ci-dessus ne peut pas fonctionner, l'idée est de « lisser » les fonctions considérées par convolution avec une gaussienne. On remplace donc f par $f_\sigma = g_\sigma * f$ avec les notations du paragraphe 3.2, et on rappelle que $\hat{f}_\sigma = \hat{g}_\sigma \hat{f}$ par la proposition 3.6. En utilisant la proposition 3.8 donnant \hat{g}_σ , et la formule de réciprocity, on calcule alors

$$\begin{aligned} \int_{\mathbb{R}^d} e^{i\langle x, \xi \rangle} \hat{f}_\sigma(\xi) \lambda_d(d\xi) &= \frac{1}{\sigma^d} \int_{\mathbb{R}^d} e^{i\langle x, \xi \rangle} g_{1/\sigma}(\xi) \hat{f}(\xi) \lambda_d(d\xi) \\ &= \frac{1}{\sigma^d} \int_{\mathbb{R}^d} \widehat{e_x g_{1/\sigma}}(y) f(y) \lambda_d(dy) \\ &= \int_{\mathbb{R}^d} g_\sigma(y - x) f(y) \lambda_d(dy) \\ &= f_\sigma, \end{aligned} \tag{3.6}$$

où l'on a utilisé à nouveau la proposition 3.8 et les propriétés élémentaires de la transformée de Fourier à l'avant-dernière étape, et la parité de la fonction g_σ à la dernière étape. Lorsque $\sigma \rightarrow 0$, on a que $f_\sigma \rightarrow f$ dans L^1 par la proposition 1.10 et les remarques effectuées après la proposition 3.7. Par ailleurs,

$$\frac{1}{\sigma^d} e^{i\langle x, \xi \rangle} g_{1/\sigma}(\xi) \hat{f}(\xi) = e^{i\langle x, \xi \rangle} e^{-\sigma^2 |\xi|^2 / 2} \hat{f}(\xi) \xrightarrow{\sigma \rightarrow 0} e^{i\langle x, \xi \rangle} \hat{f}(\xi),$$

la convergence étant dominée par $|\hat{f}(\xi)|$, qui est dans L^1 par hypothèse. On en conclut que le membre de gauche de (3.9) converge ponctuellement vers $g(x)$ lorsque $\sigma \rightarrow 0$. On en conclut bien que $g = f$ presque partout, et c'est ce qu'on voulait démontrer. \square

Corollaire 3.10. *La transformée de Fourier est injective : si $f, g \in L^1$ sont telles que $\hat{f} = \hat{g}$, alors $f = g$.*

Remarque. Attention au fait que cette égalité est valide dans L^1 ! Si f, g sont de vraies fonctions dans \mathcal{L}^1 , alors l'énoncé dit seulement que $\hat{f} = \hat{g}$ implique que $f = g$, λ_d -presque partout.

Démonstration. Si $\hat{f} = \hat{g}$, alors $\widehat{f - g} = 0$ par linéarité, et comme 0 est évidemment intégrable, on a que $f - g = R\hat{0} = 0$. \square

3.4 La transformée de Fourier L^2

La formule d'inversion de la transformée de Fourier L^1 est un analogue direct de la proposition 2.10 pour les séries de Fourier. On peut se demander s'il existe également un analogue de la théorie hilbertienne de ces séries.

Rappelons que la théorie L^2 des séries de Fourier stipule que l'application $f \mapsto (c_n(f), n \in \mathbb{Z})$ est une isométrie, en fait un isomorphisme d'espaces de Hilbert, de $L^2(\mathbb{T}, \mathcal{B}(\mathbb{T}), \lambda)$ sur $\ell^2(\mathbb{Z})$ par la formule de Bessel-Parseval :

$$\|f\|_2 = \sum_{n \in \mathbb{Z}} |c_n(f)|^2.$$

Une différence notable entre séries et transformée de Fourier vient du fait que $L^2(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d), \lambda_d)$ n'est pas inclus dans $L^1(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d), \lambda_d)$, du fait que la mesure λ_d est infinie, et que la transformée de Fourier d'une fonction $f \in L^2(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d), \lambda_d)$ n'est pas un objet bien défini *a priori*. Néanmoins, on a bien la propriété d'isométrie suivante.

Proposition 3.11. *Soit $f \in \mathcal{C}_c^\infty(\mathbb{R}^d, \mathbb{C})$ une application infiniment dérivable, et à support compact. Alors \hat{f} appartient à \mathcal{L}^2 , et de plus l'application $f \mapsto \hat{f}$ de $\mathcal{C}_c^\infty(\mathbb{R}^d, \mathbb{C})$ dans L^2 , est une isométrie si l'on munit ces deux espaces de la norme $\|\cdot\|_2$.*

Démonstration. Comme f et ses dérivées sont continues à support compact, elles sont dans \mathcal{L}^1 , donc la proposition 3.1 et le corollaire 3.5 impliquent que \hat{f} est dans tous les ensembles \mathcal{L}^p pour $p \in [1, \infty]$. Pour montrer la propriété d'isométrie, on écrit, pour $f \in \mathcal{C}_c^\infty(\mathbb{R}^d, \mathbb{C})$, (et avec $Rf(x) = f(-x)$),

$$\int_{\mathbb{R}^d} \hat{f}(\xi) \overline{\hat{f}(\xi)} d\xi = \int_{\mathbb{R}^d} \hat{f}(\xi) R\hat{f}(\xi) d\xi = \int_{\mathbb{R}^d} f(x) R\hat{f}(x) dx = \int_{\mathbb{R}^d} f(x) \overline{f(x)} dx$$

où l'on a utilisé (3.4), la formule de réciprocity, puis la formule d'inversion, qui s'applique puisque $\hat{f} \in \mathcal{L}^1$. \square

Comme $\mathcal{C}_c^\infty(\mathbb{R}^d, \mathbb{C})$ est dense dans L^2 , et qu'une isométrie est uniformément continue, on en déduit qu'il existe un unique prolongement continu de $f \mapsto \hat{f}$ de L^2 dans L^2 , qui demeure une isométrie linéaire. Pour le distinguer de la transformée de Fourier classique, on le note $\mathcal{F}: L^2 \rightarrow L^2$, que l'on appelle parfois la transformée de Fourier-Plancherel.

Théorème 3.12. *L'application \mathcal{F} prolonge la transformée de Fourier L^1 , au sens où, si $f \in L^1 \cap L^2$, alors $\mathcal{F}f = \hat{f}$ presque partout. De plus, \mathcal{F} est une isométrie de L^2 sur lui-même.*

Démonstration. Soit $f \in L^1 \cap L^2$, et $f_\sigma = g_\sigma * f$ où g_σ est la densité gaussienne du paragraphe 3.2. On sait par la proposition 1.10 que f_σ converge vers f dans L^1 et dans L^2 , et de plus, la fonction f_σ est de classe $\mathcal{C}^\infty(\mathbb{R}^d, \mathbb{C})$ par une application aisée du théorème de dérivation sous le signe somme, en utilisant que g_σ et toutes ses dérivées partielles sont bornées.

Pour qu'on puisse lui appliquer la proposition précédente, on tronque f_σ en introduisant une fonction ψ de $\mathcal{C}_c^\infty(\mathbb{R}^d, \mathbb{C})$, à valeurs dans $[0, 1]$, telle que $\psi(x) = 1$ si $|x| \leq 1$ et $\psi(x) = 0$ si $|x| \geq 2$. On laisse au lecteur le soin de construire une telle fonction explicitement. Pour tout entier $n \geq 1$, posons alors $h_n(x) = \psi(x/n)f_{1/n}(x)$, de sorte que $h_n \in \mathcal{C}_c^\infty(\mathbb{R}^d)$ pour tout n . On a alors, pour $p \in [1, \infty[$,

$$\begin{aligned} \|h_n - f\|_p &\leq \|\psi(\cdot/n)(f_{1/n} - f)\|_p + \|(1 - \psi(\cdot/n))f\|_p \\ &\leq \|f_{1/n} - f\|_p + \left(\int_{\mathbb{R}^d} |f(x)|^p \mathbf{1}_{\{|x| \geq n\}} \lambda_d(dx) \right)^{1/p} \end{aligned}$$

ce qui converge vers 0 lorsque $n \rightarrow \infty$. Ceci est valable en particulier pour $p = 1$ et $p = 2$, et de la continuité de $\mathcal{F}: L^2 \rightarrow L^2$ et de $\hat{\cdot}: L^1 \rightarrow L^\infty$, on en déduit que $\mathcal{F}h_n = \hat{h}_n$ converge respectivement dans L^2 et dans L^∞ vers $\mathcal{F}f$ et \hat{f} . En particulier, ces deux fonctions sont égales presque partout.

Il ne reste plus qu'à démontrer la surjectivité de \mathcal{F} . Or on peut construire une seconde application $\tilde{\mathcal{F}}$ prolongeant la transformée de Fourier inverse $f \mapsto R\hat{f}$ de $\mathcal{C}_c^\infty(\mathbb{R}^d, \mathbb{C})$ dans L^2 , qui est une isométrie linéaire pour les mêmes raisons que précédemment. On a alors $\mathcal{F}\tilde{\mathcal{F}}f = f$ pour toute fonction $f \in \mathcal{C}_c^\infty(\mathbb{R}^d, \mathbb{C})$, et par densité et continuité, on en conclut que la même chose est vraie pour tout $f \in L^2$. Donc \mathcal{F} est inversible à droite, et en particulier, elle est surjective. \square

Remarque. Attention, si $f \in L^2$ on n'a pas en général la formule

$$\mathcal{F}f(\xi) = \frac{1}{(2\pi)^{d/2}} \int_{\mathbb{R}^d} e^{-i\langle \xi, x \rangle} f(x) dx, \quad (3.7)$$

car cette formule n'a pas toujours de sens ! En revanche, on peut noter que pour tout $A > 0$, la formule

$$\mathcal{F}_A f(\xi) = \frac{1}{(2\pi)^{d/2}} \int_{[-A, A]^d} e^{-i\langle \xi, x \rangle} f(x) dx, \quad \xi \in \mathbb{R}^d$$

a bien un sens, puisqu'une fonction dans L^2 est localement intégrable par l'inégalité de Cauchy-Schwarz. Comme $f \mathbf{1}_{[-A, A]^d}$ converge dans L^2 vers f lorsque $A \rightarrow \infty$, on en déduit que $\mathcal{F}_A f = \mathcal{F}(f \mathbf{1}_A)$ converge dans L^2 vers $\mathcal{F}f$, ce qui est une façon de donner un sens à l'intégrale impropre (3.7).

3.5 Transformée de Fourier d'une mesure signée

Comme on l'a mentionné brièvement un peu plus haut, la transformée de Fourier s'étend naturellement au-delà des fonctions de L^1 , et peut être définie même pour des objets qui ne sont pas des fonctions (les distributions). Même si nous n'allons pas traiter de cela ici, il nous sera très utile en théorie des probabilités de manipuler la transformée de Fourier de mesures de probabilités. Il est cependant utile de traiter le cas plus général des mesures signées. Si μ est une telle mesure, on note $|\mu|$ la mesure de variation totale associée, et on rappelle qu'il s'agit d'une mesure positive finie, dont la masse totale est notée $|\mu|(\mathbb{R}^d) = \|\mu\|$, et appelée norme de variation totale de μ .

Soit donc μ une mesure signée sur \mathbb{R}^d . On définit la transformée de Fourier de μ par la formule

$$\hat{\mu}(\xi) = \frac{1}{(2\pi)^{d/2}} \int_{\mathbb{R}^d} e^{-i\langle \xi, x \rangle} \mu(dx), \quad \xi \in \mathbb{R}^d.$$

Un cas particulier important sera celui des mesures à densité (sous-entendu par rapport à λ_d), c'est-à-dire des mesures μ qui s'écrivent sous la forme

$$\mu(dx) = f(x)dx / (2\pi)^{d/2}$$

pour une fonction $f \in L^1$ à valeurs réelles^{3.2}. Il est immédiat par définition que dans ce cas on a $\hat{\mu}(\xi) = \hat{f}(\xi)$. Ceci permet d'étendre strictement le cadre des fonctions L^1 , au moins à valeurs réelles. De fait, un certain nombre de propriétés que nous avons étudiées dans le cadre L^1 restent vraies ici.

Proposition 3.13. *La transformée de Fourier d'une mesure de signée μ sur \mathbb{R}^d est une fonction continue, et bornée par $\|\mu\|$. De plus, si l'on a $\int_{\mathbb{R}^d} |x|^k |\mu|(dx) < \infty$, alors $\hat{\mu}$ est de classe $C^k(\mathbb{R}^d, \mathbb{C})$, et l'on a pour tout multi-indice α tel que $|\alpha| \leq k$,*

$$\partial_{x^\alpha} \hat{\mu}(\xi) = \int_{\mathbb{R}^d} (-ix)^\alpha e^{-i\langle \xi, x \rangle} \mu(dx).$$

Enfin, si μ et ν sont deux mesures de probabilités, on a la formule de réciprocity

$$\int_{\mathbb{R}^d} \hat{\mu}(\xi) \nu(d\xi) = \int_{\mathbb{R}^d} \hat{\nu}(x) \mu(dx).$$

La preuve est exactement la même que pour les fonctions L^1 , et est laissée en exercice.

Remarque. Attention, le lemme de Riemann-Lebesgue **n'est plus vrai** dans ce contexte! Par exemple, la mesure de Dirac δ_0 vérifie $\hat{\delta}_0 = 1$, et n'est donc pas de limite nulle en l'infini.

3.2. En fait, on pourrait aussi définir la transformée de Fourier d'une mesure complexe, c'est-à-dire d'une application s'écrivant sous la forme $\mu = \mu_1 + i\mu_2$, avec μ_1 et μ_2 des mesures signées. Ceci permettrait d'étendre strictement le cadre L^1 étudié ici, et les énoncés de ce paragraphe restent tous vrais dans cette situation. On renvoie au chapitre 6 du livre de Rudin, *Real and complex analysis* pour les rudiments sur les mesures complexes.

Une autre propriété importante qui est conservée est celle de morphisme multiplicatif par rapport à la convolution. Si μ est une mesure signée, rappelons qu'elle peut s'écrire de façon unique sous la forme $\mu = \mu_+ - \mu_-$ où μ_+ et μ_- sont deux mesures positives finies de supports disjoints (décomposition de Jordan), auquel cas on a $|\mu| = \mu_+ + \mu_-$. La mesure produit de deux mesures signées μ et ν est alors définie comme

$$\mu \otimes \nu = \mu_+ \otimes \nu_+ - \mu_+ \otimes \nu_- - \mu_- \otimes \nu_+ + \mu_- \otimes \nu_-,$$

et définit une nouvelle mesure signée, de variation totale $|\mu| \otimes |\nu|$. Dans ce cadre, le théorème de Fubini s'énonce ainsi :

Théorème 3.14. *Soit μ et ν deux mesures signées, et soit f une fonction mesurable intégrable par rapport à $|\mu| \otimes |\nu|$, alors on a*

$$\int \mu(dx) \int f(x, y) \nu(dy) = \int \nu(dy) \int f(x, y) \mu(dx) = \int f(x, y) \mu \otimes \nu(dx dy).$$

Définition 3.15. *La convolution de deux mesures signées μ et ν est la mesure image par l'application $(x, y) \mapsto x + y$ de la mesure produit $\mu \otimes \nu$, et on la note $\mu * \nu$. La mesure $\mu * \nu$ est caractérisée par le fait que*

$$\int_{\mathbb{R}^d} f(z) \mu * \nu(dz) = \int_{\mathbb{R}^d \times \mathbb{R}^d} f(x + y) \mu(dx) \nu(dy)$$

pour toute fonction f mesurable bornée.

On laisse au lecteur le soin de montrer que l'opération $*$ est associative et commutative sur l'ensemble des mesures signées, et admet δ_0 pour élément neutre. Si μ est une mesure à densité, $\mu(dx) = f(x) \lambda_d(dx)$, notons que pour toute fonction h mesurable bornée, on a par le théorème de Tonelli

$$\begin{aligned} \int_{\mathbb{R}^d} h(z) \mu * \nu(dz) &= \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} h(x + y) f(x) \lambda_d(dx) \nu(dy) \\ &= \int_{\mathbb{R}^d} h(z) \lambda_d(dz) \int_{\mathbb{R}^d} f(z - y) \nu(dy), \end{aligned}$$

ce qui signifie que $\mu * \nu$ admet une densité par rapport à λ_d , cette densité étant

$$f * \nu(x) = \int_{\mathbb{R}^d} f(x - y) \nu(dy).$$

On notera que si à son tour ν est à densité, disons $\nu(dx) = g(x) \lambda_d(dx)$, alors $f * \nu = f * g$.

Proposition 3.16. *Si μ et ν sont deux mesures signées, on a $\widehat{\mu * \nu} = \widehat{\mu} \widehat{\nu}$.*

La preuve est laissée en exercice. On a également une sorte de généralisation de la formule d'inversion.

Théorème 3.17. *Soit μ une mesure signée telle que $\widehat{\mu} \in L^1$. Alors μ admet une densité par rapport à λ_d , qui est égale presque partout à la fonction*

$$\frac{1}{(2\pi)^{d/2}} \int_{\mathbb{R}^d} e^{i(x, \xi)} \widehat{\mu}(\xi) d\xi, \quad x \in \mathbb{R}^d.$$

Démonstration. La preuve est similaire à celle du théorème 3.9. On remplace à nouveau la mesure μ par la fonction $\mu_\sigma = g_\sigma * \mu$, où g_σ est la densité gaussienne du paragraphe 3.2. On a alors $\hat{\mu}_\sigma = \hat{g}_\sigma \hat{\mu}$. Alors les mêmes manipulations que pour la preuve du théorème 3.9 donnent

$$\begin{aligned} \int_{\mathbb{R}^d} e^{i\langle x, \xi \rangle} \hat{\mu}_\sigma(\xi) \lambda_d(d\xi) &= \frac{1}{\sigma^d} \int_{\mathbb{R}^d} e^{i\langle x, \xi \rangle} g_{1/\sigma}(\xi) \hat{\mu}(\xi) \lambda_d(d\xi) \\ &= \frac{1}{\sigma^d} \int_{\mathbb{R}^d} \widehat{e_x g_{1/\sigma}}(y) \mu(dy) \\ &= \int_{\mathbb{R}^d} g_\sigma(y - x) \mu(dy) \\ &= \mu_\sigma(x). \end{aligned}$$

Et toujours comme auparavant, le théorème de convergence dominée (par $|\hat{\mu}|$) montre que la seconde intégrale de cette chaîne d'égalités converge lorsque $\sigma \rightarrow 0$ vers

$$\frac{1}{(2\pi)^{d/2}} \int_{\mathbb{R}^d} e^{i\langle x, \xi \rangle} \hat{\mu}(\xi) d\xi,$$

que l'on notera $f(x)$. Notons aussi que ces intégrales sont toutes majorées par $\|\hat{\mu}\|_1$. Il reste à montrer que f est la densité de μ par rapport à λ_d . Pour cela, soit h une fonction continue à support compact. On a alors, par une nouvelle application du théorème de Fubini

$$\int_{\mathbb{R}^d} h(x) \mu_\sigma(x) \lambda_d(dx) = \int_{\mathbb{R}^d} g_\sigma * h(x) \mu(dx).$$

Par ce qui précède, et le théorème de convergence dominée, le membre de gauche converge vers $\int_{\mathbb{R}^d} h(x) f(x) \lambda_d(dx)$. Par ailleurs, comme $g_\sigma * h$ converge vers h uniformément sur les compacts par la proposition 1.9, tout en restant bornée par $\|h\|_\infty$, on conclut par convergence dominée que le membre de droite converge vers $\int_{\mathbb{R}^d} h(x) \mu(dx)$. Comme ceci est valable pour tout choix de h , continue à support compact, on déduit par un argument de densité que $\mu(dx) = f(x) \lambda_d(dx)$. \square

Corollaire 3.18. *La transformée de Fourier définie sur les mesures signées est une application injective : si μ et ν sont deux mesures signées telles que $\hat{\mu} = \hat{\nu}$, alors $\mu = \nu$.*

Démonstration. Sous ces hypothèses, on a $\widehat{\mu - \nu} = 0$, qui est dans L^1 , et on déduit que $\mu - \nu$ est à densité, et que cette densité est nulle. Donc $\mu = \nu$. \square

3.6 Une application à l'analyse de l'équation de la chaleur

Soit f une fonction intégrable sur \mathbb{R}^d . L'équation de la chaleur sur $\mathbb{R}^d \times \mathbb{R}_+$ avec condition initiale f est le système d'équations suivant, d'inconnue une fonction $u = u(x, t)$ sur $\mathbb{R}^d \times \mathbb{R}_+$:

$$\begin{cases} \partial_t u = \frac{1}{2} \Delta u & \text{sur } \mathbb{R}^d \times]0, \infty[\\ u(\cdot, 0) = f \end{cases} \quad (3.8)$$

Ici, Δ désigne le laplacien de \mathbb{R}^d agissant sur la première coordonnée :

$$\Delta u(x, t) = \sum_{j=1}^d \frac{\partial^2 u}{\partial x_j^2}(x, t).$$

Cette équation décrit la température $u(x, t)$ au point x et à l'instant t d'un matériau homogène, étant donnée la condition initiale $u(x, 0) = f(x)$.

En supposant que la solution u soit assez « régulière » pour qu'on puisse en prendre la transformée de Fourier $\hat{u} = \hat{u}(\xi, t)$, en la variable x , et pour que les formules usuelles sur la transformée de Fourier des dérivées partielles de u soient autorisées, on obtient que l'équation de la chaleur revient à

$$\begin{cases} \partial_t \hat{u} = -\frac{|\xi|^2}{2} \hat{u} & \text{sur } \mathbb{R}^d \times]0, \infty[\\ \hat{u}(\cdot, 0) = \hat{f} \end{cases}$$

On résout cette équation en

$$\hat{u}(\xi, t) = e^{-t\frac{|\xi|^2}{2}} \hat{f}(\xi) = \widehat{g_{\sqrt{t}}(\xi) f(\xi)} = \widehat{g_{\sqrt{t}} * f(\xi)},$$

où g_σ est comme d'habitude la densité gaussienne. Par injectivité de la transformée de Fourier, cela conduit à $u = g_{\sqrt{t}} * f$.

Rétrospectivement, le théorème de dérivation sous l'intégrale, joint au fait que la fonction $(x, t) \mapsto g_{\sqrt{t}}(x)$ vérifie la première équation de (3.8), implique que $u = g_{\sqrt{t}} * f$ la satisfait aussi. En revanche, il convient de s'interroger sur la mesure dans laquelle la condition initiale est bien vérifiée, puisque g_0 n'est pas définie a priori. Cependant, les résultats sur les approximations de l'identité montrent que $g_{\sqrt{t}} * f$ converge dans L^1 vers f , ce qui peut s'interpréter comme une version faible de la condition initiale. Si f est plus régulière, par exemple continue et bornée sur \mathbb{R}^d , alors la proposition 1.9 montre que la convergence est au sens ponctuel, et même uniforme sur les compacts. On dit que la fonction $(x, t) \mapsto g_{\sqrt{t}}(x)$ est la *solution fondamentale* de l'équation de la chaleur dans \mathbb{R}^d ,

Bien sûr, nous n'avons pas parlé ici du problème d'*unicité* de la solution, qui n'est d'ailleurs pas vérifiée ici.

La transformée de Fourier est un outil tout aussi commode pour analyser d'autres équations aux dérivées partielles, comme l'équation des ondes

$$\begin{cases} \partial_{tt} u = \frac{1}{2} \Delta u & \text{sur } \mathbb{R}^d \times]0, \infty[\\ u(\cdot, 0) = f, \quad \partial_t u(\cdot, 0) = g \end{cases},$$

ou l'équation de Schrödinger

$$\begin{cases} i \partial_t u + \frac{1}{2} \Delta u = 0 & \text{sur } \mathbb{R}^d \times]0, \infty[\\ u(\cdot, 0) = f \end{cases}.$$

On laisse au lecteur le soin de proposer des solutions (plus ou moins) explicites de ces équations.

Remarque. La méthode de résolution d'une équation différentielle ou aux dérivées partielles consistant à « passer aux transformées de Fourier » présuppose toujours une régularité et une décroissance à l'infini *a priori* sur les solutions, et elle peut donner en principe que des solutions particulières. On pourra par exemple s'en convaincre en essayant de trouver une solution à l'équation différentielle $y' = y + f$ par cette méthode.

Chapitre 4

Changement de variables

Ce court chapitre a pour objet de faire quelques commentaires sur la notion de mesure image et de changement de variables.

4.1 Mesure image

Rappelons que si (E, \mathcal{E}) et (F, \mathcal{F}) sont deux espaces mesurables, si μ est une mesure (disons positive, même si la discussion peut s'étendre naturellement au cas signé) sur (E, \mathcal{E}) et si $f: E \rightarrow F$ est une fonction mesurable (par rapport aux tribus \mathcal{E} et \mathcal{F}), alors on peut définir une nouvelle mesure $f_*\mu$ sur (F, \mathcal{F}) , appelée la mesure-image de μ par f , par la formule

$$f_*\mu(A) = \mu(f^{-1}(A)), \quad A \in \mathcal{F}.$$

De façon équivalente, pour toute fonction $g: F \rightarrow \mathbb{R}_+$ mesurable positive, on a que

$$\int_F g(y) f_*\mu(dy) = \int_E g(f(x)) \mu(dx).$$

Pour le voir, constatons que cette formule est exactement équivalente à la définition dans le cas où $g = \mathbf{1}_A$ est une indicatrice, avec $A \in \mathcal{F}$, et que l'on déduit la formule ci-dessus par un argument standard, en commençant par les fonctions étagées, puis en passant par une limite monotone à toutes les fonctions mesurables positives.

On voit avec cette formule qu'un calcul de mesure image est un « changement de variables », où la variable d'intégration y devient $f(x)$.

4.2 Coordonnées polaires dans \mathbb{R}^d

Pour calculer des intégrales sur \mathbb{R}^d , il est souvent commode d'avoir recours aux coordonnées polaires, surtout lorsque la fonction intégrée ne dépend que de la norme (euclidienne) $f(x) = \tilde{f}(|x|)$. C'est-à-dire qu'on veut décrire un point $x \in \mathbb{R}^d \setminus \{0\}$ par sa norme et son « argument » $(|x|, x/|x|)$, la quantité $x/|x|$ étant un point de la sphère unité $\mathbb{S}^{d-1} = \{x \in \mathbb{R}^d: \sum_{j=1}^d x_j^2 = 1\}$. On munit ce dernier espace de la topologie induite par \mathbb{R}^d (ainsi un ouvert de \mathbb{S}^{d-1} est de la forme $U \cap \mathbb{S}^{d-1}$, où U est un ouvert de \mathbb{R}^d) et de la tribu borelienne $\mathcal{B}(\mathbb{S}^{d-1})$ associée. La fonction $\theta: x \mapsto x/|x|$ est continue, donc mesurable, de $\mathbb{R}^d \setminus \{0\}$ dans \mathbb{S}^{d-1} . On la prolonge de façon arbitraire au point 0, ce qui ne jouera pas de rôle dans ce qui suit.

Définition 4.1. La mesure uniforme sur \mathbb{S}^{d-1} est la mesure image de $d dx \mathbf{1}_{B^d(0,1)}(x)$ par l'application θ , où $B^d(0,1)$ est la boule euclidienne fermée $\{x \in \mathbb{R}^d : |x| \leq 1\}$. On la note ω_d . Plus explicitement, pour $A \in \mathcal{B}(\mathbb{S}^{d-1})$, on a

$$\omega_d(A) = d \text{Leb}(\{rx : x \in A, r \in [0, 1]\}).$$

La masse totale de ω_d est donnée par

$$\omega_d(\mathbb{S}^{d-1}) = d \text{Leb}(B^d(0, 1)) = \frac{2\pi^{d/2}}{\Gamma(d/2)},$$

de sorte que $(\omega_d(\mathbb{S}^{d-1}), d \geq 1) = (2, 2\pi, 4\pi, 2\pi^2, \dots)$. En particulier

$$\omega_1 = \delta_{-1} + \delta_1,$$

ω_2 est l'image de la mesure de Lebesgue sur $]-\pi, \pi]$ par l'application $x \mapsto e^{ix}$, si l'on identifie \mathbb{R}^d à \mathbb{C} (ce qu'on laisse en exercice), et sa masse totale est 2π , la circonférence du cercle unité.

Proposition 4.2. La mesure image de la mesure de Lebesgue sur \mathbb{R}^d par l'application $\varphi : x \mapsto (|x|, x/|x|)$ (définie de façon arbitraire en $x=0$) de \mathbb{R}^d dans $\mathbb{R}_+ \times \mathbb{S}^{d-1}$ est la mesure produit $r^{d-1} dr \mathbf{1}_{\{r \geq 0\}} \omega_d(du)$.

Démonstration. Fixons $0 \leq a < b$ et $A \in \mathcal{B}(\mathbb{S}^{d-1})$, et montrons que

$$\varphi_* \text{Leb}([a, b] \times A) = \int_{]a, b]} r^{d-1} dr \int_A \omega_d(du),$$

la conclusion s'ensuira par un lemme de classe monotone, puisque les ensembles de la forme $]a, b] \times A$ ci-dessus engendrent la tribu produit, et que l'on peut recouvrir $\mathbb{R}^d \setminus \{0\}$ par une réunion dénombrable de tels ensembles, par exemple $]2^n, 2^{n+1}] \times \mathbb{S}^{d-1}, n \in \mathbb{Z}$. Or

$$\varphi^{-1}([a, b] \times A) = bC(A) \setminus aC(A)$$

où $C(A) = \{x \in B(0, 1) : x/|x| \in A\}$ est tel que $\text{Leb}(C(A)) = \omega_d(A)/d$ par définition. Par les propriétés élémentaires de la mesure de Lebesgue, on a donc

$$\varphi_* \text{Leb}([a, b] \times A) = \frac{b^d - a^d}{d} \omega_d(A) = \int_{]a, b]} r^{d-1} dr \omega_d(A)$$

comme voulu. □

On en déduit le changement de variables en « coordonnées polaires » dans \mathbb{R}^d :

$$\int_{\mathbb{R}^d} f(x) dx = \int_{\mathbb{R}_+} r^{d-1} dr \int_{\mathbb{S}^{d-1}} \omega_d(du) f(ru)$$

pour toute fonction f mesurable positive de \mathbb{R}^d dans \mathbb{R} .

4.3 Changement de variables linéaire

La mesure image de la mesure de Lebesgue par un isomorphisme affine s'exprime très simplement.

Proposition 4.3. Soit $M \in \mathcal{M}_d(\mathbb{R})$ une matrice carrée, et $a \in \mathbb{R}^d$. Alors pour tout $A \in \mathcal{B}(\mathbb{R}^d)$, on a que

$$\text{Leb}(MA + a) = |\det(M)|\text{Leb}(A),$$

où $MA + a = \{Mx + a : x \in A\}$.

En particulier, pour $a = 0$ et $A = [0, 1]^d$, on obtient que la mesure de Lebesgue du parallélépipède déterminé par les vecteurs colonne de M , c'est-à-dire l'ensemble $M[0, 1]^d$, a pour volume $|\det(M)|$. Un déterminant est un volume !

Démonstration. Si M n'est pas inversible, son image est contenue dans un hyperplan de \mathbb{R}^d , dont la mesure est nulle (par une transformation orthogonale, on peut toujours ramener un tel hyperplan sur $\{x_1 = 0\}$, dont la mesure est nulle par le théorème de Fubini). On obtient bien le résultat dans ce cas. Donc supposons $M \in \text{GL}_d(\mathbb{R})$. Par invariance de la mesure de Lebesgue par les translations, on peut aussi supposer que $a = 0$.

La formule $\mu(A) = \text{Leb}(MA)$, avec $A \in \mathcal{B}(\mathbb{R}^d)$, définit de façon évidente une mesure (car M est inversible), qui est de plus invariante par translation puisque pour tout $b \in \mathbb{R}^d$,

$$\mu(A + b) = \text{Leb}(MA + Mb) = \text{Leb}(MA) = \mu(A).$$

Par conséquent, μ est nécessairement un multiple scalaire de la mesure de Lebesgue, c'est-à-dire qu'il existe $c \geq 0$ tel que $\mu(A) = c \text{Leb}(A)$ pour tout A . Clairement, on a $c > 0$. Il reste à montrer que $c = |\det(M)|$.

Si $M \in O_d(\mathbb{R})$ est orthogonale, c'est l'invariance de la mesure de Lebesgue par les transformations orthogonales que nous avons déjà utilisé ci-dessus, mais que nous pouvons redémontrer facilement en constatant que $\text{Leb}(MB(0, 1)) = \text{Leb}(B(0, 1)) > 0$, puisqu'une transformation orthogonale préserve la norme euclidienne. On en tire immédiatement que $c = 1 = |\det(M)|$.

Si $M \in S_d^{++}(\mathbb{R})$ est symétrique définie positive, le théorème spectral stipule qu'on peut trouver une matrice orthogonale U telle que $MP = P\Delta$, où $\Delta = \text{diag}(a_1, \dots, a_d)$ est la matrice diagonale dont les coefficients diagonaux sont les valeurs propres de M , qui sont strictement positives par hypothèse. On obtient alors que

$$\mu(P[0, 1]^d) = \text{Leb}(MP[0, 1]^d) = \text{Leb}(P\Delta[0, 1]^d) = \text{Leb}(\Delta[0, 1]^d)$$

où l'on a utilisé le cas précédent. Cette dernière quantité est la mesure du pavé $[0, a_1] \times \dots \times [0, a_d]$, de mesure $a_1 \dots a_d = \det(M)$. Mais par ailleurs, ceci vaut $c \text{Leb}(P[0, 1]^d) = c \text{Leb}([0, 1]^d) = c$, en utilisant encore le cas orthogonal. Le résultat s'ensuit.

Dans le cas général, on peut écrire $M \in \text{GL}_d(\mathbb{R})$ de façon unique sous la forme $M = OS$ avec $O \in O_d(\mathbb{R})$ orthogonale, et $S \in S_d^{++}(\mathbb{R})$ symétrique définie positive. C'est la *décomposition polaire* des matrices : il est facile de voir que S est l'unique racine carrée symétrique positive de la matrice M^*M , et que $O = MS^{-1}$ est bien orthogonale dans ce cas... On déduit des deux cas précédents que

$$\mu(A) = \text{Leb}(OSA) = \text{Leb}(SA) = \det(S)\text{Leb}(A) = |\det(M)|\text{Leb}(A).$$

D'où le résultat. □

4.4 Changement de variables \mathcal{C}^1

Nous concluons par un théorème très utile en pratique.

Théorème 4.4. *Soit D un ouvert de \mathbb{R}^d et f un \mathcal{C}^1 -difféomorphisme de D sur son image $f(D)$. Alors pour toute fonction $g: \mathbb{R}^d \rightarrow \mathbb{R}_+$ mesurable, on a*

$$\int_D g(f(x)) J_f(x) dx = \int_{f(D)} g(y) dy,$$

où le Jacobien $J_f(x) = |\det Df_x|$ est la valeur absolue du déterminant de la différentielle de f en x . Autrement dit, la mesure image de $dx \mathbf{1}_D(x)$ par f est la mesure $dy \mathbf{1}_{f(D)}(y) / J_f(f^{-1}(y))$.

La preuve est omise.

Partie II

Bases des probabilités

Chapitre 5

Bases de la théorie des probabilités

Dans ce chapitre, nous donnons les premières notions de la théorie moderne des probabilités (issue des années 1930)

5.1 Espaces de probabilités, variables aléatoires

L'objet de base de la théorie des probabilités est un *espace de probabilités*, souvent noté

$$(\Omega, \mathcal{F}, \mathbb{P}),$$

qui est un espace mesuré, où \mathbb{P} est une mesure de probabilités sur (Ω, \mathcal{F}) , c'est-à-dire une mesure positive telle que $\mathbb{P}(\Omega) = 1$. Intuitivement, les éléments de Ω , parfois appelés « événements élémentaires », correspondent aux issues possibles d'une expérience aléatoire donnée, et les parties mesurables $A \subset \mathcal{F}$ sont appelés des *événements*. La quantité $\mathbb{P}(A)$ désigne la probabilité que l'événement A soit observé. On dit que l'événement A est *presque sûr* si $\mathbb{P}(A) = 1$.

Premiers exemples d'espaces de probabilités. L'exemple le plus familier et le plus élémentaire d'expérience aléatoire consiste à choisir uniformément un élément dans un ensemble E fini. Un choix naturel d'espace de probabilités adapté est de prendre $\Omega = E$, $\mathcal{F} = 2^\Omega$ la tribu des parties de E , et \mathbb{P} la *mesure uniforme* sur E , définie par

$$\mathbb{P} = \frac{1}{\text{card}(E)} \sum_{x \in E} \delta_x,$$

ou plus explicitement,

$$\mathbb{P}(A) = \frac{\text{card}(A)}{\text{card}(E)}, \quad A \subset E.$$

Par exemple, si l'on cherche à modéliser un jeu de pile-ou-face équilibré, on peut poser $\Omega = \{0, 1\}$ et $\mathcal{F} = \{\emptyset; \{0\}; \{1\}; \{0, 1\}\} = 2^\Omega$, et enfin $\mathbb{P} = (\delta_0 + \delta_1)/2$ est la mesure de Bernoulli. Un lancer de n pièces successivement est modélisé par l'espace produit $\Omega = \{0, 1\}^n$, muni de la tribu des parties 2^Ω et de la mesure uniforme

$$\mathbb{P}(\{\omega_1, \dots, \omega_n\}) = \frac{1}{2^n},$$

qui est aussi la mesure produit des mesures de Bernoulli.

De même, un lancer de dé équilibré à 6 faces peut être modélisé en posant $\Omega = \{1, 2, 3, 4, 5, 6\}$, $\mathcal{F} = 2^\Omega$ et \mathbb{P} la mesure uniforme sur Ω : $\mathbb{P} = 6^{-1} \sum_{i=1}^6 \delta_i$, et n lancers successifs correspondent à l'espace produit n fois.

Une infinité de lancers de pièces ? Il est plus difficile de considérer l'expérience (de pensée !) consistant à jeter une pièce ou un dé une infinité de fois. Il est naturel de considérer l'espace produit $\Omega = \{0, 1\}^{\mathbb{N}^*}$ de toutes les suites à valeurs dans $\{0, 1\}$. Mais cette fois on ne va pas considérer la σ -algèbre de tous les sous-ensembles, qui est trop grande. Si $\varepsilon_1, \dots, \varepsilon_n \in \{0, 1\}$, notons

$$A_{\varepsilon_1, \dots, \varepsilon_n} = \{\omega = (\omega_1, \omega_2, \dots) \in \Omega : \omega_1 = \varepsilon_1, \dots, \omega_n = \varepsilon_n\}.$$

Soit \mathcal{F} la plus petite σ -algèbre rendant mesurable les ensembles $A_{\varepsilon_1, \dots, \varepsilon_n}$ pour tout choix de $\varepsilon_1, \dots, \varepsilon_n$. C'est aussi la plus petite σ -algèbre rendant mesurable les applications de projection $X_i: \Omega \rightarrow \{0, 1\}$ définies par $X_i(\omega) = \omega_i$. On dit que \mathcal{F} est la σ -algèbre produit sur l'espace produit Ω . Enfin, on munit l'espace (Ω, \mathcal{F}) de la mesure produit \mathbb{P} , qui est l'unique mesure de probabilités vérifiant

$$\mathbb{P}(A_{\varepsilon_1, \dots, \varepsilon_n}) = \frac{1}{2^n}, \quad \text{pour tout } \varepsilon_1, \dots, \varepsilon_n \in \{0, 1\}.$$

Notons que, si elle existe, cette mesure est bien unique par le lemme de classe monotone, les événements $A_{\varepsilon_1, \dots, \varepsilon_n}$ formant une famille stable par intersection finie, et qui engendre \mathcal{F} . L'existence de \mathbb{P} est plus subtile. Nous allons la construire explicitement à l'aide d'un espace de probabilités annexe $([0, 1[, \mathcal{B}([0, 1[), \lambda)$ où λ est la mesure de Lebesgue sur $[0, 1[$.

À tout $x \in [0, 1[$, on associe une suite $(\omega_1(x), \omega_2(x), \dots)$ donnant le développement dyadique de x :

$$x = \sum_{i=1}^{\infty} \frac{\omega_i(x)}{2^i}.$$

Pour que cette suite soit définie de façon unique, on suppose que l'on choisit si nécessaire la suite $(\omega_i(x), i \geq 1)$ de sorte qu'elle ne stationne pas à 1. Ainsi, par exemple, on associe la suite $(1, 0, 0, 0, \dots)$ au nombre $1/2$, et non la suite $(0, 1, 1, 1, \dots)$.

Les applications $\omega_i: [0, 1[\rightarrow \{0, 1\}$ sont alors mesurables. En effet, si l'on pose $\theta(x) = 2x - \lfloor 2x \rfloor$ la partie fractionnaire de $2x$, on a la formule explicite $\omega_1(x) = \lfloor 2x \rfloor$, et pour tout $i \geq 1$,

$$\omega_i(x) = \lfloor 2\theta^{o(i-1)}(x) \rfloor = \lfloor 2^i x \rfloor - 2 \lfloor 2^{i-1} x \rfloor,$$

où $\theta^{o i}$ est la composée de θ avec elle-même i fois. De ce fait, l'application $\varphi: x \mapsto (\omega_1(x), \omega_2(x), \dots)$ de $([0, 1[, \mathcal{B}([0, 1[)$ dans (Ω, \mathcal{F}) est mesurable, puisque les applications coordonnées $\omega_i: x \mapsto \omega_i(x)$ le sont. On laisse au lecteur le soin de vérifier ces assertions. Posons alors \mathbb{P} la mesure image de λ par φ . On a que pour tout $\varepsilon_1, \dots, \varepsilon_n \in \{0, 1\}$,

$$\mathbb{P}(A_{\varepsilon_1, \dots, \varepsilon_n}) = \lambda(\varphi^{-1}(A_{\varepsilon_1, \dots, \varepsilon_n})) = \lambda(\{x \in [0, 1[: \omega_1(x) = \varepsilon_1, \dots, \omega_n(x) = \varepsilon_n\}) = \frac{1}{2^n},$$

la dernière égalité s'obtenant en remarquant que $\varphi^{-1}(A_{\varepsilon_1, \dots, \varepsilon_n})$ est le sous-intervalle

$$\left[\sum_{i=1}^n \frac{\varepsilon_i}{2^i}, \sum_{i=1}^n \frac{\varepsilon_i}{2^i} + \frac{1}{2^n} \right[.$$

Ainsi, la mesure \mathbb{P} répond bien à la définition de la mesure produit, ce qui donne l'existence de cette dernière. Notons que l'espace $(\Omega, \mathcal{F}, \mathbb{P})$ que nous venons de construire est d'une complexité similaire à l'espace $([0, 1[, \mathcal{B}([0, 1[), \lambda)$, au sens précis que φ réalise entre ces espaces un isomorphisme d'espaces mesurés. L'inverse de φ est en effet mesurable à son tour, et l'on a $\lambda = (\varphi^{-1})_*\mathbb{P}$. En ce sens, construire un espace de probabilités modélisant une infinité de lancers de pièces équilibrées est donc du même ordre de difficulté que de construire la mesure de Lebesgue sur \mathbb{R} .

Exercice. Construire un espace de probabilités modélisant une infinité de lancers de pile-ou-face, mais où la probabilité d'obtenir pile est un nombre $p \in [0, 1]$ quelconque.

Variables aléatoires. Le choix d'un espace de probabilités correspondant à une situation concrète donnée n'est pas (jamais !) unique. Si l'on peut considérer les deux exemples ci-dessus comme « minimaux » en un sens, considérons par exemple l'expérience consistant à jeter deux dés à 6 faces et à observer le résultat de la somme des chiffres indiqués. Une première possibilité consiste à lister les résultats possibles, qui sont $\Omega_1 = \{2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12\}$. On munit Ω_1 de la tribu \mathcal{F}_1 de l'ensemble des parties. Le choix de la mesure de probabilités \mathbb{P}_1 nécessite un temps de réflexion : l'on pose

$$\mathbb{P}_1 = \frac{\delta_2 + 2\delta_3 + 3\delta_4 + 4\delta_5 + 5\delta_6 + 6\delta_7 + 5\delta_8 + 4\delta_9 + 3\delta_{10} + 2\delta_{11} + \delta_{12}}{36}.$$

En effet, parmi les 36 possibilités de résultats des deux dés $(i, j) \in \{1, 2, 3, 4, 5, 6\}^2$, où la première coordonnée donne le résultat du premier dé et la seconde coordonnée donne le résultat du second dé, il y en a une, $(1, 1)$, qui donne le résultat 2, deux, $(1, 2)$ et $(2, 1)$, qui donnent le résultat 3, et ainsi de suite. On a donc fait une petite excursion par un autre espace de probabilités, qui est l'espace $\Omega_2 = \{1, 2, 3, 4, 5, 6\}^2$, muni de la tribu \mathcal{F}_2 des parties, et de la mesure de probabilité uniforme \mathbb{P}_2 . Cette tribu modélise le résultat des deux lancers de dés, et quand on s'intéresse seulement à la probabilité $\mathbb{P}_1(\{k\})$ que la somme des résultats fasse $k \in \Omega_1$, on voit que ceci est la probabilité de l'événement $\{(i, j) \in \Omega_2 : i + j = k\}$, c'est-à-dire que l'on pose :

$$\mathbb{P}_1(\{k\}) = \mathbb{P}_2(\{(i, j) \in \Omega_2 : i + j = k\}).$$

On peut réexprimer cela en introduisant l'application $X: \Omega_2 \rightarrow \Omega_1$ définie par $X((i, j)) = i + j$, en disant que pour tout k ,

$$\mathbb{P}_1(\{k\}) = \mathbb{P}_2(X^{-1}(\{k\})),$$

et ceci reste évidemment valable si l'on remplace $\{k\}$ par n'importe quelle partie de Ω_1 . Autrement dit, *la mesure \mathbb{P}_1 est la mesure image de \mathbb{P}_2 par l'application X* . On appelle les applications mesurables entre espaces de probabilités des *variables aléatoires*, et ces dernières sont, en quelque sorte, les objets qui permettent de passer d'un choix d'espace de probabilités à un autre. Souvent, c'est l'étude des variables aléatoires qui est prépondérante en probabilités, bien plus que l'espace $(\Omega, \mathcal{F}, \mathbb{P})$, qu'il suffit en général de choisir « assez gros ». Nous reviendrons sur ces considérations plus tard.

Noter que dans le choix de Ω_2 , nous avons *distingué* les deux dés, comme s'il y en avait un rouge et un noir par exemple. On aurait pu procéder autrement, et poser $\Omega_3 = \{\{i, j\} : i, j \in \{1, 2, 3, 4, 5, 6\}\}$ l'ensemble des parties de $\{1, 2, 3, 4, 5, 6\}$ à au plus deux éléments : l'événement élémentaire $\{i\}$ est interprété par le fait que les deux dés ont donné le résultat i . On pose alors $\mathbb{P}_3(\{i, j\}) = 1/18$ si $i \neq j$ et $\mathbb{P}_3(\{i\}) = 1/36$. La variable aléatoire donnant la somme des dés est l'application X' telle que $X'(\{i, j\}) = i + j$ si $i \neq j$ et $X_3(\{i\}) = 2i$. Bien sûr, ce choix est plutôt maladroit, mais il donne le même résultat que ci-dessus : \mathbb{P}_1 est aussi la mesure image de \mathbb{P}_3 par X' .

Exercice. Trouver une variable aléatoire $X'' : \Omega_2 \rightarrow \Omega_3$ envoyant la mesure \mathbb{P}_2 sur \mathbb{P}_3 , et telle que $X = X' \circ X''$.

Définition 5.1. Une variable aléatoire sur l'espace de probabilités $(\Omega, \mathcal{F}, \mathbb{P})$ est une application mesurable $X : (\Omega, \mathcal{F}) \rightarrow (E, \mathcal{E})$ où (E, \mathcal{E}) est un ensemble mesurable.

Remarque. Certains ouvrages appellent *variable aléatoire* une application mesurable d'un espace de probabilités dans \mathbb{R} ou \mathbb{C} , et *vecteur aléatoire* une application mesurable d'un espace de probabilités dans \mathbb{R}^d ou \mathbb{C}^d , nous préférons ici nous placer dans la plus grande généralité.

Définition 5.2. La loi de la variable aléatoire X est alors la mesure image $\mathbb{P}_X = X_*\mathbb{P}$ de \mathbb{P} par X , définie par

$$\mathbb{P}_X(A) = \mathbb{P}(X^{-1}(A)), \quad A \in \mathcal{E}.$$

C'est une mesure de probabilités sur (E, \mathcal{E}) , définissant un nouvel espace de probabilités $(E, \mathcal{E}, \mathbb{P}_X)$.

Remarque. Noter que, *stricto sensu*, une variable aléatoire est seulement définie sur un espace mesurable (Ω, \mathcal{F}) plutôt que sur un espace de probabilités (la mesure \mathbb{P} n'intervient pas). En revanche, la loi d'une variable aléatoire en dépend de façon cruciale, et c'est pourquoi on fait rentrer \mathbb{P} dans la définition de X pour parler de la loi de X . Les puristes diront qu'une variable aléatoire est un couple (X, \mathbb{P}) , ou un sextuplet $(\Omega, \mathcal{F}, \mathbb{P}, E, \mathcal{E}, X)$, etc...

Remarque. On adopte souvent l'écriture suivante

$$\mathbb{P}_X(A) = \mathbb{P}(X \in A),$$

c'est-à-dire qu'on écrit $X^{-1}(A) = \{\omega : X(\omega) \in A\} = X \in A$, ce qui est encore une façon d'effacer le rôle de Ω . Par extension, cela donne l'écriture parfois utilisée $\mathbb{P}_X(dx) = \mathbb{P}(X \in dx)$ pour la loi de X .

Variations aléatoires discrètes. Notons que, si X est à valeurs dans un ensemble dénombrable E (on parle de variable aléatoire discrète), muni de la tribu 2^E , alors on a, pour tout $A \subset E$,

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

Autrement dit, la connaissance de la loi de X revient à celle des quantités $p_x = \mathbb{P}(X = x)$, et l'on a

$$\mathbb{P}_X(dx) = \sum_{x \in E} p_x \delta_x.$$

Variabes aléatoires à densité. Soit X une variable aléatoire à valeurs dans \mathbb{R}^d . On dit que X est à densité si sa loi est absolument continue par rapport à λ_d . Dans ce cas, le théorème de Radon-Nikodym implique qu'il existe une fonction mesurable positive $f_X: \mathbb{R}^d \rightarrow \mathbb{R}_+$ d'intégrale 1 par rapport à λ_d , telle que

$$\mathbb{P}_X(dx) = f_X(x)dx.$$

On appelle cette fonction la densité de (la loi de) X , même si elle n'est définie que λ_d -presque partout.

Remarque. Soit (E, \mathcal{E}) un espace mesurable, et μ une mesure de probabilités sur (E, \mathcal{E}) . On peut naturellement poser la question suivante : existe-t-il un espace de probabilités $(\Omega, \mathcal{F}, \mathbb{P})$ et une variable aléatoire $X: \Omega \rightarrow E$ sur cet espace, telle que X a pour loi μ , c'est-à-dire que $\mathbb{P}_X = \mu$. La réponse est oui : il suffit de prendre $(\Omega, \mathcal{F}, \mathbb{P}) = (E, \mathcal{E}, \mu)$ et de prendre pour X la fonction identité de E . On a bien $\mathbb{P}_X = X_*\mathbb{P} = X_*\mu = \mu$.

Définition 5.3. *L'application X construite ci-dessus est appelée la variable aléatoire canonique de loi μ .*

5.2 Espérance d'une variable aléatoire

a. Définition et formule de transfert

Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace de probabilités et $X: \Omega \rightarrow \mathbb{R}$ une variable aléatoire à valeurs réelles. Si X est positive, ou si $X \in L^1(\Omega, \mathcal{F}, \mathbb{P})$ est intégrable, on note

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

l'intégrale de X par rapport à \mathbb{P} , et on l'appelle *l'espérance de X* , ou encore la *moyenne* de X . On définit de même l'espérance d'une variable aléatoire intégrable à valeurs complexes. Plus généralement, si $X: \Omega \rightarrow \mathbb{C}^d$ est à valeurs vectorielles, on note $\mathbb{E}[X] = (\mathbb{E}[X_1], \dots, \mathbb{E}[X_d])$ où X_1, \dots, X_d sont les coordonnées de X , qui sont à leur tour des variables aléatoires.

Si par exemple $X = \mathbf{1}_A$ est l'indicatrice de l'événement $A \in \mathcal{F}$, on a

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

C'est ce dont on s'est servi pour montrer la formule d'inclusion-exclusion.

L'espérance hérite des propriétés de linéarité et de positivité de l'intégrale.

Proposition 5.4. *Pour tout scalaire a et toutes variables aléatoires X et Y , on a*

$$\mathbb{E}[aX + Y] = a\mathbb{E}[X] + \mathbb{E}[Y]$$

dès lors que X, Y sont intégrables, ou bien que X, Y, a sont positifs.

De plus, si $X \geq 0$, on a $\mathbb{E}[X] \geq 0$ avec égalité si et seulement si $X = 0$ \mathbb{P} -presque partout.

La preuve est immédiate.

Proposition (Formule de transfert). Soit $X: \Omega \rightarrow E$ est une variable aléatoire à valeurs dans un espace mesurable (E, \mathcal{E}) , et si $f: E \rightarrow \mathbb{R}_+$ est une application mesurable positive, alors $f(X) = f \circ X$ est une variable aléatoire positive, et on a

$$\mathbb{E}[f(X)] = \int_{\Omega} f(X(\omega)) \mathbb{P}(d\omega) = \int_E f(x) \mathbb{P}_X(dx).$$

Si maintenant $f: E \rightarrow \mathbb{R}_+$ est mesurable, on a que $f(X) \in L^1(\Omega, \mathcal{F}, \mathbb{P})$ si et seulement si $f \in L^1(E, \mathcal{E}, \mathbb{P}_X)$, et dans ce cas cette même formule reste valable.

Démonstration. Pour f de la forme $\mathbf{1}_A$ avec $A \in \mathcal{E}$ c'est juste la définition de la mesure image, et par linéarité on déduit la formule pour toutes les fonctions étagées, puis toutes les fonctions mesurables positives par un argument de convergence monotone. Enfin, on obtient le résultat pour toutes les fonctions f intégrables en décomposant $f = f^+ - f^-$, où l'on rappelle que $x^+ = x \vee 0$ et $x^- = (-x)^+$. \square

Noter que le dernier membre de la formule de transfert ne fait plus intervenir X qu'à travers sa loi \mathbb{P}_X , et en particulier, elle ne fait pas intervenir l'espace $(\Omega, \mathcal{F}, \mathbb{P})$. Comme mentionné plus haut, les probabilistes aiment bien se débarrasser de l'espace de probabilités à la première occasion !

b. Caractérisation de la loi à l'aide de l'espérance

Notons que la formule de transfert permet de caractériser efficacement la loi d'une variable aléatoire.

Proposition 5.5. Soit X une variable aléatoire à valeurs dans un espace mesurable E . Alors la loi de X est caractérisée par les espérances $\mathbb{E}[f(X)]$, où f décrit l'ensemble des fonctions mesurables bornées.

Démonstration. La preuve est triviale : si X et X' sont deux telles variables aléatoires telles que $\mathbb{E}[f(X)] = \mathbb{E}[f(X')]$ pour toute fonction f mesurable bornée, on prend $f = \mathbf{1}_A$ avec $A \in \mathcal{E}$ pour obtenir que $\mathbb{P}_X(A) = \mathbb{P}_{X'}(A)$, et donc $\mathbb{P}_X = \mathbb{P}_{X'}$, puisque ceci est valable pour tout $A \in \mathcal{E}$. \square

Par exemple, soit $X = (X_1, X_2, \dots, X_d)$ une variable aléatoire à valeurs dans \mathbb{R}^d . On suppose que la loi de X admet une densité f_X . Alors la loi de X_i admet à son tour une densité, donnée par

$$f_{X_i}(x) = \int_{\mathbb{R}^{d-1}} f_X(x_1, \dots, x_{i-1}, x, x_{i+1}, \dots, x_d) dx_1 \dots dx_{i-1} dx_{i+1} \dots dx_d.$$

En effet, si $h: \mathbb{R} \rightarrow \mathbb{R}$ est une fonction mesurable bornée, on a par le théorème de Fubini (et en notant $\widehat{dx}_i = dx_1 \cdots dx_{i-1} dx_{i+1} \cdots dx_d$)

$$\begin{aligned} \mathbb{E}[h(X_i)] &= \int_{\mathbb{R}^d} h(x_i) f_X(x_1, \dots, x_d) dx_1 \cdots dx_d \\ &= \int_{\mathbb{R}} h(x) dx \int_{\mathbb{R}^{d-1}} f_X(x_1, \dots, x_{i-1}, x, x_{i+1}, \dots, x_d) \widehat{dx}_i \\ &= \int_{\mathbb{R}} h(x) f_{X_i}(x) dx \end{aligned}$$

et l'on reconnaît la formule de $\mathbb{E}[h(Y)]$ où Y est une variable aléatoire de densité f_{X_i} . Noter que f_{X_i} est bien une fonction positive d'intégrale 1, ce qu'on obtient en prenant $h = 1$ dans le calcul précédent.

Remarque (lois marginales). En général, si $X = (X_1, \dots, X_d)$ est une variable aléatoire à valeurs dans un espace produit $E_1 \times \dots \times E_d$, la loi de X_i , qui est une loi sur E_i , est appelée la *i*-ème loi marginale de X . Il est vrai que la loi de X caractérise la loi de ses marginales : la *i*-ème loi marginale est en effet la mesure image de \mathbb{P}_X par la *i*-ème application de projection $E_1 \times \dots \times E_d \rightarrow E_i$. En revanche, **la réciproque est fautive !** Pour s'en convaincre, on peut prendre $X_1(i, j) = i$ et $X_2(i, j) = j$, variables aléatoires définies sur $\Omega = \{0, 1\}^2$, muni de la mesure produit $(\delta_0/2 + \delta_1/2)^{\otimes 2}$, et constater que $X = (X_1, X_2)$ et $Y = (X_1, X_1)$ ont les mêmes lois marginales (uniformes sur $\{0, 1\}$), tandis que X et Y n'ont certainement pas même loi, puisque

$$\mathbb{P}(X_1 = X_2) = 1/2 \neq 1 = \mathbb{P}(X_1 = X_1).$$

En pratique, il est utile de caractériser la loi d'une variable aléatoire X en calculant $\mathbb{E}[f(X)]$ pour le moins de fonctions f possible. Par exemple :

Exercice. Soit $d \geq 1$ un entier fixé, et H un sous-ensemble de $\mathcal{C}_c(\mathbb{R}^d, \mathbb{R})$ dense pour la norme uniforme. Si X est une variable aléatoire à valeurs dans \mathbb{R}^d , alors sa loi est caractérisée par les espérances $(\mathbb{E}[f(X)], f \in H)$.

On reviendra plus abondamment au paragraphe 5.3 sur d'autres critères de caractérisation de la loi d'une variable aléatoire.

c. Moments d'une variable aléatoire

Définition 5.6. Soit X une variable aléatoire à valeurs dans \mathbb{R} ou \mathbb{C} , et $k > 0$ un nombre entier. Si $X \in L^k(\Omega, \mathcal{F}, \mathbb{P})$, le nombre $\mathbb{E}[X^k]$ est appelé le moment d'ordre k de X . On appelle également la quantité $\mathbb{E}[|X|^k]$ le moment absolu d'ordre k de X .

Notons encore une fois que le moment d'ordre k ne dépend que de la loi de X , puisque

$$\mathbb{E}[X^k] = \int_{\mathbb{R}} x^k \mathbb{P}_X(dx)$$

dès que cette intégrale a un sens, par la formule de transfert. On parle donc aussi des moments d'une mesure de probabilités μ sur \mathbb{R} , égaux à $\int_{\mathbb{R}} x^k \mu(dx)$ si $x \mapsto x$ est dans $L^k(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mu)$.

Ainsi, le moment d'ordre 1 de X n'est autre que l'espérance de X , lorsqu'elle est bien définie. Notons que, comme une mesure de probabilités est une mesure finie, l'inégalité de Hölder implique que les espaces $L^p(\Omega, \mathcal{F}, \mathbb{P})$ forment une famille décroissante en $p \in [0, \infty]$ (par définition on note $L^0(\Omega, \mathcal{F}, \mathbb{P})$ l'ensemble de toutes les fonctions mesurables, définies \mathbb{P} -presque partout). En particulier, si X admet un moment d'ordre k , alors X admet un moment d'ordre k' pour tout entier $k' \leq k$.

Les moments de variables aléatoires sont des outils très utiles pour étudier la *queue de distribution* d'une variable aléatoire réelle X , que l'on définit comme la fonction $x \mapsto \mathbb{P}(X \geq x)$.

Proposition (inégalité de Markov). *Soit X une variable aléatoire positive. Alors on a, pour tout réel $p > 0$,*

$$\mathbb{P}(X \geq x) \leq \frac{\mathbb{E}[X^p]}{x^p}, \quad x > 0.$$

Démonstration. Quitte à remplacer X par X^p , il suffit de traiter le cas $p = 1$. On a alors

$$\mathbb{P}(X \geq x) = \mathbb{E}\left[\frac{X}{x} \mathbf{1}_{\{X \geq x\}}\right] \leq \frac{\mathbb{E}[X \mathbf{1}_{\{X \geq x\}}]}{x},$$

et on conclut en majorant l'indicatrice par 1. Noter cependant que cette étape intermédiaire est parfois utile. \square

Par ailleurs, en appliquant l'inégalité de Markov à la variable aléatoire positive $e^{\lambda X}$, on obtient

Corollaire (inégalité de Chernov). *Soit X une variable aléatoire à valeurs réelles, alors pour tout $\lambda > 0$ et tout $x \in \mathbb{R}$ on a*

$$\mathbb{P}(X \geq x) \leq e^{-\lambda x} \mathbb{E}[e^{\lambda X}].$$

Cette borne est le point de départ de la théorie des grandes déviations de sommes de variables aléatoires indépendantes, dont nous toucherons un mot plus bas.

d. Variance et covariance

Définition 5.7. *Soit $X \in L^2(\Omega, \mathcal{F}, \mathbb{P})$ une variable aléatoire de carré intégrable. La quantité*

$$\text{Var}(X) = \mathbb{E}[(X - \mathbb{E}[X])^2] = \|X - \mathbb{E}[X]\|_2^2$$

est appelée la variance de X . La quantité $\sqrt{\text{Var}(X)} = \|X - \mathbb{E}[X]\|_2$ est appelée l'écart-type de X .

Notons que l'écart-type de X est la distance au sens L^2 de X à la constante $\mathbb{E}[X]$. Comme par définition $\mathbb{E}[(X - \mathbb{E}[X]) \cdot 1] = 0$, on en déduit que $X - \mathbb{E}[X]$ est orthogonale (au sens du produit scalaire dans $L^2(\Omega, \mathcal{F}, \mathbb{P})$) au sous-espace des fonctions constantes. Par conséquent, l'écart-type est également la distance de X (au sens L^2) à ce sous-espace, et

$$\text{Var}(X) = \inf_{c \in \mathbb{R}} \mathbb{E}[(X - c)^2].$$

De plus, cet infimum est atteint uniquement en $\mathbb{E}[X]$, qui est donc la meilleure approximation possible de X par une constante, au sens des moindres carrés.

En développant le carré dans la définition de la variance, et par linéarité de l'espérance, on trouve la formule utile suivante :

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

valable pour tout $X \in L^2(\Omega, \mathcal{F}, \mathbb{P})$. Notons que, par définition, une variable aléatoire de carré intégrable est constante presque sûrement si et seulement si $\text{Var}(X) = 0$.

En appliquant l'inégalité de Markov à la variable aléatoire $|X - \mathbb{E}[X]|$ et avec $p = 2$, on obtient le résultat très utile suivant.

Corollaire (inégalité de Bienaymé-Chebychev). *Soit X une variable aléatoire dans $L^2(\Omega, \mathcal{F}, \mathbb{P})$. Alors on a, pour tout $x > 0$,*

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

Ainsi, la variance permet d'estimer la probabilité qu'une variable aléatoire s'écarte de sa moyenne.

Définition 5.8. *Soit $X, Y \in L^2(\Omega, \mathcal{F}, \mathbb{P})$ des variables aléatoires de carré intégrable. La covariance de X et Y est définie par*

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

La covariance est donc le produit scalaire dans $L^2(\Omega, \mathcal{F}, \mathbb{P})$ des variables aléatoires $X - \mathbb{E}[X]$ et $Y - \mathbb{E}[Y]$. Une formule alternative est donnée par

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

On a évidemment $\text{Cov}(X, X) = \text{Var}(X)$ et $\text{Cov}(X, Y) = \text{Cov}(Y, X)$.

On appelle parfois *corrélation* de X et Y la quantité normalisée

$$\text{Corr}(X, Y) = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}},$$

qu'on peut interpréter comme le cosinus de l'angle formé dans l'espace $L^2(\Omega, \mathcal{F}, \mathbb{P})$ par les vecteurs $X - \mathbb{E}[X]$ et $Y - \mathbb{E}[Y]$. C'est une quantité dans $[-1, 1]$ par l'inégalité de Cauchy-Schwarz.

Si maintenant X est une variable aléatoire à valeurs dans \mathbb{R}^d pour un entier $d \geq 1$, on note $X_i, 1 \leq i \leq d$ les applications coordonnées de X , qui sont à leur tour des variables aléatoires à valeurs dans \mathbb{R} . La *matrice de variance-covariance* de X est alors donnée par

$$\Sigma_X = (\text{Cov}(X_i, X_j))_{1 \leq i, j \leq d} \in \mathfrak{M}_d(\mathbb{R}).$$

Il s'agit d'une matrice symétrique positive. En effet, pour tout $x \in \mathbb{R}^d$, on a, par linéarité,

$$\langle x, \Sigma_X x \rangle = \text{Var}(\langle x, X \rangle) \geq 0$$

e. Médiane et quantiles

Soit X une variable aléatoire à valeurs réelles. Soit $\alpha \in]0, 1[$. On dit que q_α est un α -quantile de la loi de X (ou simplement de X) si

$$\mathbb{P}(X \leq q_\alpha) \geq \alpha \quad \text{et} \quad \mathbb{P}(X \geq q_\alpha) \geq 1 - \alpha.$$

Notons que tout nombre q est un $F_X(q)$ -quantile de X , puisque

$$\mathbb{P}(X \leq q) + \mathbb{P}(X \geq q) = 1 + \mathbb{P}(X = q) \geq 1.$$

Plus précisément, un même nombre q peut être un α -quantile de X pour différentes valeurs de α . Cela arrive si et seulement si $\{q\}$ est un atome de \mathbb{P}_X , et dans ce cas les valeurs correspondantes de α sont $[F_X(q-), F_X(q)]$.

Par ailleurs, en général, il n'y a pas unicité d'un α -quantile. Plus exactement, si l'on note

$$q_\alpha^- = \sup \{x \in \mathbb{R} : F_X(x) < \alpha\} \quad \text{et} \quad q_\alpha^+ = \inf \{x \in \mathbb{R} : F_X(x) > \alpha\}$$

alors l'ensemble des α -quantiles de X est l'intervalle $[q_\alpha^-, q_\alpha^+]$. Si $q_\alpha^- < q_\alpha^+$, alors $]q_\alpha^-, q_\alpha^+[$ est l'intervalle ouvert maximal où F_X est constante égale à α , et l'on a $\mathbb{P}(q_\alpha^- < X < q_\alpha^+) = 0$.

Enfin, si q_α et q_β sont respectivement un α -quantile et un β -quantile de X avec $\alpha < \beta$, alors $q_\alpha < q_\beta$.

Définition 5.9. On appelle médiane de la loi de X (ou plus simplement médiane de X) un $1/2$ -quantile de X . De façon équivalente, m est une médiane de X si

$$\mathbb{P}(X \geq m) \geq 1/2 \quad \text{et} \quad \mathbb{P}(X \leq m) \geq 1/2.$$

En général il n'y a pas unicité de la médiane de X .

Nous avons vu que l'espérance s'interprète comme meilleure approximation de X par une fonction constante au sens des moindres carrés. Une interprétation similaire de la médiane est possible, mais au sens L^1 .

Proposition 5.10. Soit X une variable aléatoire dans L^1 . Alors m est une médiane de X si et seulement si

$$\mathbb{E}[|X - m|] = \inf_{c \in \mathbb{R}} \mathbb{E}[|X - c|],$$

c'est-à-dire si m réalise la distance pour la norme L^1 de X à l'espace des fonctions constantes.

Démonstration. Notons que la fonction $\phi : c \mapsto \mathbb{E}[|X - c|]$ est une fonction convexe sur \mathbb{R} , et comme $\mathbb{E}[|X - c|] \geq |c| - \mathbb{E}[|X|]$ cette fonction tend vers $+\infty$ quand $|c| \rightarrow +\infty$. Par conséquent, elle atteint son minimum sur un intervalle $[a, b]$.

Si l'on dérive cette fonction formellement au point c , on obtient $\mathbb{E}[\mathbf{1}_{\{X < c\}} - \mathbf{1}_{\{X > c\}}] = \mathbb{P}(X < c) - \mathbb{P}(X > c)$. Soyons plus précis : on a pour tout $h \neq 0$, et $x, c \in \mathbb{R}$,

$$\left| \frac{|x - (c + h)| - |x - c|}{h} \right| \leq 1$$

et l'accroissement $(|x - (c + h)| - |x - c|)/h$ tend vers $\mathbf{1}_{\{x < c\}} - \mathbf{1}_{\{x \geq c\}}$ lorsque $h \downarrow 0$, et vers $\mathbf{1}_{\{x \leq c\}} - \mathbf{1}_{\{x > c\}}$ lorsque $h \uparrow 0$. On en déduit par la convergence dominée que ϕ est dérivable à gauche et à droite en c , de dérivées

$$\phi'_g(c) = \mathbb{P}(X < c) - \mathbb{P}(X \geq c) \quad \text{et} \quad \phi'_d(c) = \mathbb{P}(X \leq c) - \mathbb{P}(X > c)$$

Ces fonctions sont bien sûr croissantes, vérifient $\phi'_g \leq \phi'_d$, et elles sont non nulles si c est un α -quantile de X avec $\alpha \neq 1/2$. Si c est une médiane de X on a $\phi'_g(c) \leq 0 \leq \phi'_d(c)$. On conclut que le minimum de ϕ est atteint exactement aux médianes de X . \square

Toute variable aléatoire admet une médiane même si elle n'admet pas d'espérance, et même si c'est le cas, les deux notions peuvent donner des résultats très différents. Le résultat suivant montre de façon quantitative que si on considère des variables L^2 , alors espérance et médiane sont proches.

Proposition 5.11. *Si X est une variable aléatoire dans L^2 et m est une médiane de X , alors $|\mathbb{E}[X] - m| \leq \sqrt{\text{Var}(X)}$.*

Démonstration. Pour toute constante c , on a $\|X - c\|_1 \leq \|X - c\|_2$ par l'inégalité de Cauchy-Schwarz. Par conséquent, on obtient $\inf_{c \in \mathbb{R}} \|X - c\|_1 \leq \inf_{c \in \mathbb{R}} \|X - c\|_2$. À gauche, on reconnaît $\mathbb{E}[|X - m|]$ pour toute médiane m , et à droite, on reconnaît $\sqrt{\text{Var}(X)}$. On conclut par inégalité triangulaire. \square

5.3 Fonctions associées à une variable aléatoire

Nous allons maintenant associer à une variable aléatoire à valeurs scalaires ou vectorielles un certain nombre de fonctions qui caractérisent la loi des variables considérées

a. Fonction de répartition

Soit X une variable aléatoire à valeurs dans \mathbb{R} , définie sur un espace de probabilités $(\Omega, \mathcal{F}, \mathbb{P})$. On définit sa fonction de répartition par la formule

$$F_X(x) = \mathbb{P}(X \leq x), \quad x \in \mathbb{R}.$$

Par la propriété de continuité des mesures de probabilités par réunion croissante et intersection décroissante, on déduit que F_X est une fonction croissante, continue à droite. Plus précisément, la limite à gauche de F_X en un point $x \in \mathbb{R}$, notée $F_X(x -)$, est donnée par

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

ou autrement dit,

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

En particulier, la fonction F_X est également continue si et seulement si la loi de X est sans atome, puisque $\mathbb{P}(X = x) = \mathbb{P}_X(\{x\})$ par définition. En termes de théorie de la mesure, la loi \mathbb{P}_X de X n'est autre que la mesure de Stieltjes dF_X associée à la fonction F_X , c'est-à-dire l'unique mesure μ sur \mathbb{R} telle que $\mu(]a, b]) = F_X(b) - F_X(a)$ pour tout $a \leq b$. En particulier la fonction de répartition d'une variable aléatoire réelle X caractérise sa loi.

Proposition 5.12. *Soit X et X' deux variables aléatoires à valeurs dans \mathbb{R} , telles que $F_X = F_{X'}$. Alors X et X' ont la même loi.*

Noter que, dans l'énoncé précédent, comme dans ceux, similaires, qui sont à venir dans ce chapitre, on ne suppose pas que X et X' sont définies sur le même espace de probabilités $(\Omega, \mathcal{F}, \mathbb{P})$.

La fonction de répartition F_X est un outil pratique pour calculer des moments.

Exercice. Soit X une variable aléatoire positive, et $p \in [1, \infty[$, alors

$$\mathbb{E}[X^p] = \int_0^\infty px^{p-1} \mathbb{P}(X \geq x) dx.$$

Notons que la fonction $G_X(x) = 1 - F_X(x-) = \mathbb{P}(X \geq x)$ a déjà été considérée plus haut, sous le nom de la queue de distribution de X .

b. Fonction génératrice

Soit X une variable aléatoire à valeurs dans $\mathbb{N} = \{0, 1, 2, \dots\}$. On définit sa fonction génératrice comme la série entière

$$g_X(z) = \mathbb{E}[z^X] = \sum_{n \in \mathbb{N}} \mathbb{P}(X = n) z^n.$$

Comme par définition la somme des coefficients $\mathbb{P}(X = n)$ vaut 1, le rayon de convergence de cette série entière est supérieur ou égal à 1, et la formule ci-dessous est bien définie pour z un nombre complexe dans le disque fermé $\bar{\mathbb{D}} = \{z \in \mathbb{C} : |z| \leq 1\}$, et définit une fonction continue sur $\bar{\mathbb{D}}$. De plus, g_X est analytique sur le disque ouvert $\mathbb{D} = \{z \in \mathbb{C} : |z| < 1\}$ et l'on a

$$\mathbb{P}(X = n) = \frac{g_X^{(n)}(0)}{n!}, \quad n \in \mathbb{N}.$$

En particulier, on voit que la fonction génératrice caractérise la loi de X , puisqu'on retrouve à partir de ces quantités toutes les probabilités $\mathbb{P}(X \in A)$ avec $A \subset \mathbb{N}$.

Proposition 5.13. *Si X et X' sont deux variables aléatoires à valeurs dans \mathbb{N} telles que $g_X(z) = g_{X'}(z)$ pour tout $z \in [0, 1]$ (ou plus généralement pour tout z dans un sous-ensemble de \mathbb{D} ayant au moins un point d'accumulation), alors X et X' ont même loi, c'est-à-dire dans ce cas que $\mathbb{P}(X = n) = \mathbb{P}(X' = n)$ pour tout $n \in \mathbb{N}$.*

Démonstration. On utilise le fait que si les zéros d'une fonction analytique sur un ouvert connexe D ont un point d'accumulation dans D , alors cette fonction est nulle. \square

Un des intérêts de la fonction génératrice est son lien avec les moments de la variable aléatoire X . En effet, le théorème de dérivation sous le signe intégrale (ou le théorème de dérivation des séries entières) montre que pour tout $z \in \mathbb{D}$, on a

$$g'_X(z) = \mathbb{E}[Xz^{X-1}] = \sum_{n \geq 1} n \mathbb{P}(X = n) z^{n-1}.$$

Si l'on prend $z \in [0, 1[$ et que l'on fait tendre $z \nearrow 1$, le théorème de convergence monotone montre que

$$\mathbb{E}[X] = g'_X(1-),$$

et cette limite à gauche existe toujours (elle peut valoir $+\infty$). Plus généralement, on a le résultat suivant.

Proposition 5.14. *Pour tout $k \geq 0$, la limite à gauche de $g_X^{(k)}$ en 1 existe dans $[0, +\infty]$, et vaut*

$$g_X^{(k)}(1-) = \mathbb{E}[X(X-1)\dots(X-k+1)].$$

Démonstration. En dérivant k fois g_X en un point z de \mathbb{D} , on trouve

$$g_X^{(k)}(z) = \sum_{n \geq k} n(n-1)\dots(n-k+1)\mathbb{P}(X=n)z^{n-1},$$

et ceci converge vers la quantité voulue lorsque z converge vers 1 le long de $[0, 1[$. Noter que l'on aurait pu faire partir la somme de $n=0$ plutôt que de $n=k$, puisque les k premiers termes sont nuls : de même, dans l'énoncé, on peut invariablement ajouter l'indicatrice de l'événement $\{X \geq k\}$ dans l'espérance. \square

c. Fonction caractéristique

Fixons $d \geq 1$ un entier, et soit X une variable aléatoire à valeurs dans \mathbb{R}^d . La fonction caractéristique de X est définie par

$$\varphi_X(\xi) = \mathbb{E}[e^{i\langle \xi, X \rangle}], \quad \xi \in \mathbb{R}^d.$$

Par la formule de transfert, ceci n'est autre que

$$\varphi_X(\xi) = \int_{\mathbb{R}^d} e^{i\langle \xi, x \rangle} \mathbb{P}_X(dx) = (2\pi)^{d/2} \widehat{\mathbb{P}}_X(-\xi),$$

où $\widehat{\mathbb{P}}_X$ est la transformée de Fourier de la loi de X . La propriété d'injectivité de la transformée de Fourier sur les mesures signées implique (c'est bien le moins) que la fonction caractéristique d'une variable aléatoire caractérise la loi de cette variable.

Proposition 5.15. *Soit X et X' deux variables aléatoires à valeurs dans \mathbb{R}^d telles que $\varphi_X = \varphi_{X'}$. Alors X et X' ont même loi.*

Remarque. Attention, on doit bien supposer que $\varphi_X(\xi) = \varphi_{X'}(\xi)$ pour **tout** $\xi \in \mathbb{R}^d$. La situation est différente de celle pour les fonctions génératrices.

Les propriétés de la transformée de Fourier que nous avons étudiées impliquent que la régularité de la fonction caractéristique est intimement liée à l'existence de moments.

Proposition 5.16. *Soit X une variable aléatoire à valeurs dans \mathbb{R} . Si X admet un moment d'ordre k , alors φ_X est de classe $\mathcal{C}^k(\mathbb{R}, \mathbb{C})$, et l'on a*

$$\mathbb{E}[X^k] = (-i)^k \varphi_X^{(k)}(0).$$

Démonstration. Sous nos hypothèses, les théorèmes de dérivation sous le signe intégrale s'appliquent et donnent $\varphi_X^{(k)}(\xi) = \mathbb{E}[(iX)^k e^{i\xi X}]$. On peut aussi appliquer la formule de transfert et invoquer les résultats de dérivation des transformées de Fourier. \square

d. Transformée de Laplace

Soit X une variable aléatoire à valeurs dans \mathbb{R}_+ . On peut alors définir sa transformée de Laplace par la formule

$$L_X(\lambda) = \mathbb{E}[e^{-\lambda X}], \quad \lambda \geq 0.$$

La transformée de Laplace est étroitement liée à la fonction génératrice : si X est à valeurs dans \mathbb{N} , on a

$$L_X(\lambda) = g_X(e^{-\lambda}).$$

Proposition 5.17. *Si X et X' sont deux variables aléatoires positives telles que $L_X = L_{X'}$, alors X et X' ont la même loi.*

Démonstration. Soit \mathcal{A} l'ensemble des fonctions de $[0, \infty]$ dans \mathbb{R} de la forme $x \mapsto \sum_{i=1}^k a_i e^{-\lambda_i x}$, avec a_1, \dots, a_k dans \mathbb{R} et $\lambda_1, \dots, \lambda_k$ dans \mathbb{R}_+ . Alors \mathcal{A} est une algèbre de fonctions continues sur le compact $[0, \infty]$, séparant les points de ce compact. Le théorème de Stone-Weierstrass montre que \mathcal{A} est dense dans l'ensemble $\mathcal{C}([0, \infty], \mathbb{R})$ des fonctions continues sur \mathbb{R}_+ admettant une limite à l'infini, pour la norme uniforme. Par linéarité, si $L_X = L_{X'}$, alors on a $\mathbb{E}[f(X)] = \mathbb{E}[f(X')]$ pour tout $f \in \mathcal{A}$, et par densité, le même résultat est valable pour $f \in \mathcal{C}([0, \infty], \mathbb{R})$. Par un argument d'approximation, on en déduit que $\mathbb{P}(X \in I) = \mathbb{P}(X' \in I)$ pour tout intervalle ouvert $I \in \mathbb{R}_+$, c'est-à-dire que $\mathbb{P}_X(I) = \mathbb{P}_{X'}(I)$ et on conclut par le lemme de classe monotone. \square

Proposition 5.18. *La transformée de Laplace d'une variable aléatoire positive est une fonction continue sur \mathbb{R}_+ , et de classe $\mathcal{C}^\infty(]0, \infty[, \mathbb{R})$. Par ailleurs, on a pour tout entier $k \geq 0$,*

$$\mathbb{E}[X^k] = (-1)^k L_X^{(k)}(0+),$$

la limite à droite en 0 de $L_X^{(k)}$.

Démonstration. La dérivée k -ième de $\lambda \mapsto e^{-\lambda x}$ est $(-x)^k e^{-\lambda x}$, et pour tout intervalle compact $I \subset]0, \infty[$, on a

$$\sup \{x^k e^{-\lambda x} : x \in \mathbb{R}_+, \lambda \in I\} < \infty.$$

On peut donc appliquer le théorème de dérivation sous le signe intégrale et obtenir le caractère \mathcal{C}^∞ sur $]0, \infty[$. La continuité sur \mathbb{R}_+ est similaire, et utilise juste que $e^{-\lambda x} \leq 1$ pour tout $\lambda, x \geq 0$. Enfin, on déduit que pour tout $\lambda > 0$, on a

$$(-1)^k L_X^{(k)}(\lambda) = \mathbb{E}[X^k e^{-\lambda X}],$$

et on conclut en faisant tendre $\lambda \searrow 0$, et par convergence monotone. \square

Noter que la définition de la transformée de Laplace s'étend *verbatim* à tout nombre complexe λ tel que $\Re(\lambda) \geq 0$, et que $L_X(i\xi) = \varphi_X(\xi)$ est la fonction caractéristique de X . De plus, la preuve précédente montre que L_X est en fait holomorphe sur le demi-plan ouvert $\{\lambda \in \mathbb{C} : \Re(\lambda) > 0\}$. En particulier, ceci implique que pour vérifier que X et X' ont même loi, il suffit de montrer que $L_X(\lambda) = L_{X'}(\lambda)$ pour tout λ dans un sous-ensemble de \mathbb{R}_+ ayant au moins un point d'accumulation dans $]0, \infty[$.

On notera aussi que s'il existe $\varepsilon > 0$ tel que $\mathbb{E}[e^{\varepsilon X}] < \infty$ (on dit alors que X admet des moments exponentiels), alors la définition de la transformée de Laplace s'étend à tout $\lambda \in [-\varepsilon, \infty[$, et même^{5.1} à $\{\lambda \in \mathbb{C} : \Re(\lambda) \geq -\varepsilon\}$. Si c'est le cas, le développement en série entière de L_X au voisinage de 0 donne, du fait de la proposition 5.18,

$$L_X(\lambda) = \sum_{k \geq 0} \mathbb{E}[X^k] \frac{(-\lambda)^k}{k!}, \quad \lambda \in]-\varepsilon, \infty[.$$

On appelle de ce fait L_X la *fonction génératrice des moments*.

Corollaire 5.19. *Soit X une variable aléatoire positive. Supposons qu'il existe $\varepsilon > 0$ tel que $\mathbb{E}[e^{\varepsilon X}] < \infty$. Alors la suite $(\mathbb{E}[X^k], k \geq 1)$ des moments de X caractérise sa loi.*

En particulier, une loi de probabilités sur \mathbb{R} à support borné est caractérisée par ses moments.

Même si nous avons considéré ici des variables aléatoires positives, toutes les considérations précédentes s'étendent *mutatis mutandis* à des variables aléatoires réelles telles que $\mathbb{E}[e^{\lambda X}] < +\infty$ pour tout λ dans un intervalle $[a, b]$ contenant 0 et non réduit à un point, auquel cas $L_X(\lambda) = \mathbb{E}[e^{-\lambda X}]$ définit une fonction holomorphe dans la bande $\{\lambda \in \mathbb{C} : \Re(\lambda) \in]-b, -a[\}$.

5.4 Exemples fondamentaux de lois de variables aléatoires

Dans toute cette partie, X désignera une variable aléatoire définie sur un espace de probabilités $(\Omega, \mathcal{F}, \mathbb{P})$.

a. Lois discrètes

Loi uniforme sur un ensemble fini

Soit E un ensemble fini, alors $X: \Omega \rightarrow E$ est de loi uniforme sur E si

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

Ceci implique évidemment que $\mathbb{P}(X \in A) = \text{card}(A) / \text{card}(E)$ pour tout $A \subset E$.

Plus généralement, si E est un ensemble fini ou dénombrable et $(p_x, x \in E)$ est une famille de nombres réels positifs de somme 1, on lui associe une loi de probabilité par la formule

$$\mathbb{P}(X = x) = p_x,$$

c'est-à-dire que $\mathbb{P}(X \in A) = \sum_{x \in A} p_x$. Noter qu'il n'y a pas de loi uniforme sur un ensemble strictement dénombrable !

^{5.1.} Dans ce cas, la fonction caractéristique φ_X s'étend donc en une fonction holomorphe sur un domaine ouvert de \mathbb{C} contenant \mathbb{R} .

Loi de Bernoulli

Soit $p \in [0, 1]$, on dit que $X: \Omega \rightarrow \{0, 1\}$ suit une loi de Bernoulli de paramètre p (ou encore, que X est une variable aléatoire de Bernoulli de paramètre p) si

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

On a alors $\mathbb{E}[X] = p$, $\text{Var}(X) = p(1 - p)$, $g_X(z) = 1 - p + pz$. Pour $p = 1/2$, c'est la loi uniforme sur $\{0, 1\}$. Ceci modélise un lancer d'une pièce, biaisée si $p \neq 1/2$.

Loi binomiale

Soit $n \in \mathbb{N}$, $p \in [0, 1]$. On dit que $X: \Omega \rightarrow \{0, 1, 2, \dots, n\}$ suit une loi binomiale de paramètres (n, p) si

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

On a alors $\mathbb{E}[X] = np$, $\text{Var}(X) = np(1 - p)$, $g_X(z) = (1 - p + pz)^n$. Ceci correspond au nombre de pile lorsqu'on lance n fois une pièce biaisée, avec probabilité p d'obtenir pile. Formellement, c'est la loi de la variable aléatoire $X: \{0, 1\}^n \rightarrow \mathbb{N}$ définie par

$$X(\omega_1, \dots, \omega_n) = \omega_1 + \dots + \omega_n$$

sur l'espace de probabilités $(\{0, 1\}^n, 2^{\{0,1\}^n}, \text{Ber}(p)^{\otimes n})$ où $\text{Ber}(p) = p\delta_1 + (1 - p)\delta_0$ est la loi de Bernoulli.

Nous anticipons un peu sur le prochain chapitre en notant qu'une telle loi est obtenue en prenant la somme de n variables de Bernoulli de paramètre p indépendantes.

Loi géométrique

Soit $p \in]0, 1]$. On dit que $X: \Omega \rightarrow \mathbb{N}$ suit une loi géométrique de paramètre p si

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

On notera que les conventions diffèrent selon les ouvrages : notre choix est motivé par la définition suivante : X a la même loi que le premier temps de succès dans une suite de tirages de variables de Bernoulli indépendantes (un succès étant interprété comme le fait que la variable aléatoire prenne la valeur 1). En effet, si (X_1, X_2, \dots) est une suite de lancers de pile-ou-face biaisés, la variable aléatoire $T = \inf\{k \geq 1: X_k = 1\}$ a bien la loi voulue, puisque

$$\mathbb{P}(T = k) = \mathbb{P}(X_1 = 0, \dots, X_{k-1} = 0, X_k = 1) = (1 - p)^{k-1} p.$$

Noter que la variable aléatoire T est à valeurs dans $\mathbb{N} \cup \{\infty\}$, puisqu'elle prend la valeur ∞ sur la suite constante égale à 0. Cependant, cette suite n'est pas chargée par la mesure produit $\text{Ber}(p)^{\mathbb{N}^*}$, et on a donc que $\mathbb{P}(T < \infty) = 1$.

On a $\mathbb{E}[X] = 1/p$, $\text{Var}(X) = (1 - p)/p^2$,

$$g_X(z) = \frac{pz}{1 - (1 - p)z}.$$

Loi de Poisson

Soit $\theta \geq 0$. On dit que $X: \Omega \rightarrow \mathbb{N}$ suit une loi de Poisson de paramètre θ si

$$\mathbb{P}(X = k) = e^{-\theta} \frac{\theta^k}{k!}, \quad k \geq 0.$$

On a que $\mathbb{E}[X] = \theta$, $\text{Var}(X) = \theta$, $g_X(z) = \exp(-\theta(1 - z))$. Mentionnons dès maintenant qu'une variable de loi de Poisson de paramètre θ peut être vue comme une variable de Bernoulli de paramètres $n, \theta/n$ pour n très grand : une loi de Poisson correspond donc au nombre d'occurrence d'un phénomène très rare (de probabilité inversement proportionnelle au nombre d'expériences réalisées), et s'appelle parfois la *loi des événements rares*. En effet, pour tout k fixé, on a, pour tout $k \geq 0$ fixé,

$$\binom{n}{k} \left(\frac{\theta}{n}\right)^k \left(1 - \frac{\theta}{n}\right)^{n-k} \xrightarrow{n \rightarrow \infty} e^{-\theta} \frac{\theta^k}{k!}.$$

b. Loïs à densité

Loi uniforme sur un sous-ensemble mesurable de \mathbb{R}^d .

Soit $A \in \mathcal{B}(\mathbb{R}^d)$ un borélien tel que $0 < \lambda_d(A) < \infty$. On dit que la variable aléatoire $X: \Omega \rightarrow \mathbb{R}^d$ est de loi uniforme sur A si

$$\mathbb{P}_X(dx) = \frac{dx}{\lambda_d(A)} \mathbf{1}_A(x).$$

Par exemple, si $d = 1$ et $A = [a, b]$ est un intervalle avec $a < b$, la loi uniforme est $dx \mathbf{1}_{[a,b]}(x) / (b - a)$. La fonction de répartition associée est

$$F_X(x) = 0 \vee \frac{x - a}{b - a} \wedge 1, \quad x \in \mathbb{R},$$

où nous notons $x \vee y = \max\{x, y\}$ et $x \wedge y = \min\{x, y\}$.

La fonction caractéristique est donnée par

$$\varphi_X(\xi) = \frac{e^{ib\xi} - e^{ia\xi}}{i\xi(b-a)} = e^{i\xi \frac{a+b}{2}} \cdot \frac{\sin((b-a)\xi)}{(b-a)\xi}, \quad \xi \in \mathbb{R} \setminus \{0\}, \quad \varphi_X(0) = 1.$$

Loïs exponentielles

Soit $\theta > 0$. La variable aléatoire $X: \Omega \rightarrow \mathbb{R}$ suit la loi exponentielle de paramètre θ si

$$\mathbb{P}_X(dx) = \theta e^{-\theta x} \mathbf{1}_{\mathbb{R}_+}(x) dx.$$

La fonction de répartition est $F_X(x) = 1 - e^{-\theta x}$ pour tout $x \geq 0$, et $F_X(x) = 0$ si $x < 0$. On travaille plutôt avec la queue de distribution $G_X(x) = e^{-\theta x}$.

La propriété fondamentale des variables exponentielles est *l'absence de mémoire* : si X a une loi exponentielle alors $G_X(x + y) = G_X(x)G_X(y)$ pour tout $x, y \geq 0$, ce qui se traduit par

$$\mathbb{P}(X > x + y) = \mathbb{P}(X > x)\mathbb{P}(X > y)$$

ou encore par

$$\mathbb{P}(X > x + y | X > x) = \mathbb{P}(X > y).$$

On parle également de « non-vieillesse » de la loi exponentielle. Cette propriété est caractéristique des lois exponentielles (si l'on accepte en plus la masse de Dirac en 0 comme loi exponentielle de paramètre $+\infty$). La fonction caractéristique de X et la transformée de Laplace sont données par

$$\varphi_X(\xi) = \frac{\theta}{\theta - i\xi}, \quad L_X(\lambda) = \frac{\theta}{\theta + \lambda}.$$

Lois gaussiennes sur \mathbb{R} .

On a vu que la densité gaussienne $g_\sigma(x) = \exp(-x^2/2\sigma^2) / (2\pi\sigma^2)^{1/2}$ est une fonction positive d'intégrale 1 par rapport à la mesure de Lebesgue sur \mathbb{R} (on parle de densité de probabilité). On dit que la variable aléatoire $X: \Omega \rightarrow \mathbb{R}$ admet une loi gaussienne de moyenne m et de variance σ^2 si sa loi est donnée par

$$\mathbb{P}_X(dx) = g_\sigma(x - m)dx.$$

Comme on l'a vu au chapitre précédent, sa fonction caractéristique est donnée par

$$\varphi_X(\xi) = \exp\left(im\xi - \frac{\sigma^2|\xi|^2}{2}\right).$$

En dérivant, on en déduit que l'espérance de X est m , et sa variance est σ^2 , ce qui est cohérent avec la définition.

Il est par ailleurs facile de constater que $\mathbb{E}[e^{\lambda X}] < \infty$ for every $\lambda \in \mathbb{R}$, puisque $e^{-x^2/2\sigma^2}$ décroît bien plus vite à l'infini que e^{-ax} , pour tout $a > 0$. Donc la transformée de Laplace existe bien (au sens étendu que nous avons donné à la fin du paragraphe sur cette transformée), et vaut

$$L_X(\lambda) = \exp\left(-\lambda m + \frac{\sigma^2\lambda^2}{2}\right).$$

On parlera plus tard, au paragraphe 8.3, de la famille des lois gaussiennes sur \mathbb{R}^d .

Chapitre 6

Indépendance

Dans tout ce chapitre, on fixe l'espace de probabilités $(\Omega, \mathcal{F}, \mathbb{P})$.

6.1 Probabilités conditionnelles élémentaires

Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace de probabilités, et $B \in \mathcal{F}$ un événement tel que $\mathbb{P}(B) > 0$. On définit alors, pour tout $A \in \mathcal{F}$,

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

et on l'appelle probabilité de A sachant B . Comme $\mathbb{P}(\Omega|B) = \mathbb{P}(\Omega) / \mathbb{P}(B) = 1$, on obtient que l'application $A \mapsto \mathbb{P}(A|B)$ est une mesure de probabilités. Intuitivement, l'espace de probabilités $(\Omega, \mathcal{F}, \mathbb{P}(\cdot|B))$ est l'espace correspondant à une expérience aléatoire pour laquelle on sait *a priori* que l'événement B est réalisé. Si A et B sont tous deux des événements tels que $\mathbb{P}(A)\mathbb{P}(B) > 0$, alors on obtient facilement la *formule de Bayes*

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

Soit I un ensemble d'indices fini ou dénombrable. Si $(B_i, i \in I)$ est une partition mesurable de Ω , c'est-à-dire que les ensembles B_i sont des événements deux-à-deux disjoints et de réunion Ω , alors pour tout événement A , on a la *formule des probabilités totales*

$$\mathbb{P}(A) = \sum_{i \in I} \mathbb{P}(A|B_i)\mathbb{P}(B_i),$$

où l'on pose par convention $\mathbb{P}(A|B_i) = 0$ si $\mathbb{P}(B_i) = 0$. Cette formule est également aisée à démontrer. À l'aide de cette formule, on peut réécrire la formule de Bayes sous la forme

$$\mathbb{P}(B|A) = \frac{\mathbb{P}(A|B)\mathbb{P}(B)}{\mathbb{P}(A|B)\mathbb{P}(B) + \mathbb{P}(A|B^c)\mathbb{P}(B^c)}.$$

Exemple. Les probabilités conditionnelles sont réputées donner des résultats parfois inattendus. En voici un exemple. Supposons qu'une certaine maladie frappe un individu sur 100, c'est-à-dire qu'un individu donné a une probabilité 0,01 d'en être affecté. On suppose que l'on dispose d'un test de dépistage de la maladie, mais qu'il n'est fiable qu'à 99%, c'est-à-dire qu'il a une probabilité 0,01 de donner un résultat positif quand on l'utilise sur un individu sain (faux positif), et une probabilité 0,01 de donner un résultat négatif quand on l'utilise sur un individu malade. Supposons qu'un individu donné soit testé positif. Quelle est la probabilité qu'il soit effectivement malade ?

En notant \oplus et \ominus les événements « être diagnostiqué » positif/négatif, et par M , S les événements « être malade/sain », les données du problème sont

$$\mathbb{P}(M) = \frac{1}{100} = 1 - \mathbb{P}(S), \quad \mathbb{P}(\oplus|M) = \frac{99}{100}, \quad \mathbb{P}(\oplus|S) = \frac{1}{100}.$$

On cherche à calculer $\mathbb{P}(M|\oplus)$, et la formule de Bayes donne

$$\mathbb{P}(M|\oplus) = \frac{\mathbb{P}(\oplus|M)\mathbb{P}(M)}{\mathbb{P}(\oplus|M)\mathbb{P}(M) + \mathbb{P}(\oplus|S)\mathbb{P}(S)} = \frac{\frac{99}{100} \cdot \frac{1}{100}}{\frac{99}{100} \cdot \frac{1}{100} + \frac{1}{100} \cdot \frac{99}{100}} = \frac{1}{2}.$$

En y réfléchissant un peu, comme peu d'individus sont effectivement malades, si toute la population fait le test, il y aura clairement beaucoup plus de faux positifs que de faux négatifs ! Ceci illustre le fait que pour qu'un test soit efficace, il vaut mieux qu'il soit pratiqué *a priori* sur une population considérée comme « à risque ».

6.2 Indépendance d'événements

Soit $A, B \in \mathcal{F}$ deux événements. On dit que A et B sont *indépendants* si

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

Autrement dit, si de plus $\mathbb{P}(B) > 0$, ceci signifie que $\mathbb{P}(A|B) = \mathbb{P}(A)$, c'est-à-dire que l'information donnée par B n'a aucune influence sur la probabilité que A ait lieu. Plus généralement, si A_1, A_2, \dots, A_n sont des événements, on dit qu'ils sont indépendants si pour tout $I \subset \{1, 2, \dots, n\}$, on a

$$\mathbb{P}\left(\bigcap_{i \in I} A_i\right) = \prod_{i \in I} \mathbb{P}(A_i). \quad (6.1)$$

Il convient de faire attention ici :

- Si les événements (A_1, \dots, A_n) sont indépendants, alors ils sont aussi indépendants deux-à-deux (prendre pour I les paires d'éléments de $\{1, \dots, n\}$), mais la réciproque n'est pas vraie. Par exemple, si l'on jette deux pièces équilibrées, et qu'on note $\omega_1, \omega_2 \in \{0, 1\}$ les résultats (0 face, 1 pile), alors les événements $\{\omega_1 = 0\}, \{\omega_2 = 0\}, \{\omega_1 = \omega_2\}$ sont indépendants deux-à-deux, mais pas indépendants. On parle parfois « d'indépendance dans leur ensemble » des événements (A_1, \dots, A_n) pour insister sur ce point.
- Dans la définition, il ne suffit pas de vérifier $\mathbb{P}(A_1 \cap \dots \cap A_n) = \mathbb{P}(A_1) \cdots \mathbb{P}(A_n)$, puisque par exemple on peut avoir $A_1 = \emptyset$.

La définition précédente s'étend à une famille quelconque $(A_j, j \in J)$ d'événements : on dit que ces événements sont indépendants si l'on a (6.1) pour tout sous-ensemble $I \subset J$ fini.

Remarque. Si A_1, A_2, \dots, A_n sont des événements indépendants, alors A_1^c, A_2, \dots, A_n sont également indépendants. En effet, pour tout $1 < i_2 < \dots < i_k \leq n$, on a

$$\begin{aligned} \mathbb{P}(A_1^c \cap A_{i_2} \cap \dots \cap A_{i_k}) &= \mathbb{P}(A_{i_2} \cap \dots \cap A_{i_k}) - \mathbb{P}(A_1 \cap A_{i_2} \cap \dots \cap A_{i_k}) \\ &= (1 - \mathbb{P}(A_1)) \prod_{j=2}^k \mathbb{P}(A_{i_j}) \\ &= \mathbb{P}(A_1^c) \mathbb{P}(A_{i_2}) \cdots \mathbb{P}(A_{i_k}). \end{aligned}$$

Cela suffit clairement pour conclure.

6.3 Indépendance de σ -algèbres

Rappelons que si \mathcal{C} est une classe de sous-ensembles d'un même ensemble Ω , on note $\sigma(\mathcal{C})$ la plus petite σ -algèbre contenant \mathcal{C} . Par ailleurs, si $(\mathcal{F}_i, i \in I)$ est une famille quelconque de sous- σ -algèbres d'une même σ -algèbre \mathcal{F} , on note

$$\bigvee_{i \in I} \mathcal{F}_i = \sigma\left(\bigcup_{i \in I} \mathcal{F}_i\right)$$

la plus petite σ -algèbre contenant toutes les $\mathcal{F}_i, i \in I$.

Soit $\mathcal{F}_1, \dots, \mathcal{F}_n$ des sous- σ -algèbres de \mathcal{F} . On dit qu'elles sont indépendantes si et seulement si l'on a

$$\mathbb{P}\left(\bigcap_{i=1}^n A_i\right) = \prod_{i=1}^n \mathbb{P}(A_i), \quad \text{pour tout } A_1 \in \mathcal{F}_1, \dots, A_n \in \mathcal{F}_n.$$

Proposition 6.1. *Les σ -algèbres $\mathcal{F}_1, \dots, \mathcal{F}_n$ sont indépendantes si et seulement si pour tout $A_1 \in \mathcal{F}_1, \dots, A_n \in \mathcal{F}_n$, les événements A_1, \dots, A_n sont indépendants.*

Démonstration. Le sens direct est le seul sens non trivial à démontrer. Supposons donc que les σ -algèbres $\mathcal{F}_1, \dots, \mathcal{F}_n$ sont indépendantes et prenons $A_1 \in \mathcal{F}_1, \dots, A_n \in \mathcal{F}_n$. Soit $I \subset \{1, \dots, n\}$, posons $B_i = A_i$ si $i \in I$ et $B_i = \Omega$ si $i \notin I$. Alors pour tout $i \in \{1, \dots, n\}$ on a $B_i \in \mathcal{F}_i$, et la définition de l'indépendance de σ -algèbres implique que

$$\mathbb{P}\left(\bigcap_{i \in I} A_i\right) = \mathbb{P}\left(\bigcap_{i=1}^n B_i\right) = \prod_{i=1}^n \mathbb{P}(B_i) = \prod_{i \in I} \mathbb{P}(A_i).$$

Ceci montre que A_1, \dots, A_n sont indépendants. □

Exercice. Montrer que A_1, \dots, A_n sont indépendants si et seulement si les σ -algèbres $\sigma(\{A_1\}), \dots, \sigma(\{A_n\})$ sont indépendantes, où $\sigma(\{A\}) = \{\emptyset, A, A^c, \Omega\}$ est la plus petite σ -algèbre contenant A .

Plus généralement, si $(\mathcal{F}_i, i \in I)$ est une famille quelconque de σ -algèbres, on dit qu'elles sont indépendantes si les σ -algèbres $(\mathcal{F}_j, j \in J)$ sont indépendantes pour toute partie finie $J \subset I$.

En pratique, il n'est pas nécessaire de vérifier la formule de factorisation ci-dessus pour tous les événements, comme le montre le résultat suivant.

Lemme 6.2. *Soit $\mathcal{C}_1, \dots, \mathcal{C}_n$ des sous-ensembles des σ -algèbres $\mathcal{F}_1, \dots, \mathcal{F}_n$, stables par intersection finie, contenant Ω , et tels que $\sigma(\mathcal{C}_i) = \mathcal{F}_i$ pour tout $i \in \{1, 2, \dots, n\}$. On suppose que pour tout choix de $C_i \in \mathcal{C}_i, i \in \{1, 2, \dots, n\}$ on a*

$$\mathbb{P}\left(\bigcap_{i=1}^n C_i\right) = \prod_{i=1}^n \mathbb{P}(C_i).$$

Alors les σ -algèbres $\mathcal{F}_1, \dots, \mathcal{F}_n$ sont indépendantes.

Démonstration. Fixons C_2, \dots, C_n dans leurs classes $\mathcal{C}_2, \dots, \mathcal{C}_n$ respectives, et notons

$$\mathcal{M}_1 = \{A \in \mathcal{F}_1 : \mathbb{P}(A \cap C_2 \cap \dots \cap C_n) = \mathbb{P}(A)\mathbb{P}(C_2)\dots\mathbb{P}(C_n)\}.$$

Alors \mathcal{M}_1 contient \mathcal{C}_1 par hypothèse (et donc $\Omega \in \mathcal{M}_1$ en particulier), et est une classe monotone : elle contient \emptyset , est stable par réunion dénombrable disjointe, et enfin par complémentaire, puisque $\Omega \cap C_2 \cap \dots \cap C_n \setminus A \cap C_1 \cap \dots \cap C_n = A^c \cap C_2 \cap \dots \cap C_n$, et $\mathbb{P}(A^c) = \mathbb{P}(\Omega) - \mathbb{P}(A)$. Par le lemme de classe monotone, on a donc $\mathcal{M}_1 = \sigma(\mathcal{C}_1) = \mathcal{F}_1$. Ensuite, on fixe $A_1 \in \mathcal{F}_1$ et C_3, \dots, C_n dans leurs classes respectives $\mathcal{C}_3, \dots, \mathcal{C}_n$, et on note

$$\mathcal{M}_2 = \{A \in \mathcal{F}_2 : \mathbb{P}(A_1 \cap A \cap C_3 \cap \dots \cap C_n) = \mathbb{P}(A_1)\mathbb{P}(A)\mathbb{P}(C_3)\dots\mathbb{P}(C_n)\}.$$

Pour les mêmes raisons que ci-dessus, on a que $\mathcal{M}_2 = \mathcal{F}_2$. En procédant ainsi par récurrence, on obtient que

$$\mathbb{P}\left(\bigcap_{i=1}^n A_i\right) = \prod_{i=1}^n \mathbb{P}(A_i), \quad \text{pour tout } A_1 \in \mathcal{F}_1, \dots, A_n \in \mathcal{F}_n,$$

ce qu'on voulait démontrer. □

On en déduit le lemme de « regroupement par paquets ».

Lemme 6.3. *Soit $(\mathcal{F}_i, i \in I)$ une famille de σ -algèbres indépendantes, et $(I_j, j \in J)$ une partition de I . Alors les σ -algèbres $(\bigvee_{i \in I_j} \mathcal{F}_i, j \in J)$ sont indépendantes.*

Démonstration. Soit $K \subset J$ un sous-ensemble fini. Il suffit de montrer que les σ -algèbres $\bigvee_{i \in I_j} \mathcal{F}_i, j \in K$ sont indépendantes. Pour cela, on définit \mathcal{C}_j comme l'ensemble des intersections finies d'éléments de $\bigcup_{i \in I_j} \mathcal{F}_i$ pour tout $j \in K$, et on applique le lemme 6.2. En effet, un élément C_j de \mathcal{C}_j s'écrit sous la forme $A_1^j \cap \dots \cap A_{k(j)}^j$ où les événements A_m^j sont dans des σ -algèbres \mathcal{F}_{i_m} distinctes, avec $i_m \in I_j$. En choisissant les C_j de cette forme, la propriété d'indépendance de $(\mathcal{F}_i, i \in I)$ implique clairement que

$$\mathbb{P}\left(\bigcap_{j \in K} C_j\right) = \prod_{j \in K} \prod_{m=1}^{k(j)} \mathbb{P}(A_m^j) = \prod_{j \in K} \mathbb{P}(C_j).$$

On conclut par le lemme 6.2. □

6.4 Indépendance de variables aléatoires

a. σ -algèbre associée à une variable aléatoire

À toute variable aléatoire X à valeurs dans un espace mesurable (E, \mathcal{E}) , on associe une σ -algèbre (tribu)

$$\sigma(X) = \{X^{-1}(A) : A \in \mathcal{E}\},$$

qui est une sous- σ -algèbre de \mathcal{F} , appelée la σ -algèbre engendrée par la variable aléatoire X . On interprète la σ -algèbre $\sigma(X)$ comme l'information contenue dans la variable X , ou encore les événements mesurables par rapport à X .

Exercice. Soit Y une variable aléatoire à valeurs dans un espace mesurable (E, \mathcal{E}) , et soit X une variable aléatoire à valeurs dans \mathbb{R} telle que X est mesurable par rapport à $\sigma(Y)$. Alors il existe une fonction mesurable f de (E, \mathcal{E}) dans $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ telle que $X = f(Y)$. [Pour montrer cela on peut commencer par supposer que X est une fonction indicatrice, et on rappelle qu'une fonction mesurable positive est limite croissante de fonctions étagées presque partout.]

Plus généralement, si $(X_i, i \in I)$ est une famille quelconque de variables aléatoires, on lui associe la tribu engendrée par cette famille comme

$$\sigma(X_i, i \in I) = \bigvee_{i \in I} \sigma(X_i).$$

b. Indépendance de variables aléatoires

Définition 6.4. Soit X_1, \dots, X_n des variables aléatoires, respectivement à valeurs dans les espaces mesurables (E_i, \mathcal{E}_i) , $1 \leq i \leq n$. On dit que ces variables aléatoires sont indépendantes si les tribus $\sigma(X_i)$, $1 \leq i \leq n$ sont indépendantes. Ceci signifie que pour tout choix d'ensembles mesurables $A_i \in \mathcal{E}_i$, $1 \leq i \leq n$, on a

$$\mathbb{P}(X_i \in A_i, 1 \leq i \leq n) = \prod_{i=1}^n \mathbb{P}(X_i \in A_i). \quad (6.2)$$

Plus généralement, on dit que les variables aléatoires d'une famille quelconque $(X_i, i \in I)$ sont indépendantes si toutes les sous-familles finies $(X_j, j \in J)$ avec J fini inclus dans I sont formées de variables aléatoires indépendantes.

Remarque. Si les σ -algèbres $(\mathcal{F}_i, i \in I)$ sont indépendantes, et si les variables aléatoires $(X_i, i \in I)$ sont telles que, pour tout $i \in I$, X_i est mesurable par rapport à \mathcal{F}_i , alors les variables aléatoires $(X_i, i \in I)$ sont indépendantes.

Proposition 6.5. Soit X_1, X_2, \dots, X_n des variables aléatoires, respectivement à valeurs dans des espaces mesurables $(E_1, \mathcal{E}_1), (E_2, \mathcal{E}_2), \dots, (E_n, \mathcal{E}_n)$. Ces variables aléatoires sont indépendantes si et seulement si la loi de (X_1, X_2, \dots, X_n) est la loi produit des marginales :

$$\mathbb{P}_{(X_1, \dots, X_n)} = \mathbb{P}_{X_1} \otimes \mathbb{P}_{X_2} \otimes \dots \otimes \mathbb{P}_{X_n}.$$

Démonstration. Supposons que X_1, X_2, \dots, X_n sont indépendantes. Par la définition de l'indépendance, les deux mesures de probabilités apparaissant dans l'énoncé de part et d'autre de l'égalité sont égales sur les pavés mesurables de $E_1 \times E_2 \times \dots \times E_n$. Elles sont donc égales partout par le lemme des classes monotones. La réciproque est immédiate. \square

Exemple. On a construit au chapitre 5.1 une suite infinie de variables aléatoires indépendantes uniformes sur $\{0, 1\}$, à l'aide de l'écriture dyadique d'une variable aléatoire de loi uniforme sur $[0, 1[$.

Corollaire 6.6. Soit X_1, \dots, X_n des variables aléatoires indépendantes, respectivement à valeurs dans un espace mesurable (E_i, \mathcal{E}_i) , et pour tout $i \in \{1, \dots, n\}$, soit $f_i: E_i \rightarrow \mathbb{R}$ une fonction mesurable. On suppose, ou bien que toutes les fonctions f_i sont positives, ou bien que $f_i \in L^1(\mathbb{P}_{X_i})$ pour tout $i \in \{1, \dots, n\}$. Alors

$$\mathbb{E} \left[\prod_{i=1}^n f_i(X_i) \right] = \prod_{i=1}^n \mathbb{E}[f_i(X_i)].$$

En particulier, si $f_i \in L^1(\mathbb{P}_{X_i})$ pour tout $i \in \{1, \dots, n\}$, alors $\prod_{i=1}^n f_i(X_i) \in L^1(\mathbb{P})$.

Par exemple, soit X_1, \dots, X_n des variables aléatoires à valeurs réelles, intégrables, et indépendantes, alors $X_1 \cdots X_n$ est aussi intégrable et $\mathbb{E}[X_1 \cdots X_n] = \mathbb{E}[X_1] \cdots \mathbb{E}[X_n]$.

Remarque. Soit X, Y deux variables aléatoires indépendantes et dans L^2 . Alors les variables aléatoires $X - \mathbb{E}[X]$ et $Y - \mathbb{E}[Y]$ sont indépendantes et dans L^2 , et l'on a

$$\text{Cov}(X, Y) = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] = 0.$$

On dit que des variables aléatoires indépendantes sont *décorrélées* (de corrélation nulle). La réciproque n'est pas vraie. Si (ε, X) sont deux variables aléatoires indépendantes, où $\mathbb{P}(\varepsilon = 1) = \mathbb{P}(\varepsilon = -1) = 1/2$ et où X est une variable aléatoire gaussienne $\mathcal{N}(0, 1)$, alors les variables aléatoires X et εX sont décorréliées. En effet, on a facilement $\mathbb{E}[X] = \mathbb{E}[\varepsilon X] = 0$, et

$$\text{Cov}(X, \varepsilon X) = \mathbb{E}[\varepsilon X^2] = \mathbb{E}[\varepsilon] \mathbb{E}[X^2] = 0.$$

Cependant, les variables aléatoires X et εX ne sont clairement pas indépendantes, sinon la loi de $(X, \varepsilon X)$ serait à densité sur \mathbb{R}^2 . Or il est clair que la loi de $(X, \varepsilon X)$ est en fait portée par les bissectrices $\{(x, x) : x \in \mathbb{R}\} \cup \{(x, -x) : x \in \mathbb{R}\}$.

c. Critères d'indépendance de variables aléatoires

Proposition 6.7. Soit X_1, \dots, X_n des variables aléatoires discrètes, à valeurs dans des ensembles E_1, \dots, E_n . Alors ces variables sont indépendantes si et seulement si l'on a, pour tout (x_1, \dots, x_n) dans $E_1 \times \dots \times E_n$

$$\mathbb{P}(X_1 = x_1, \dots, X_n = x_n) = \mathbb{P}(X_1 = x_1) \cdots \mathbb{P}(X_n = x_n).$$

La preuve de ce résultat est évidente.

Proposition 6.8. *Soit X_1, \dots, X_n des variables aléatoires à valeurs dans \mathbb{R} . Alors elles sont indépendantes si et seulement si l'on a, pour tout $x_1, \dots, x_n \in \mathbb{R}$,*

$$\mathbb{P}(X_1 \leq x_1, \dots, X_n \leq x_n) = \prod_{i=1}^n F_{X_i}(x_i).$$

Démonstration. Pour le voir, il suffit d'appliquer le lemme 6.2 en prenant pour \mathcal{C}_i la classe des intervalles de la forme $]-\infty, x]$, ou égaux à \mathbb{R} tout entier. Noter que la factorisation ci-dessus a encore lieu si l'on prend certains $x_i = \infty$, par la continuité des mesures de probabilités par limite monotone. La réciproque est évidente. \square

Proposition 6.9. *Soit X_1, \dots, X_n sont des variables aléatoires réelles, alors elles sont indépendantes si et seulement si pour tout $\xi = (\xi_1, \dots, \xi_n) \in \mathbb{R}^n$, on a*

$$\varphi_{(X_1, \dots, X_n)}(\xi) = \prod_{j=1}^n \varphi_{X_j}(\xi_j).$$

Ceci est une application immédiate du lemme 6.6 et de l'injectivité de la transformée de Fourier.

Exemple : variables aléatoires gaussiennes. Soit $\sigma > 0$ et $m = (m_1, \dots, m_d) \in \mathbb{R}^d$. On dit que la variable aléatoire $X = (X_1, \dots, X_d)$ à valeurs dans \mathbb{R}^d est gaussienne de moyenne m et de matrice de variance-covariance $\sigma^2 I_d$ si la loi de X a pour densité $g_\sigma(x - m)$ par rapport à λ_d , où $g_\sigma(x) = (2\pi\sigma^2)^{-d/2} \exp(-|x|^2/2\sigma^2)$. On note généralement $\mathcal{N}(m, \sigma^2 I_d)$ cette loi. Par les propriétés de la transformée de Fourier des densités gaussiennes, on a que

$$\varphi_X(\xi) = \exp\left(im\xi - \frac{\sigma^2|\xi|^2}{2}\right) = \prod_{j=1}^d \exp\left(im_j \xi_j - \frac{\sigma^2 \xi_j^2}{2}\right). \quad (6.3)$$

En prenant $\xi = \xi' e_j$, où $\xi' \in \mathbb{R}$ et où e_j est le j -ème vecteur de la base canonique de \mathbb{R}^d , on en déduit en particulier que

$$\varphi_{X_j}(\xi') = \exp\left(im_j \xi' - \frac{\sigma^2(\xi')^2}{2}\right),$$

c'est-à-dire que X_j suit la loi gaussienne $\mathcal{N}(m_j, \sigma^2)$ sur \mathbb{R} . De plus, la formule 6.3 implique que les variables aléatoires X_1, \dots, X_d sont indépendantes.

Remarquons enfin que l'espérance de X est bien égale à m , et sa matrice de variance-covariance est $\Sigma_X = \sigma^2 I_d$. Ceci peut s'obtenir en dérivant la fonction caractéristique, et nous laissons la vérification en exercice. Pour simplifier, noter que X suit la loi $\mathcal{N}(m, \sigma^2 I_d)$ si et seulement si $X - m$ suit la loi $\mathcal{N}(0, \sigma^2 I_d)$. En effet, pour toute fonction $f: \mathbb{R}^d \rightarrow \mathbb{R}$ mesurable bornée,

$$\mathbb{E}[f(X - m)] = \int_{\mathbb{R}^d} f(x - m) g_\sigma(x - m) dx = \int_{\mathbb{R}^d} f(x) g_\sigma(x) dx,$$

et la réciproque est similaire. De ce fait, à une translation près par un vecteur de \mathbb{R}^d , on peut se ramener à l'étude des lois $\mathcal{N}(0, \sigma^2 I_d)$.

Enfin, voici un dernier exemple de critère d'indépendance pratique pour des variables aléatoires à densité.

Proposition 6.10. Soit $X = (X_1, \dots, X_d)$ une variable aléatoire à valeurs dans \mathbb{R}^d . On suppose que la loi de X admet une densité f_X , et qu'il existe des fonctions mesurables $f_i: \mathbb{R} \rightarrow \mathbb{R}_+$ pour $1 \leq i \leq d$ telles que

$$f_X(x) = \prod_{i=1}^d f_i(x_i).$$

Alors les variables aléatoires X_1, \dots, X_d sont indépendantes, et il existe des nombres réels $c_i > 0$, $1 \leq i \leq d$ tels que $f_{X_i} = c_i f_i$, où f_{X_i} est la densité de la loi de X_i .

Rappelons que sous nos hypothèses, les variables X_i sont bien à densité, du fait de la discussion suivant la proposition 5.5.

Démonstration. Soit $h_1, \dots, h_d: \mathbb{R} \rightarrow \mathbb{R}$ des fonctions mesurables bornées. Alors on a

$$\mathbb{E} \left[\prod_{i=1}^d h_i(X_i) \right] = \int_{\mathbb{R}^d} \prod_{i=1}^d h_i(x_i) f_i(x_i) dx_i = \prod_{i=1}^d \int_{\mathbb{R}^d} h_i(y) f_i(y) dy,$$

par le théorème de Fubini. Posons $c_i = \left(\int_{\mathbb{R}} f_i(y) dy \right)^{-1}$. En prenant toutes les fonctions $h_i \equiv 1$ dans l'équation précédente, on obtient $c_1 \cdots c_d = 1$, et en particulier, les nombres c_i sont tous strictement positifs et finis. On réécrit donc l'équation précédente sous la forme

$$\mathbb{E} \left[\prod_{i=1}^d h_i(X_i) \right] = \prod_{i=1}^d \int_{\mathbb{R}^d} h_i(y) c_i f_i(y) dy.$$

En prenant toutes les fonctions $h_i \equiv 1$ sauf une (disons h_j), on déduit de ceci que la loi de X_j admet pour densité la fonction $c_j f_j$. Enfin, on a obtenu que

$$\mathbb{E} \left[\prod_{i=1}^d h_i(X_i) \right] = \prod_{i=1}^d \mathbb{E}[h_i(X_i)],$$

pour toutes les fonctions h_i mesurables bornées, ce qui montre l'indépendance des variables aléatoires X_1, \dots, X_d . \square

Comme exemple d'application, on peut montrer une nouvelle fois que les composantes d'une variable aléatoire gaussienne $\mathcal{N}(m, \sigma^2 I_d)$ sont indépendantes.

6.5 Sommes de variables aléatoires indépendantes

Soit X et Y deux variables aléatoires à valeurs dans \mathbb{R}^d , définies sur un espace de probabilités $(\Omega, \mathcal{F}, \mathbb{P})$. On suppose X et Y indépendantes.

Lemme 6.11. La loi de la variable aléatoire $X + Y$ est la convolée $\mathbb{P}_X * \mathbb{P}_Y$.

Démonstration. Soit f une fonction mesurable bornée. Comme X et Y sont indépendantes, la loi de (X, Y) est la mesure $\mathbb{P}_X \otimes \mathbb{P}_Y$, et donc

$$\mathbb{E}[f(X + Y)] = \int_{\mathbb{R}^d \times \mathbb{R}^d} f(x + y) \mathbb{P}_X(dx) \mathbb{P}_Y(dy) = \int_{\mathbb{R}^d} f(z) (\mathbb{P}_X * \mathbb{P}_Y)(dz),$$

par définition de la convolée. D'où le résultat. \square

Par récurrence, on en déduit que si X_1, X_2, \dots, X_n sont des variables aléatoires indépendantes à valeurs dans \mathbb{R}^d , la loi de la somme $X_1 + X_2 + \dots + X_n$ est la convolée $\mathbb{P}_{X_1} * \mathbb{P}_{X_2} * \dots * \mathbb{P}_{X_n}$. En particulier, la fonction caractéristique de $X_1 + \dots + X_n$ est donnée par

$$\varphi_{X_1 + \dots + X_n}(\xi) = \prod_{i=1}^n \varphi_{X_i}(\xi), \quad \xi \in \mathbb{R}^d.$$

Si l'on a une suite de variables aléatoires X_1, X_2, \dots indépendantes et de même loi (on abrège cela en i.i.d., pour « indépendantes et identiquement distribuées »), la suite des sommes partielles

$$S_n = \sum_{i=1}^n X_i, \quad n \geq 0$$

est appelée une *marche aléatoire* à pas i.i.d.

Notons que si l'on suppose que les variables aléatoires réelles X_1, \dots, X_n sont de carré intégrable, alors

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

En particulier, si les variables aléatoires X_1, \dots, X_n sont de plus indépendantes, on a

$$\text{Var}(X_1 + \dots + X_n) = \sum_{i=1}^n \text{Var}(X_i).$$

Corollaire (loi faible L^2 des grands nombres). Soit X_1, X_2, \dots une suite de variables aléatoires réelles indépendantes et de même loi, telles que $\mathbb{E}[X_1^2] < \infty$. Alors

$$\frac{X_1 + \dots + X_n}{n} \xrightarrow[n \rightarrow \infty]{} \mathbb{E}[X_1],$$

la convergence ayant lieu dans l'espace $L^2(\Omega, \mathcal{F}, \mathbb{P})$.

Démonstration. On constate simplement que, comme $\mathbb{E}[X_1] = \mathbb{E}[X_2] = \dots$,

$$\begin{aligned} \mathbb{E} \left[\left| \frac{X_1 + \dots + X_n}{n} - \mathbb{E}[X_1] \right|^2 \right] &= \mathbb{E} \left[\left| \frac{X_1 + \dots + X_n - \mathbb{E}[X_1 + \dots + X_n]}{n} \right|^2 \right] \\ &= \text{Var} \left(\frac{X_1 + \dots + X_n}{n} \right) \\ &= \frac{1}{n^2} \sum_{i=1}^n \text{Var}(X_i) = \frac{1}{n} \text{Var}(X_1), \end{aligned}$$

ce qui tend vers 0 lorsque $n \rightarrow \infty$. \square

On remarquera que l'on a utilisé uniquement le fait que les variables aléatoires X_1, \dots, X_n ont même espérance et variance, et sont décorrélées, c'est-à-dire que $\text{Cov}(X_i, X_j) = 0$ pour tout $i \neq j$. Comme on l'a vu, cette condition est plus faible que la condition i.i.d.

6.6 Lemme de Borel-Cantelli

Si A_1, A_2, \dots est une suite d'événements, on définit

$$\limsup_{n \rightarrow \infty} A_n = \bigcap_{k \geq 1} \bigcup_{n \geq k} A_n$$

qu'on peut voir comme l'ensemble des $\omega \in \Omega$ qui appartiennent à une infinité des événements A_n . De même, on pose

$$\liminf_{n \rightarrow \infty} A_n = \bigcup_{k \geq 1} \bigcap_{n \geq k} A_n$$

qui est l'ensemble des $\omega \in \Omega$ qui appartiennent à tous les événements A_n , sauf peut-être un nombre fini d'entre eux. Les sous-ensembles $\limsup_{n \rightarrow \infty} A_n$ et $\liminf_{n \rightarrow \infty} A_n$ sont eux-mêmes des événements. Par ailleurs on a clairement

$$\limsup_{n \rightarrow \infty} A_n^c = \left(\liminf_{n \rightarrow \infty} A_n \right)^c, \quad \liminf_{n \rightarrow \infty} A_n^c = \left(\limsup_{n \rightarrow \infty} A_n \right)^c.$$

6.6.1 L'énoncé, et un exemple

Le lemme de Borel-Cantelli est une observation simple mais extrêmement utile.

Lemme 6.12. *Soit A_1, A_2, \dots une suite d'événements. Si $\sum_{n \geq 1} \mathbb{P}(A_n) < \infty$, alors $\mathbb{P}(\limsup_{n \rightarrow \infty} A_n) = 0$.*

Démonstration. L'hypothèse stipule que $\mathbb{E}[\sum_{n \geq 1} \mathbf{1}_{A_n}] < \infty$. Ceci implique que $\sum_{n \geq 1} \mathbf{1}_{A_n} < \infty$ presque sûrement, c'est-à-dire que presque tout $\omega \in \Omega$ n'appartient qu'à un nombre fini des événements A_n . Autrement dit, $\mathbb{P}(\liminf_{n \rightarrow \infty} A_n^c) = 1$, et on conclut en passant au complémentaire. \square

La conclusion importante du lemme est que $\mathbb{P}(\liminf_{n \rightarrow \infty} A_n^c) = 1$, c'est-à-dire que les événements A_n^c ont lieu à partir d'un certain rang. En pratique, si l'on cherche à montrer que des événements ont lieu à partir d'un certain rang, on estime donc les probabilités des complémentaires (A_n est donc un « mauvais » événement) et on essaie de montrer que ces probabilités sont petites (au sens où elles sont sommables). Illustrons ceci par un exemple instructif.

Exemple. Nombre de « pile » consécutifs

Soit (X_1, X_2, \dots) une suite de variables aléatoires de Bernoulli de paramètre $1/2$ indépendantes. Pour tout $n \geq 1$, on note R_n le nombre maximal des X_i consécutifs valant 1, à partir de l'indice n . Formellement

$$R_n = \sup \{m \geq 1 : X_n = X_{n+1} = \dots = X_{n+m-1} = 1\}.$$

Remarquons alors que pour tout $K > 0$,

$$\mathbb{P}(R_n > K) \leq \mathbb{P}(X_n = X_{n+1} = \dots = X_{n+K-1} = 1) = \frac{1}{2^K}.$$

Prenons $K = K(n) = \lfloor (1 + \varepsilon) \log_2(n) \rfloor$ où $\varepsilon > 0$ est fixé. On voit alors que si $A_n = \{R_n > K(n)\}$, on a

$$\mathbb{P}(A_n) \leq \frac{1}{2^{(1+\varepsilon)\log_2(n)-1}} \leq \frac{2}{n^{1+\varepsilon}}$$

Par conséquent, le lemme de Borel-Cantelli implique que $R_n \leq (1 + \varepsilon) \log_2(n)$ pour tout n assez grand, disons $n \geq n_0(\varepsilon)$ où $n_0(\varepsilon)$ est aléatoire, mais fini presque sûrement. En prenant ε de la forme 2^{-k} pour $k \geq 0$, on obtient que presque sûrement,

$$\limsup_{n \rightarrow \infty} \frac{R_n}{\log_2(n)} \leq 1.$$

Notons $M_n = \max(R_1, R_2, \dots, R_n)$. Avec la notation ci-dessus, pour tout $n \geq n_0(\varepsilon)$,

$$M_n \leq \max(R_1, \dots, R_{n_0(\varepsilon)}) \vee (1 + \varepsilon) \log_2(n)$$

et on déduit que l'on a également

$$\limsup_{n \rightarrow \infty} \frac{M_n}{\log_2(n)} \leq 1.$$

Nous montrons à présent le résultat suivant.

Proposition 6.13. *On a presque sûrement*

$$\lim_{n \rightarrow \infty} \frac{M_n}{\log_2(n)} = 1.$$

Par exemple, une suite de 2^n lancers de pile ou face contient au moins un bloc ayant environ n valeurs 1 consécutives si n est assez grand. Par exemple, pour 1000 valeurs consécutives, on devrait vraisemblablement trouver des blocs d'environ 10 valeurs identiques consécutives, mais pas beaucoup plus.

Démonstration. Fixons $\varepsilon \in]0, 1[$, et notons $l_n = \lfloor (1 - \varepsilon) \log_2(n) \rfloor$, et $N_n = \lfloor n / l_n \rfloor$. On note alors $B_j = \{jl_n + 1, jl_n + 2, \dots, (j+1)l_n\}$ pour $0 \leq j \leq N_n - 1$, de sorte que les ensembles B_j sont disjoints deux-à-deux, de cardinal l_n et de réunion incluse dans $\{1, 2, \dots, n\}$. Clairement, s'il existe $j \in \{0, \dots, N_n - 1\}$ tel que $X_i = 1$ pour tout $i \in B_j$, alors on a $M_n \geq l_n$. Montrons donc que cela arrive presque sûrement à partir d'un certain rang.

Pour cela, notons

$$A_n = \{\forall j \in \{0, \dots, N_n - 1\}, \exists i \in B_j : X_i = 0\},$$

et notons que par le lemme de regroupement par paquets, les tribus $\sigma(X_i, i \in B_j)$, $j \in \{0, \dots, N_n - 1\}$ sont indépendantes. Par conséquent

$$\begin{aligned} \mathbb{P}(A_n) &= \prod_{j=0}^{N_n-1} \mathbb{P}(\exists i \in B_j : X_i = 0) \\ &= \prod_{j=0}^{N_n-1} (1 - \mathbb{P}(\forall i \in B_j : X_i = 1)) \\ &= \left(1 - \frac{1}{2^{l_n}}\right)^{N_n} \leq \left(1 - \frac{1}{2n^{1-\varepsilon}}\right)^{N_n}, \end{aligned}$$

où l'on a utilisé une nouvelle fois l'indépendance des X_i à la dernière étape. Par conséquent

$$\mathbb{P}(A_n) = \exp\left(N_n \ln\left(1 - \frac{1}{2n^{1-\varepsilon}}\right)\right) = \exp\left(-\left(\frac{n}{\log_2(n)} - 1\right) \frac{1}{2n^{1-\varepsilon}}(1 + o(1))\right)$$

ce que l'on peut borner par $\exp(-n^{\varepsilon/2})$ pour tout n assez grand. Ce majorant est sommable, et le lemme de Borel-Cantelli permet de conclure que $M_n \geq l_n$ à partir d'un certain rang, comme on l'a vu. Ainsi, on obtient que pour tout $\varepsilon \in]0, 1[$ on a

$$\liminf_{n \rightarrow \infty} \frac{M_n}{\log_2(n)} \geq 1 - \varepsilon$$

presque sûrement, et on conclut. \square

6.6.2 Lemme « réciproque »

Noter que le lemme de Borel-Cantelli ne fait aucunement intervenir une hypothèse d'indépendance des événements A_n ! En revanche, une telle hypothèse est nécessaire pour l'énoncé « réciproque » ci-dessous.

Lemme 6.14. *Soit A_1, A_2, \dots des événements indépendants. Si $\sum_{n \geq 1} \mathbb{P}(A_n) = \infty$, alors $\mathbb{P}(\limsup_{n \rightarrow \infty} A_n) = 1$.*

Démonstration. Pour tout $k \geq 1$, on a par le théorème de convergence dominée,

$$\mathbb{P}\left(\bigcap_{n \geq k} A_n^c\right) = \mathbb{E}\left[\prod_{n \geq k} \mathbf{1}_{A_n^c}\right] = \lim_{N \rightarrow \infty} \mathbb{E}\left[\prod_{n=k}^N \mathbf{1}_{A_n^c}\right] = \lim_{N \rightarrow \infty} \prod_{n=k}^N \mathbb{E}[\mathbf{1}_{A_n^c}],$$

où l'on a utilisé l'indépendance à la dernière étape. Cette limite vaut

$$\prod_{n \geq k} (1 - \mathbb{P}(A_n)) \leq \exp\left(\sum_{n \geq k} -\mathbb{P}(A_n)\right) = 0,$$

où l'on a utilisé la borne $1 - x \leq e^{-x}$, et l'hypothèse de divergence de $\sum \mathbb{P}(A_n)$. On en déduit que

$$\mathbb{P}\left(\liminf_{n \rightarrow \infty} A_n^c\right) \leq \sum_{k \geq 1} \mathbb{P}\left(\bigcap_{n \geq k} A_n^c\right) = 0,$$

et on conclut par passage au complémentaire. \square

Exercice. L'hypothèse d'indépendance de ce lemme est indispensable ! Donner un ou des contre-exemples naturels si les événements ne sont pas indépendants.

Nous donnons maintenant deux applications de la seconde version du lemme de Borel-Cantelli.

Une mesure « uniforme » sur \mathbb{N} ?

Comme première application, on montre qu'il n'existe pas de mesure de probabilités \mathbb{P} « bien répartie » sur \mathbb{N} , au sens où $\mathbb{P}(n\mathbb{N}) = 1/n$ pour tout $n \geq 1$. En effet, si tel était le cas, les événements $A_p = p\mathbb{N}$ seraient tous indépendants lorsque p décrit l'ensemble des nombres premiers. En effet, on aurait, pour tous les nombres p_1, \dots, p_k premiers distincts,

$$\mathbb{P}(A_{p_1} \cap \dots \cap A_{p_k}) = \mathbb{P}(p_1 \cdots p_k \mathbb{N}) = \frac{1}{p_1 \cdots p_k} = \prod_{i=1}^k \mathbb{P}(A_{p_i}).$$

Comme $\sum_p 1/p = \infty$, où la somme porte sur l'ensemble des nombres premiers, on en déduirait, par la seconde version du lemme de Borel-Cantelli que \mathbb{P} -presque tout entier n est dans une infinité des ensembles $p\mathbb{N}$ avec p premier, ce qui est clairement impossible.

Motifs dans une suite de pile ou face

Donnons une autre application simple de ce lemme. Considérons une suite X_1, X_2, \dots de variables aléatoires de Bernoulli de paramètres respectifs p_i . En posant $A_i = \{X_i = 1\}$ et en appliquant les deux lemmes précédents, on obtient

- Si $\sum_{i=1}^{\infty} p_i < \infty$ alors presque sûrement, seul un nombre fini des X_i est non nul.
- Si $\sum_{i=1}^{\infty} p_i = \infty$ **et si** les variables aléatoires X_1, X_2, \dots sont indépendantes, alors, presque sûrement, une infinité des variables X_i valent 1.

Ceci a une conséquence intéressante sur les nombres réels. Rappelons que, si U est une variable aléatoire de loi uniforme sur $[0, 1[$, alors la suite $X = (X_1, X_2, \dots)$ de son développement dyadique propre

$$U = \sum_{i \geq 1} \frac{X_i}{2^i}$$

est une suite de variables aléatoires de loi de Bernoulli de paramètre $1/2$, indépendantes. Fixons $\varepsilon = (\varepsilon_1, \dots, \varepsilon_n) \in \{0, 1\}^n$ une suite finie quelconque, et un entier $k \geq 0$, et posons $A_\varepsilon(k) = \{\omega \in \{0, 1\}^{\mathbb{N}^*} : (\omega_{k+1}, \dots, \omega_{k+n}) = \varepsilon\}$. Soit $B_\varepsilon = \limsup_{k \rightarrow \infty} \{X \in A_\varepsilon(k)\}$ l'événement que la suite ε apparaisse une infinité de fois dans le développement dyadique propre de X .

Les événements $\{X \in A_\varepsilon(k)\} = \{X_{k+1} = \varepsilon_1, \dots, X_{k+n} = \varepsilon_n\}$, $k \geq 0$ ne sont pas indépendants. En revanche, les événements $\{X \in A_\varepsilon(kn)\}$, $k \geq 0$ sont respectivement mesurables par rapport aux σ -algèbres $\sigma(X_{kn+1}, \dots, X_{(k+1)n})$, qui sont indépendantes par le regroupement par paquets. De plus, $\mathbb{P}(X \in A_\varepsilon(kn)) = 1/2^n$ pour tout ε et tout k . La somme sur k des probabilités de ces événements est donc infinie, et le deuxième lemme de Borel-Cantelli implique donc que

$$\mathbb{P}(B_\varepsilon) \geq \mathbb{P}\left(\limsup_{k \rightarrow \infty} \{X \in A_\varepsilon(kn)\}\right) = 1.$$

Comme l'ensemble des suites finies de 0 et de 1 est dénombrable, on en déduit que, si B est l'intersection des événements B_ε sur toutes les suites finies ε , alors $\mathbb{P}(B) = 1$. Nous avons montré qu'avec probabilité 1, toute suite finie apparaît une infinité de fois dans le développement dyadique de la variable aléatoire U . Autrement dit, presque tout nombre (pour la mesure de Lebesgue) satisfait cette propriété, et contient en particulier une infinité de fois tous les romans du monde codés en binaire, ainsi que ceux qui restent à écrire. On peut y trouver également des tentatives de preuve ou de réfutation assez convaincantes de l'hypothèse de Riemann...

6.7 Loi du 0-1 de Kolmogorov

Soit $(\mathcal{F}_1, \mathcal{F}_2, \dots)$ une suite de σ -algèbres. Notons

$$\mathcal{G}_n = \bigvee_{k \geq n} \mathcal{F}_k \quad \text{et} \quad \mathcal{G}_\infty = \bigcap_{n \geq 1} \mathcal{G}_n.$$

On dit que \mathcal{G}_∞ est la σ -algèbre asymptotique associée à la suite $(\mathcal{F}_1, \mathcal{F}_2, \dots)$.

Théorème 6.15. *Si les σ -algèbres $\mathcal{F}_1, \mathcal{F}_2, \dots$ sont indépendantes, alors la tribu asymptotique est triviale au sens où pour tout $A \in \mathcal{G}_\infty$, on a $\mathbb{P}(A) \in \{0, 1\}$.*

Démonstration. Par le lemme de regroupement par paquets, on a que \mathcal{G}_{n+1} est indépendante de \mathcal{F}_i , ce pour tout $n \geq 1$ et $i \in \{1, 2, \dots, n\}$. Comme $\mathcal{G}_\infty \subset \mathcal{G}_{n+1}$, on en déduit que \mathcal{G}_∞ est indépendante de \mathcal{F}_i pour tout $i \geq 1$. Donc \mathcal{G}_∞ est indépendante de $\bigvee_{i \geq 1} \mathcal{F}_i = \mathcal{G}_1$, à nouveau par le lemme de regroupement par paquets. Mais comme $\mathcal{G}_\infty \subset \mathcal{G}_1$, on en déduit que \mathcal{G}_∞ est indépendante d'elle-même, ce qui signifie que $\mathbb{P}(A) = \mathbb{P}(A \cap A) = \mathbb{P}(A)^2$ pour tout $A \in \mathcal{G}_\infty$, CQFD. \square

Comme exemple d'application, on en déduit le résultat suivant sur une marche aléatoire $(S_n, n \geq 0)$ à pas i.i.d. X_1, X_2, \dots

Proposition 6.16. *Soit $(X_n, n \geq 1)$ une suite de variables aléatoires réelles indépendantes, et soit $S_n = X_1 + \dots + X_n$ pour tout $n \geq 1$, $S_0 = 0$. Alors les événements*

$$\left\{ \limsup_{n \rightarrow \infty} S_n = \infty \right\}, \quad \left\{ \liminf_{n \rightarrow \infty} S_n = -\infty \right\},$$

sont de probabilité 0 ou 1.

Démonstration. Il suffit de montrer le résultat pour le premier événement, quitte à changer la suite $(X_n, n \geq 1)$ en leurs opposés.

On a que pour tout $k \geq 1$, $\{\limsup_{n \rightarrow \infty} S_n = \infty\} = \{\limsup_{n \rightarrow \infty} (S_n - S_k) = \infty\}$, et cette dernière limite supérieure est

$$\limsup_{n \rightarrow \infty} (X_{k+1} + X_{k+2} + \dots + X_n),$$

qui est mesurable par rapport à $\mathcal{G}_{k+1} = \sigma(X_{k+1}, X_{k+2}, \dots) = \bigvee_{i \geq k+1} \sigma(X_i)$. En particulier, $\{\limsup_{n \rightarrow \infty} S_n = \infty\}$ est mesurable par rapport à la σ -algèbre \mathcal{G}_k pour tout $k \geq 1$, et donc par rapport à la tribu asymptotique \mathcal{G}_∞ . Comme les σ -algèbres $\sigma(X_i), i \geq 1$ sont indépendantes par hypothèse, la loi du 0-1 de Kolmogorov s'applique et donne le résultat. \square

Corollaire 6.17. *Supposons que les variables $X_n, n \geq 1$ sont i.i.d. et que*

$$\mathbb{P}(X_1 = 1) = \mathbb{P}(X_1 = -1) = \frac{1}{2}.$$

Alors la marche aléatoire $S_n = X_1 + \dots + X_n$ oscille, au sens où

$$\mathbb{P}\left(\limsup_{n \rightarrow \infty} S_n = \infty, \liminf_{n \rightarrow \infty} S_n = -\infty\right) = 1$$

Démonstration. On sait que les événements $\{\liminf_{n \rightarrow \infty} S_n = \infty\}, \{\limsup_{n \rightarrow \infty} S_n = -\infty\}$ sont de probabilité 0 ou 1. Mais comme la loi de $(S_n, n \geq 0)$ est la même que celle de $(-S_n, n \geq 0)$ par symétrie de la loi des variables aléatoires $X_n, n \geq 0$, ces deux événements ont aussi la même probabilité.

Nous allons montrer que la suite $(S_n, n \geq 0)$ n'est pas bornée avec probabilité 1. Ceci signifie que

$$\mathbb{P}\left(\left\{\limsup_{n \rightarrow \infty} S_n = \infty\right\} \cup \left\{\liminf_{n \rightarrow \infty} S_n = -\infty\right\}\right) = 1.$$

Par conséquent, l'un au moins des deux événements est de probabilité strictement positive. Par les remarques ci-dessus, la probabilité est en fait 1 pour ces deux événements.

Il reste à montrer que $(S_n, n \geq 0)$ n'est pas bornée avec probabilité 1. Introduisons l'événement

$$A_K = \{|S_n| \leq K \text{ pour tout } n \geq 0\},$$

où K est un entier donné. Posons $N = 2K + 1$, et constatons que

$$\bigcup_{k \geq 0} \{X_{kN+1} = 1, \dots, X_{(k+1)N} = 1\} \subset A_K^c.$$

Or $\mathbb{P}(X_{kN+1} = 1, \dots, X_{(k+1)N} = 1) = 1/2^N > 0$, et de plus les événements $\{X_{kN+1} = 1, \dots, X_{(k+1)N} = 1\}, k \geq 0$ sont indépendants. Par conséquent,

$$\mathbb{P}\left(\bigcap_{k \geq 0} \{X_{kN+1} = 1, \dots, X_{(k+1)N} = 1\}^c\right) = \prod_{k \geq 0} \left(1 - \frac{1}{2^N}\right) = 0,$$

et on en déduit que $\mathbb{P}(A_K) = 0$ pour tout K . En prenant la réunion sur K , on en déduit le résultat. \square

Exercice. Généraliser le résultat précédent à une marche aléatoire $S_n = X_1 + \dots + X_n$ à pas i.i.d. dont la loi est symétrique, c'est-à-dire que X_1 et $-X_1$ ont même loi, dès que cette loi n'est pas δ_0 .

6.8 Complément : existence d'une suite de variables aléatoires indépendantes

Soit (E, \mathcal{E}, μ) un espace de probabilités. Rappelons que l'on peut toujours construire une variable aléatoire X de loi μ , en choisissant par exemple la variable aléatoire canonique. Le théorème suivant montre qu'on peut faire beaucoup mieux.

Théorème 6.18. *Soit $(E_n, \mathcal{E}_n, \mu_n), n \geq 1$ une suite d'espaces de probabilités. Alors il existe un espace de probabilités sur lequel est définie une suite de variables aléatoires $(X_n, n \geq 1)$ indépendantes, de lois respectives $\mathbb{P}_{X_n} = \mu_n$.*

Ce théorème est un résultat d'existence de la mesure produit $\mu = \bigotimes_{n \geq 1} \mu_n$ sur l'espace produit $E = E_1 \times E_2 \times \dots$ muni de la tribu produit $\mathcal{E} = \mathcal{E}_1 \otimes \mathcal{E}_2 \otimes \dots$, qui est la plus petite tribu rendant mesurables les applications de projection de $X_n: E \rightarrow E_n$. Si elle existe, la mesure produit μ est définie par

$$\mu(X_1 \in A_1, \dots, X_n \in A_n) = \prod_{i=1}^n \mu_i(A_i), \quad A_1 \in \mathcal{E}_1, \dots, A_n \in \mathcal{E}_n,$$

et une telle mesure est nécessairement unique par le théorème de classe monotone. Sous réserve que cette mesure existe bien, il suffit de prendre pour $X = (X_1, X_2, \dots)$ la variable canonique sur (E, \mathcal{E}, μ) .

Nous n'allons pas montrer ce théorème en toute généralité, mais expliquons comment on peut l'obtenir pour $E_n = \mathbb{R}$ pour tout $n \geq 1$. Rappelons que l'on a construit au chapitre 5.1 une suite infinie (Y_1, Y_2, \dots) de variables aléatoires indépendantes uniformes dans $\{0, 1\}$. On peut alors réindexer cette suite en $(Z_{n,m} : n, m \geq 1)$, en prenant une bijection $\varphi: \mathbb{N}^* \times \mathbb{N}^* \rightarrow \mathbb{N}^*$ et en posant

$$Z_{n,m} = Y_{\varphi(n,m)}.$$

La famille $(Z_{n,m} : n, m \geq 1)$ est clairement formée de variables aléatoires indépendantes uniformes dans $\{0, 1\}$. Pour tout $n \geq 1$, on pose alors

$$U_n = \sum_{m \geq 1} \frac{Z_{n,m}}{2^m},$$

ce qui définit une suite de variables uniformes $(U_n, n \geq 1)$ qui sont respectivement mesurables par rapport à $(Z_{n,m} : m \geq 1)$, et dont indépendantes par regroupement par paquets. Ces variables aléatoires sont de plus clairement uniformes dans $[0, 1]$ par lemme de classe monotone, puisque la probabilité que U_n soit dans l'intervalle dyadique $[k2^{-m}, (k+1)2^{-m}[$ est égale à

$$\mathbb{P}(Z_{n,1} = \omega_1, \dots, Z_{n,m} = \omega_m) = \frac{1}{2^m}$$

où $k2^{-m}$ s'écrit $\sum_{i=1}^m \omega_i 2^{-i}$. Pour conclure, on peut utiliser la technique de simulation de variables aléatoires réelles par l'inverse de la fonction de répartition. Soit $F_n(x) = \mu_n(]-\infty, x])$ la fonction de répartition d'une variable aléatoire de loi μ_n , et

$$F_n^{-1}(u) = \inf \{x \in \mathbb{R} : F_n(x) \geq u\}, \quad u \in]0, 1[.$$

Notons que pour tout $x \in \mathbb{R}$ et $u \in]0, 1[$, on a $F_n^{-1}(u) \leq x$ si et seulement si $u \leq F_n(x)$. En effet, si $F_n^{-1}(u) \leq x$ alors par définition et croissance de F_n , cela implique que $F_n(y) \geq u$ pour tout $y > x$. En faisant tendre y vers x par valeurs supérieures, on obtient $F_n(x) \geq u$ par continuité à droite de F_n . La réciproque est immédiate par définition. Posons alors

$$X_n = F_n^{-1}(U_n), \quad n \geq 1,$$

ce qui définit une suite de variables aléatoires indépendantes, puisqu'elles sont respectivement mesurables par rapport aux σ -algèbres indépendantes $\sigma(U_n)$. Mais par ailleurs, on a, pour tout $x \in \mathbb{R}$,

$$\mathbb{P}(X_n \leq x) = \mathbb{P}(U_n \leq F_n(x)) = F_n(x),$$

ce qui signifie que X_n a pour fonction de répartition F_n , et donc que X_n a pour loi μ_n . Le théorème est donc démontré dans le cas où les mesures μ_n sont définies sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$.

Remarque. L'approche présentée ci-dessus peut se généraliser (avec un certain effort !) au cas où les espaces $(E_n, \mathcal{E}_n, \mu_n)$ sont des espaces métriques séparés complets, munis de leurs tribus boréliennes et d'une mesure de probabilités. Pour autant, le théorème reste vrai sans cette contrainte en plus, et se démontre en toute généralité par une application du théorème de prolongement de mesures de Carathéodory.

Chapitre 7

Lois des grands nombres

Nous allons maintenant étudier un résultat fondamental en probabilités, stipulant essentiellement qu'une somme de variables aléatoires indépendantes se comporte en première approximation comme sa moyenne. Nous avons déjà vu au chapitre précédent que si l'on a des variables aléatoires X_1, X_2, \dots dans L^2 qui sont indépendantes (ou décorrélées) et de même loi, alors

$$\frac{X_1 + \dots + X_n}{n} \xrightarrow[n \rightarrow \infty]{L^2} \mathbb{E}[X_1].$$

Nous allons voir de nombreuses variantes de cette loi « faible » des grands nombres. Comme il s'agit d'un résultat asymptotique, nous allons dans un premier temps décrire quelques-uns des modes de convergence de variables aléatoires qui sont usuellement considérés en théorie des probabilités.

7.1 Différentes notions de convergence pour des variables aléatoires

Soit $(X_n, n \geq 1)$ une suite de variables aléatoires, et X une autre variable aléatoire, toutes étant définies sur un espace de probabilités $(\Omega, \mathcal{F}, \mathbb{P})$, et à valeurs dans \mathbb{R} ou \mathbb{C} .

a. Convergence presque sûre

On dit que $(X_n, n \geq 1)$ converge vers X presque sûrement, et on note $X_n \xrightarrow{p.s.} X$, si l'événement $\{\lim_{n \rightarrow \infty} X_n = X\} = \{\limsup_{n \rightarrow \infty} |X_n - X| = 0\}$ est presque sûr, c'est-à-dire

$$\mathbb{P}\left(\left\{\omega \in \Omega : \lim_{n \rightarrow \infty} X_n(\omega) = X(\omega)\right\}\right) = 1.$$

Il s'agit d'un mode de convergence intuitif d'un point de vue probabiliste, puisqu'il énonce une propriété (la convergence d'une suite de variables aléatoires) vraie « ω par ω ». C'est (à un ensemble de probabilité nulle près) la convergence ponctuelle des fonctions mesurables X_n vers X .

b. Convergence L^p

Soit $p \in [1, \infty]$. On dit que $(X_n, n \geq 1)$ converge vers X dans L^p si $\mathbb{E}[|X_n - X|^p] \rightarrow_{n \rightarrow \infty} 0$. Il s'agit de la convergence usuelle dans l'espace de Banach $L^p(\Omega, \mathcal{F}, \mathbb{P})$ muni de la norme $\|\cdot\|_p$.

À l'exception du cas où $p = \infty$, la convergence dans L^p n'implique pas la convergence presque sûre. En revanche, les résultats classiques sur les espaces L^p impliquent le résultat suivant.

Proposition 7.1. *Soit $(X_n, n \geq 0)$ une suite de variables aléatoires convergeant vers X dans L^p . Alors il existe une extraction $(n_k, k \geq 1)$ telle que $(X_{n_k}, k \geq 1)$ converge presque sûrement vers X .*

De même, la convergence p.s. de la suite $(X_n, n \geq 0)$ n'implique pas la convergence dans L^p . Néanmoins, pour $p \in [1, \infty[$, le théorème de convergence dominée implique que, si $|X_n| < Y$ avec $Y \in L^p(\Omega, \mathcal{F}, \mathbb{P})$, alors X_n converge vers X dans L^p .

c. Convergence en probabilité

On dit que $(X_n, n \geq 1)$ converge vers X en probabilité, si pour tout $\varepsilon > 0$ on a

$$\mathbb{P}(|X_n - X| > \varepsilon) \xrightarrow[n \rightarrow \infty]{} 0.$$

On note alors

$$X_n \xrightarrow[n \rightarrow \infty]{\mathbb{P}} X.$$

Proposition 7.2. *Si $(X_n, n \geq 1)$ converge vers X presque sûrement ou dans L^p (pour un $p \geq 1$ donné), alors on a aussi convergence en probabilité.*

Démonstration. Si $(X_n, n \geq 1)$ converge presque sûrement vers X , alors pour tout $\varepsilon > 0$, on peut appliquer le théorème de convergence dominée dans $\mathbb{P}(|X_n - X| > \varepsilon) = \mathbb{E}[\mathbf{1}_{\{|X_n - X| > \varepsilon\}}]$, en constatant que l'indicatrice est de limite 0 presque sûrement.

Si l'on a convergence dans L^p avec $p \in [1, \infty[$ on applique l'inégalité de Markov :

$$\mathbb{P}(|X_n - X| > \varepsilon) \leq \frac{\mathbb{E}[|X_n - X|^p]}{\varepsilon^p} \xrightarrow[n \rightarrow \infty]{} 0.$$

Si $p = \infty$ c'est encore plus simple puisque la convergence dans L^∞ implique la convergence presque sûre. Dans tous les cas, on a montré la convergence en probabilité. \square

Nous montrons maintenant que la convergence en probabilité est associée à une topologie sur les variables aléatoires.

Proposition 7.3. *Soit $L^0(\Omega, \mathcal{F}, \mathbb{P})$ l'ensemble des variables aléatoires (réelles ou complexes) sur l'espace de probabilités $(\Omega, \mathcal{F}, \mathbb{P})$, définies à égalité p.s. près. Définissons, pour $X, Y \in L^0(\Omega, \mathcal{F}, \mathbb{P})$,*

$$d^0(X, Y) = \mathbb{E}[|X - Y| \wedge 1].$$

Alors d^0 est une distance sur $L^0(\Omega, \mathcal{F}, \mathbb{P})$, et l'on a que $(X_n, n \geq 1)$ converge en probabilité vers X si et seulement si cette même suite converge dans $(L^0(\Omega, \mathcal{F}, \mathbb{P}), d^0)$. De plus, cet espace métrique est complet.

On laisse en exercice le fait que d^0 est une distance. Si la suite $(X_n, n \geq 1)$ converge en probabilité, on a pour tout $\varepsilon > 0$

$$\mathbb{E}[|X_n - X| \wedge 1] = \mathbb{E}[(|X_n - X| \wedge 1) \cdot (\mathbf{1}_{\{|X_n - X| \leq \varepsilon\}} + \mathbf{1}_{\{|X_n - X| > \varepsilon\}})] \leq \varepsilon + \mathbb{P}(|X_n - X| > \varepsilon),$$

et donc $\limsup_{n \rightarrow \infty} \mathbb{E}[|X_n - X| \wedge 1] \leq \varepsilon$. Donc $d^0(X_n, X) \rightarrow 0$. D'autre part, on a clairement, pour $\varepsilon \in]0, 1[$,

$$\mathbb{P}(|X_n - X| > \varepsilon) = \mathbb{P}(|X_n - X| \wedge 1 > \varepsilon) \leq \frac{d^0(X_n, X)}{\varepsilon}$$

par l'inégalité de Markov. La réciproque s'ensuit immédiatement.

Montrons le caractère complet. Pour cela, soit $(X_n, n \geq 1)$ une suite de Cauchy pour la distance d^0 . On peut trouver une extraction $(n_k, k \geq 1)$ telle que

$$d^0(X_{n_{k+1}}, X_{n_k}) \leq 2^{-k}.$$

On voit alors que

$$\mathbb{E} \left[\sum_{k \geq 1} (|X_{n_{k+1}} - X_{n_k}| \wedge 1) \right] = \sum_{k \geq 1} d^0(X_{n_{k+1}}, X_{n_k}) < \infty,$$

de sorte que presque sûrement, on a $\sum_{k \geq 1} (|X_{n_{k+1}} - X_{n_k}| \wedge 1) < \infty$, et donc aussi $\sum_{k \geq 1} |X_{n_{k+1}} - X_{n_k}| < \infty$. On pose alors

$$X = X_{n_1} + \sum_{k \geq 1} (X_{n_{k+1}} - X_{n_k}),$$

cette série convergeant avec probabilité 1 par ce qui précède. En particulier, X_{n_k} converge presque sûrement vers X . On en déduit que X_{n_k} converge aussi en probabilité vers X , et donc pour la distance d^0 . Comme $(X_n, n \geq 1)$ est une suite de Cauchy admettant une suite extraite qui converge, on en conclut que X est la limite de $(X_n, n \geq 1)$ dans l'espace (L^0, d^0) , et que ce dernier est complet.

Au cours de la preuve, nous avons montré le résultat suivant, qui est un résultat analogue à un théorème du cours sur les espaces L^p .

Proposition 7.4. *Si $(X_n, n \geq 1)$ converge en probabilité vers X , alors il existe une extraction $(n_k, k \geq 1)$ telle que $(X_{n_k}, k \geq 1)$ converge presque sûrement vers X .*

Remarque. Nous avons vu que la convergence L^p et la convergence en probabilité sont toutes les deux issues de la convergence dans un espace métrique. Il n'en est pas de même (en général) pour la convergence p.s.

Exercice. Montrer que si (X, d) est un espace métrique, et si $(x_n, n \geq 1)$ est une suite de X , alors $(x_n, n \geq 1)$ converge vers $x \in X$ si et seulement si de toute sous-suite, on peut réextraire une sous-sous-suite qui converge vers x .

Déterminer une suite de variables aléatoires $(X_n, n \geq 1)$ telle que de toute sous-suite on puisse réextraire une sous-sous-suite qui converge presque sûrement vers 0, mais telle que X_n ne converge pas presque sûrement. On pourra penser à des variables aléatoires de Bernoulli bien choisies, définies sur l'espace $([0, 1], \mathcal{B}([0, 1]), \lambda)$ où λ est la mesure de Lebesgue.

On peut enfin « remonter » de la convergence en probabilité à une convergence L^p si l'on a une hypothèse de moments.

Proposition 7.5. *Soit $q > 1$ et $(X_n, n \geq 1)$ une suite de variables aléatoires bornée dans L^q , c'est-à-dire que $\sup_{n \geq 1} \mathbb{E}[|X_n|^q] < \infty$. On suppose que X_n converge en probabilité vers X . Alors pour tout $p \in [1, q[$ on a que X_n converge vers X dans L^p .*

Démonstration. Tout d'abord, notons que la limite X est dans L^q . En effet, par la proposition 7.4, il existe une sous-suite $(X_{n_k}, k \geq 1)$ convergeant vers X presque sûrement. Le lemme de Fatou donne alors

$$\mathbb{E}[|X|^q] \leq \liminf_{k \rightarrow \infty} \mathbb{E}[|X_{n_k}|^q] < \infty \tag{7.1}$$

puisque la suite $(X_n, n \geq 1)$ est bornée dans L^q . Comme $L^q \subset L^p$ on déduit que $X \in L^p$ pour tout $p \in [1, q[$. On écrit alors, pour tout $\varepsilon > 0$,

$$\begin{aligned} \mathbb{E}[|X_n - X|^p] &= \mathbb{E}[|X_n - X|^p \mathbf{1}_{\{|X_n - X| > \varepsilon\}}] + \mathbb{E}[|X_n - X|^p \mathbf{1}_{\{|X_n - X| \leq \varepsilon\}}] \\ &\leq \mathbb{E}[|X_n - X|^q]^{p/q} \mathbb{P}(|X_n - X| > \varepsilon)^{(q-p)/q} + \varepsilon^p \end{aligned}$$

où l'on a utilisé à la seconde étape l'inégalité de Hölder pour les exposants q/p et son conjugué $q/(q-p)$. Comme $\mathbb{E}[|X_n - X|^q]^{p/q}$ est borné par hypothèse et par (7.1), on en déduit que pour tout $\varepsilon > 0$, on a

$$\limsup_{n \rightarrow \infty} \mathbb{E}[|X_n - X|^p] \leq \varepsilon^p.$$

C'est ce qu'il fallait démontrer. □

On obtient le diagramme d'implications suivant :

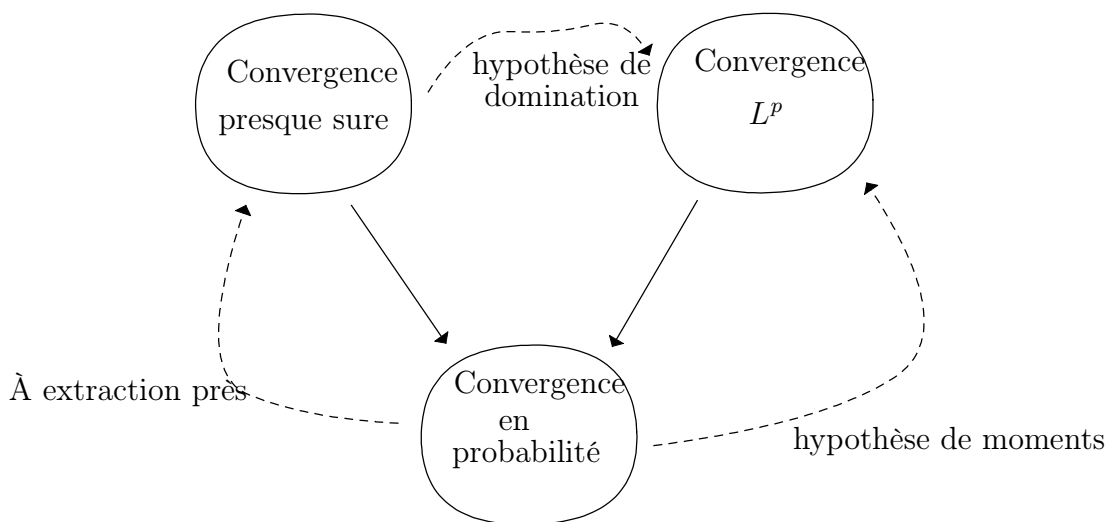


Figure 7.1. Diagramme d'implications

7.2 La loi forte des grands nombres

Le but de cette partie est de montrer le théorème suivant.

Théorème (Loi forte des grands nombres). Soit X_1, X_2, \dots une suite de variables aléatoires réelles indépendantes et de même loi, dans L^1 . Alors

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

Remarquons que ce théorème est également vrai pour des variables aléatoires à valeurs dans \mathbb{R}^d , pour tout $d \geq 1$: pour le voir, il suffit d'appliquer le théorème précédent coordonnée par coordonnée. Par ailleurs, on a également la convergence au sens L^1 , ce qui est un résultat beaucoup plus facile.

Proposition 7.6. *Sous les mêmes hypothèses, on a également*

$$\frac{X_1 + \dots + X_n}{n} \xrightarrow[n \rightarrow \infty]{L^1} \mathbb{E}[X_1].$$

Démonstration. Remarquons que si $X_1 \in L^2$, alors le résultat est une conséquence de la loi faible L^2 des grands nombres, puisque la convergence dans L^2 implique celle dans L^1 . Dans le cas général, pour K donné, et $i \geq 1$, notons $Y_i = X_i \mathbf{1}_{\{|X_i| \leq K\}}$ et $Z_i = X_i \mathbf{1}_{\{|X_i| > K\}}$, de sorte que $X_i = Y_i + Z_i$. Comme les variables aléatoires (Y_i) sont i.i.d. dans L^2 , et que les (Z_i) sont i.i.d. également et dans L^1 , on a que pour tout K ,

$$\begin{aligned} \mathbb{E} \left[\left| \frac{S_n}{n} - \mathbb{E}[X_1] \right| \right] &\leq \mathbb{E} \left[\left| \frac{\sum_{i=1}^n Y_i}{n} - \mathbb{E}[Y_1] \right| \right] + \mathbb{E} \left[\left| \frac{\sum_{i=1}^n Z_i}{n} - \mathbb{E}[Z_1] \right| \right] \\ &\leq \mathbb{E} \left[\left| \frac{\sum_{i=1}^n Y_i}{n} - \mathbb{E}[Y_1] \right| \right] + 2\mathbb{E}[|Z_1|]. \end{aligned}$$

Donc on a

$$\limsup_{n \rightarrow \infty} \mathbb{E} \left[\left| \frac{S_n}{n} - \mathbb{E}[X_1] \right| \right] \leq 2\mathbb{E}[|Z_1|].$$

Mais par convergence dominée, on a que ce majorant tend vers 0 lorsque $K \rightarrow \infty$. D'où le résultat. \square

Nous allons donner plusieurs approches de la loi forte des grands nombres, sous des hypothèses de moins en moins restrictives, pour donner une idée des diverses méthodes qui permettent d'approcher un tel résultat. D'autres preuves standard de ce résultat font appel à des résultats plus poussés de la théorie ergodique (théorème de Birkhoff) ou de la théorie des martingales.

Jusqu'à la fin de la partie 7.2, on suppose que les variables aléatoires X_1, X_2, \dots sont i.i.d. On notera $S_n = X_1 + \dots + X_n$ pour simplifier.

7.2.1 Le cas L^4

Supposons dans un premier temps que $\mathbb{E}[X_1^4] < \infty$. Notons alors que, quitte à changer X_n en $X_n - \mathbb{E}[X_1]$, on peut supposer ces variables aléatoires centrées, c'est-à-dire que $\mathbb{E}[X_1] = 0$. Nous allons montrer que presque sûrement, on a

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

Ceci impliquera alors clairement que S_n/n converge vers 0 presque sûrement. Pour cela, il suffit de montrer que $\mathbb{E}[\Sigma] < \infty$. Nous estimons donc

$$\mathbb{E} \left[\left(\frac{S_n}{n} \right)^4 \right] = \frac{1}{n^4} \sum_{i_1, i_2, i_3, i_4=1}^n \mathbb{E}[X_{i_1} X_{i_2} X_{i_3} X_{i_4}].$$

Dans cette dernière somme, notons que si l'un des indices est distinct des trois autres, par exemple $i_1 \notin \{i_2, i_3, i_4\}$, alors l'indépendance implique que $\mathbb{E}[X_{i_1}X_{i_2}X_{i_3}X_{i_4}] = \mathbb{E}[X_{i_1}]\mathbb{E}[X_{i_2}X_{i_3}X_{i_4}] = 0$. Ne restent dans la somme que les indices pour lesquels $\{i_1, i_2, i_3, i_4\}$ est de cardinal 1 ou 2. On en déduit

$$\begin{aligned}\mathbb{E}\left[\left(\frac{S_n}{n}\right)^4\right] &= \frac{1}{n^4}\sum_{i=1}^n \mathbb{E}[X_i^4] + 3\cdot\frac{2}{n^4}\sum_{1\leq i<j\leq n} \mathbb{E}[X_i^2]\mathbb{E}[X_j^2] \\ &= \frac{1}{n^3}\mathbb{E}[X_1^4] + \frac{3(n-1)}{n^3}\mathbb{E}[X_1^2]^2.\end{aligned}$$

Cette dernière quantité est sommable en $n \geq 1$. On en déduit donc que S_n/n converge vers 0 presque sûrement et dans L^4 .

7.2.2 Le cas L^2

Supposons à présent que $\mathbb{E}[X_1^2] < \infty$. Une fois encore, on peut supposer que $\mathbb{E}[X_1] = 0$. Nous avons déjà vu que S_n/n converge vers 0 dans L^2 (donc dans L^1) : c'est la loi faible des grands nombres. On en déduit ainsi qu'il existe une sous-suite le long de laquelle S_n/n converge p.s. vers 0. Essayons d'être plus précis, en considérant explicitement l'extraction $(k^2, k \geq 1)$. On a

$$\mathbb{E}\left[\left(\frac{S_{k^2}}{k^2}\right)^2\right] = \frac{1}{k^4}\text{Var}(S_{k^2}) = \frac{1}{k^2}\text{Var}(X_1).$$

Ceci étant sommable en k , on en déduit comme précédemment que S_{k^2}/k^2 converge vers 0 dans L^2 et presque sûrement.

Pour se débarrasser de l'extraction, il faut contrôler la suite S_n/n entre deux valeurs consécutives de la suite extraite. Plus exactement, pour $\varepsilon > 0$, on considère

$$A_k = \left\{ \max_{k^2 \leq n < (k+1)^2} |S_n - S_{k^2}| > \varepsilon k^2 \right\}, \quad k \geq 1.$$

Notons que

$$\begin{aligned}\mathbb{P}(A_k) &\leq \sum_{k^2 \leq n < (k+1)^2} \mathbb{P}(|S_n - S_{k^2}| > \varepsilon k^2) \\ &\leq \sum_{k^2 \leq n < (k+1)^2} \frac{\text{Var}(S_n - S_{k^2})}{\varepsilon^2 k^4}\end{aligned}$$

par l'inégalité de Bienaymé-Chebychev. Or, $S_n - S_{k^2} = X_{k^2+1} + X_{k^2+2} + \dots + X_n$ a même loi que $S_{n-k^2} = X_1 + \dots + X_{n-k^2}$. On en déduit que $\text{Var}(S_n - S_{k^2}) \leq (n - k^2)\text{Var}(X_1)$ et donc

$$\mathbb{P}(A_k) \leq \frac{((k+1)^2 - k^2)^2}{\varepsilon^2 k^4} \text{Var}(X_1).$$

Comme $(k+1)^2 - k^2 = 2k+1$, on voit que le majorant est sommable en k . Le lemme de Borel-Cantelli implique donc que presque sûrement, pour tout k assez grand, on a que

$$M_k = \max_{k^2 \leq n < (k+1)^2} |S_n - S_{k^2}| \leq \varepsilon k^2.$$

Soit alors $n \in \mathbb{N}^*$, et $k = k(n) \in \mathbb{N}$ l'unique entier tel que $k^2 \leq n < (k+1)^2$. On a alors

$$\left| \frac{S_n}{n} \right| \leq \frac{k^2}{n} \left(\left| \frac{S_{k^2}}{k^2} \right| + \left| \frac{M_k}{k^2} \right| \right)$$

et donc, presque sûrement,

$$\limsup_{n \rightarrow \infty} \left| \frac{S_n}{n} \right| \leq \varepsilon.$$

Ceci étant valable pour tout nombre rationnel $\varepsilon > 0$, on déduit que S_n/n converge vers 0 presque sûrement.

7.2.3 Le cas L^1 par la méthode d'écrêtement

Nous allons donner maintenant une première preuve du théorème 7.2 due à Etemadi. Dans un premier temps, nous remarquons que si l'on écrit $X_n = X_n^+ - X_n^-$, alors

$$\frac{S_n}{n} = \frac{X_1^+ + \dots + X_n^+}{n} - \frac{X_1^- + \dots + X_n^-}{n}$$

où les variables X_1^+, X_2^+, \dots d'une part et X_1^-, X_2^-, \dots d'autre part sont i.i.d. et positives, d'espérances finies. Il suffit donc de montrer le théorème pour des variables aléatoires positives, ce que l'on suppose maintenant. Attention, on prendra garde au fait que l'on ne peut plus dès lors supposer les variables centrées !

Nous utilisons maintenant la notion de variable aléatoire tronquée, et posons

$$Y_n = X_n \mathbf{1}_{\{X_n \leq n\}}.$$

Lemme 7.7. *Presque sûrement, on a que $Y_n = X_n$ pour tout n assez grand.*

Démonstration. Notons que $\{X_n = Y_n\} = \{X_n \leq n\}$. Si l'on pose $A_n = \{X_n > n\}$ alors

$$\sum_{n \geq 1} \mathbb{P}(A_n) = \sum_{n \geq 1} \mathbb{P}(X_1 > n) \leq \int_0^\infty \mathbb{P}(X_1 > x) dx = \mathbb{E}[X_1] < \infty$$

où l'on a utilisé la comparaison entre une somme et une intégrale, puis l'exercice apparaissant en dessous de la proposition 5.12. On en déduit par le lemme de Borel-Cantelli que, presque sûrement, $X_n \leq n$ pour tout n assez grand, et on conclut. \square

Ainsi, en notant $T_n = Y_1 + \dots + Y_n$, il suffit de montrer que T_n/n converge vers $\mathbb{E}[X_1]$ presque sûrement pour obtenir que S_n/n converge également vers $\mathbb{E}[X_1]$ p.s. À ce stade, on utilise une idée similaire à la preuve de la loi forte des grands nombres dans le cas L^2 : nous montrons la convergence de T_n/n le long d'une sous-suite. Cette fois, on fixe $\alpha > 1$ et on pose $k(n) = \lfloor \alpha^n \rfloor$. On a alors, par l'inégalité de Bienaymé-Chebychev,

$$\begin{aligned} \sum_{n \geq 1} \mathbb{P}(|T_{k(n)} - \mathbb{E}[T_{k(n)}]| > \varepsilon k(n)) &\leq \frac{1}{\varepsilon^2} \sum_{n \geq 1} \frac{\text{Var}(T_{k(n)})}{k(n)^2} \\ &\leq \frac{1}{\varepsilon^2} \sum_{n \geq 1} \frac{1}{k(n)^2} \sum_{m=1}^{k(n)} \text{Var}(Y_m) \\ &\leq \frac{1}{\varepsilon^2} \sum_{m \geq 1} \text{Var}(Y_m) \sum_{n: k(n) \geq m} \frac{1}{k(n)^2}. \end{aligned}$$

Comme $k(n) \geq \alpha^n / 2$ pour tout $n \geq 1$, on peut comparer la dernière somme à une série géométrique : si n_0 est le plus petit entier tel que $\lfloor \alpha^n \rfloor \geq m$

$$\sum_{n: k(n) \geq m} \frac{1}{k(n)^2} \leq 4 \sum_{n \geq n_0} \frac{1}{\alpha^{2n}} = \frac{4}{1 - \alpha^{-2}} \alpha^{-2n_0} \leq \frac{4}{1 - \alpha^{-2}} \cdot \frac{1}{m^2}.$$

et l'on obtient qu'il existe une constante C dépendant seulement de α et ε telle que

$$\sum_{n \geq 1} \mathbb{P}(|T_{k(n)} - \mathbb{E}[T_{k(n)}]| > \varepsilon k(n)) \leq C \sum_{m \geq 1} \frac{\text{Var}(Y_m)}{m^2}. \quad (7.2)$$

Montrons que cette dernière somme est finie. Pour cela, on écrit (en utilisant une nouvelle fois l'exercice après la proposition 5.12)

$$\text{Var}(Y_m) \leq \mathbb{E}[Y_m^2] = \int_0^\infty 2y \mathbb{P}(Y_m > y) dy = \int_0^m 2y \mathbb{P}(X_m > y) dy = \int_0^m 2y \mathbb{P}(X_1 > y) dy.$$

On en déduit

$$\begin{aligned} \sum_{m \geq 1} \frac{\text{Var}(Y_m)}{m^2} &\leq \sum_{m \geq 1} \frac{1}{m^2} \int_0^\infty 2y \mathbf{1}_{[0, m]}(y) \mathbb{P}(X_1 > y) dy \\ &\leq \int_0^\infty \left(\sum_{m = \lceil y \rceil}^\infty \frac{1}{m^2} \right) 2y \mathbb{P}(X_1 > y) dy \\ &\leq C' \int_0^\infty \mathbb{P}(X_1 > y) dy = C' \mathbb{E}[X_1] < \infty \end{aligned}$$

pour une constante universelle C' , et en utilisant que la somme de Riemann $\sum_{m \geq \lceil y \rceil} m^{-2}$ est équivalente à y^{-1} lorsque $y \rightarrow \infty$. En combinant cela avec (7.2), le lemme de Borel-Cantelli montre que presque sûrement, pour tout n assez grand, on a

$$\frac{|T_{k(n)} - \mathbb{E}[T_{k(n)}]|}{k(n)} \leq \varepsilon.$$

Comme cela est valide pour tout $\varepsilon > 0$ rationnel, on en déduit que $|T_{k(n)} - \mathbb{E}[T_{k(n)}]| / k(n)$ converge presque sûrement vers 0. Ensuite, on déduit que $T_{k(n)} / k(n)$ converge p.s. vers $\mathbb{E}[X_1]$, puisque par convergence dominée on a $\mathbb{E}[Y_n] \rightarrow \mathbb{E}[X_1]$, et donc par le lemme de Cesaro, $\mathbb{E}[T_{k(n)}] / k(n) \rightarrow \mathbb{E}[X_1]$. Enfin, si l'on se donne $n \geq 1$, soit $m = m(n)$ l'unique entier tel que $k(m) \leq n < k(m+1)$. On a alors

$$\frac{T_{k(m)}}{k(m)} \cdot \frac{k(m)}{k(m+1)} \leq \frac{T_n}{n} \leq \frac{T_{k(m+1)}}{k(m+1)} \cdot \frac{k(m+1)}{k(m)}.$$

En faisant tendre $n \rightarrow \infty$ on en déduit que presque sûrement,

$$\frac{1}{\alpha} \mathbb{E}[X_1] \leq \liminf_{n \rightarrow \infty} \frac{T_n}{n} \leq \limsup_{n \rightarrow \infty} \frac{T_n}{n} \leq \alpha \mathbb{E}[X_1].$$

Comme α peut être n'importe quel nombre rationnel strictement plus grand que 1, on conclut.

7.2.4 Le cas L^1 : une seconde preuve

Nous proposons maintenant une preuve entièrement différente du théorème 7.2. Fixons $a > \mathbb{E}[X_1]$ et notons

$$M_k = \max_{0 \leq n \leq k} (S_n - na).$$

Alors la suite $(M_k, k \geq 0)$ de variables aléatoires est clairement croissante, et converge presque sûrement vers une limite $M = \sup_{n \geq 0} (S_n - na)$ à valeurs dans $[0, \infty]$. Notons alors que pour tout $k \geq 0$, $\{M = \infty\} = \{\sup_{n \geq k} (X_{k+1} + X_{k+2} + \dots + X_n - na) = \infty\}$ est un événement mesurable par rapport à $\sigma(X_{k+1}, X_{k+2}, \dots)$. Ainsi, ce même événement appartient à la tribu asymptotique des variables aléatoires X_1, X_2, \dots . De ce fait, on a $\mathbb{P}(M = \infty) \in \{0, 1\}$ par la loi du 0-1 de Kolmogorov.

Supposons par l'absurde que l'on ait $\mathbb{P}(M = \infty) = 1$. On note alors $S'_0 = 0$ et $S'_n = X_2 + \dots + X_{n+1}$ pour tout $n \geq 1$, de sorte que par regroupement par paquets, la suite $(S'_n, n \geq 0)$ soit de même loi que $(S_n, n \geq 0)$, et indépendante de X_1 (puisque clairement mesurable par rapport à $\sigma(X_2, X_3, \dots)$). On a alors, pour tout $k \geq 1$,

$$\begin{aligned} M_k &= 0 \vee \max_{1 \leq n \leq k} (S_n - na) \\ &= 0 \vee \max_{0 \leq n \leq k-1} (S_{n+1} - (n+1)a) \\ &= 0 \vee \max_{0 \leq n \leq k-1} (X_1 - a + S'_n - na) \\ &= 0 \vee ((X_1 - a) + M'_{k-1}), \end{aligned}$$

où $M'_{k-1} = \max_{0 \leq n \leq k-1} (S'_n - na)$. Finalement, cela implique

$$M_k - M'_{k-1} = (X_1 - a) \vee (-M'_{k-1}).$$

Notons que M_k est clairement d'espérance finie comme maximum d'un nombre fini de variables aléatoires intégrables. De plus, M'_k a même loi que M_k (pour s'en convaincre, on peut écrire M_k comme une fonction mesurable de X_1, \dots, X_k , et constater que M'_k s'exprime comme la même fonction de X_2, \dots, X_{k+1}) et donc

$$\mathbb{E}[M_k - M'_{k-1}] = \mathbb{E}[M_k] - \mathbb{E}[M'_{k-1}] = \mathbb{E}[M_k] - \mathbb{E}[M_{k-1}] = \mathbb{E}[M_k - M_{k-1}] \geq 0,$$

la suite $(M_k, k \geq 0)$ étant croissante. D'un autre côté, l'on a que $(X_1 - a) \vee (-M'_{k-1})$ est dominée par $(X_1 - a)^+$, et converge presque sûrement vers $(X_1 - a) \vee (-M')$, où M' est la limite de M'_k lorsque $k \rightarrow \infty$. Mais notons que pour tout $x \in \mathbb{R}$,

$$\mathbb{P}(M' \leq x) = \lim_k \mathbb{P}(M'_k \leq x) = \lim_k \mathbb{P}(M_k \leq x) = \mathbb{P}(M \leq x) = 0$$

où l'on a utilisé le fait que les événements $\{M' \leq x\}$ et $\{M \leq x\}$ sont les réunions décroissantes des événements $\{M'_k \leq x\}$ et $\{M_k \leq x\}$ lorsque $k \rightarrow \infty$, le fait que M_k et M'_k ont même loi, et enfin le fait que $\mathbb{P}(M < \infty) = 0$ par hypothèse. On déduit que $\mathbb{P}(M' = \infty) = 1$ également, et donc $(X_1 - a) \vee (-M') = X_1 - a$ presque sûrement. De cela, on déduit par convergence dominée que

$$\mathbb{E}[(X_1 - a) \vee (-M'_{k-1})] \xrightarrow[k \rightarrow \infty]{} \mathbb{E}[X_1 - a] < 0,$$

par hypothèse sur a . Comme on a montré par ailleurs que ces espérances sont toutes positives, on a une contradiction. On en déduit que presque sûrement, pour tout a rationnel strictement plus grand que $\mathbb{E}[X_1]$, on a que $\sup_{n \geq 0} (S_n - na) < \infty$, et par conséquent

$$\limsup_{n \rightarrow \infty} \frac{S_n}{n} \leq a.$$

En faisant tendre a vers $\mathbb{E}[X_1]$, on déduit que $\limsup_{n \rightarrow \infty} S_n/n \leq \mathbb{E}[X_1]$, et quitte à changer X_n en $-X_n$, on déduit également que $\liminf_{n \rightarrow \infty} S_n/n \geq \mathbb{E}[X_1]$ p.s. D'où le résultat.

7.2.5 Quelques ramifications de la loi des grands nombres

Cas d'une espérance bien définie, mais infinie

La loi forte des grands nombres reste valable dès lors que $\mathbb{E}[X_1]$ est bien définie, et éventuellement infinie. En effet, si par exemple $\mathbb{E}[X_1^+] = \infty$ et $\mathbb{E}[X_1^-] < \infty$, et en écrivant que $S_n \geq (X_1^+ \wedge K + \dots + X_n^+ \wedge K) - (X_1^- + \dots + X_n^-)$, on voit bien que pour tout $K > 0$,

$$\liminf_{n \rightarrow \infty} \frac{S_n}{n} \geq \mathbb{E}[X_1^+ \wedge K] - \mathbb{E}[X_1^-]$$

presque sûrement. Lorsque $K \rightarrow \infty$, ce minorant converge vers $\mathbb{E}[X_1] = \infty$ par convergence monotone.

Cas où l'espérance n'existe plus nécessairement

Si X est une variable aléatoire intégrable, alors on a, pour tout $x > 0$,

$$\mathbb{P}(|X| \geq x) \leq \frac{\mathbb{E}[|X| \mathbf{1}_{\{|X| \geq x\}}]}{x},$$

ce qui s'obtient comme étape intermédiaire dans la preuve de l'inégalité de Markov. Comme le numérateur du majorant tend vers 0 quand $x \rightarrow \infty$ par convergence dominée (par $|X|$), on obtient que $\mathbb{P}(|X| \geq x) = o(1/x)$. En revanche, cette dernière condition n'implique pas en général que X soit intégrable ! On a néanmoins le résultat suivant, que nous énonçons sans preuve.

Théorème 7.8. *Soit X_1, X_2, \dots une suite i.i.d. de variables aléatoires, vérifiant*

$$x\mathbb{P}(X_1 \geq x) \xrightarrow{x \rightarrow \infty} 0.$$

Notons $S_n = X_1 + \dots + X_n$ et $m_n = \mathbb{E}[X_1 \mathbf{1}_{\{|X_1| \leq n\}}]$. Alors

$$\frac{S_n}{n} - m_n \xrightarrow[n \rightarrow \infty]{\mathbb{P}} 0.$$

7.3 Quelques applications

7.3.1 Marches aléatoires non centrées

On obtient comme conséquence immédiate de la loi forte des grands nombres qu'une marche aléatoire non centrée tend vers l'infini. En effet, soit X_1, X_2, \dots des variables aléatoires indépendantes et de même loi, dans L^1 , et $S_n = X_1 + \dots + X_n$.

Si $\mathbb{E}[X_1] \neq 0$, on a immédiatement que $|S_n| \rightarrow \infty$ presque sûrement lorsque $n \rightarrow \infty$ par la loi forte des grands nombres. On pourra mettre ce résultat en contraste avec le corollaire 6.17.

7.3.2 Approximation d'intégrales par la méthode de Monte-Carlo

Supposons que l'on ait à calculer l'intégrale suivante

$$I(f) = \int_{[0,1]^d} f(x) \lambda(dx)$$

où $f: [0, 1]^d \rightarrow \mathbb{R}$ est une fonction intégrable. Soit U_1, U_2, \dots une suite de variables aléatoires indépendantes de loi uniforme sur le cube $[0, 1]^d$. Alors les variables aléatoires $f(U_1), f(U_2), \dots$ sont indépendantes, intégrables et d'espérance $I(f)$. La loi des grands nombres garantit donc que

$$\lim_{n \rightarrow \infty} \frac{f(U_1) + \dots + f(U_n)}{n} = I(f).$$

Ceci fournit donc une méthode d'approximation d'une intégrale sur $[0, 1]^d$, puisque les variables aléatoires U_i sont aisées à simuler, dès lors que l'on dispose d'une fonction « rand » renvoyant (au moins théoriquement) une suite de variables aléatoires indépendantes uniformes sur $[0, 1]$. En effet, les coordonnées des variables aléatoires U_i sont elles-mêmes indépendantes et uniformes sur $[0, 1]$.

La loi des grands nombres ne dit cependant rien *a priori* sur la vitesse de convergence vers l'intégrale. On verra au chapitre suivant qu'elle est de l'ordre de $1/\sqrt{n}$. Cette vitesse est relativement mauvaise si l'on compare aux méthodes numériques usuelles disponibles pour $d = 1$ (méthode des trapèze, de Simpson, etc...). Cependant, la méthode présentée ici, dite *méthode de Monte Carlo*, présente plusieurs avantages :

- son efficacité ne dépend ni de la régularité de f , ni de la dimension
- si l'on connaît $I_n(f) = (f(U_1) + \dots + f(U_n))/n$, le calcul de $I_{n+1}(f)$ demande très peu d'opérations, là où des méthodes plus standard demanderaient de raffiner des partitions de $[0, 1]$.

Le premier point est particulièrement important, et de fait, la méthode de Monte Carlo et ses variantes est utilisée presque systématiquement pour estimer des intégrales de grande dimension.

Chapitre 8

Convergence en loi et théorème central limite

Au chapitre précédent, on a vu que la loi des grands nombres donnait un comportement asymptotique « au premier ordre » de la somme d'une suite de variables aléatoires. Par exemple, si X_1, X_2, \dots est une suite de variables aléatoires indépendantes, de loi de Bernoulli de paramètre $1/2$, alors on a presque sûrement, lorsque $n \rightarrow \infty$,

$$\frac{X_1 + \dots + X_n}{n} = \frac{1}{2} + o(1).$$

La limite est la constante déterministe $1/2$: un des aspects surprenants de ce résultat est l'émergence d'un « ordre » dans le « désordre » apparent d'une suite de variables aléatoires indépendantes (voir par exemple notre discussion sur l'apparition de tous les motifs finis dans la suite X_1, X_2, \dots). Il est alors légitime de se demander si l'on peut déterminer l'ordre supérieur de la convergence, c'est-à-dire expliciter le terme $o(1)$. C'est là que le caractère aléatoire refait surface : le théorème central limite stipule que pour n grand,

$$\frac{X_1 + \dots + X_n}{n} \stackrel{\text{loi}}{\simeq} \frac{1}{2} + \frac{N}{2\sqrt{n}}$$

où N est une variable aléatoire gaussienne centrée de variance 1. Le terme d'erreur $o(1)$ ci-dessus est donc intrinsèquement aléatoire. Cependant, nous allons devoir expliquer le sens de l'approximation ci-dessus.

8.1 Convergence étroite, convergence en loi

La convergence en loi est pour ainsi dire le quatrième mode de convergence fondamental utilisé en théorie des probabilités, avec ceux discutés dans le chapitre précédent. Elle tient pourtant une place à part, car elle concerne non pas les variables aléatoires à proprement parler, mais plutôt leurs lois.

Nous notons $\mathcal{C}_b(\mathbb{R}^d)$ l'espace des fonctions continues bornées sur \mathbb{R}^d et à valeurs dans \mathbb{R} .

Définition 8.1. Une suite (μ_n) de mesures de probabilités sur \mathbb{R}^d converge étroitement vers une mesure de probabilités μ sur \mathbb{R}^d si pour toute fonction $f \in \mathcal{C}_b(\mathbb{R}^d)$, on a

$$\int_{\mathbb{R}^d} f(x) \mu_n(dx) \xrightarrow{n \rightarrow \infty} \int_{\mathbb{R}^d} f(x) \mu(dx).$$

On dit qu'une suite (X_n) de variables aléatoires à valeurs dans \mathbb{R}^d converge en loi vers la variable aléatoire X sur \mathbb{R}^d si la suite (\mathbb{P}_{X_n}) converge étroitement vers \mathbb{P}_X .

Autrement dit, la suite (X_n) converge en loi vers X si pour toute fonction $f \in \mathcal{C}_b(\mathbb{R}^d)$ l'on a que

$$\mathbb{E}[f(X_n)] \xrightarrow[n \rightarrow \infty]{} \mathbb{E}[f(X)].$$

Comme remarqué plus haut, la convergence en loi de variables aléatoires est une propriété de leurs lois, plutôt que des variables aléatoires elles-mêmes. Ainsi, si (X_n) converge en loi vers X , alors (X_n) converge également en loi vers n'importe quelle variable aléatoire X' de même loi que X !

On notera respectivement

$$\mu_n \xrightarrow{(e)} \mu, \quad X_n \xrightarrow{\text{loi}} X,$$

pour dire qu'une suite de mesures de probabilités converge étroitement vers μ (respectivement, qu'une suite de variables aléatoires converge en loi vers X).

8.1.1 Exemples élémentaires

Lois sur \mathbb{N}

Proposition 8.2. Une suite (μ_n) de mesures de probabilités sur \mathbb{N} converge étroitement vers la mesure de probabilités μ sur \mathbb{N} si et seulement si $\mu_n(k) \rightarrow \mu(k)$ pour tout $k \in \mathbb{N}$. De façon équivalente, une suite de variables aléatoires (X_n) à valeurs dans \mathbb{N} converge en loi vers la variable aléatoire X à valeurs dans \mathbb{N} , si et seulement si

$$\mathbb{P}(X_n = k) \xrightarrow[n \rightarrow \infty]{} \mathbb{P}(X = k) \quad \text{pour tout } k \in \mathbb{N}. \quad (8.1)$$

Démonstration. Nous montrons la deuxième formulation de la proposition, en termes de convergence en loi. La condition nécessaire est facile à vérifier. Si (X_n) converge en loi vers X et si $f(x) = (1 - |x - k|)_+$, qui est une fonction continue bornée valant 1 en un entier k et 0 en tout autre entier, on a bien

$$\mathbb{P}(X_n = k) = \mathbb{E}[f(X_n)] \xrightarrow{} \mathbb{E}[f(X)] = \mathbb{P}(X = k)$$

lorsque $n \rightarrow \infty$.

Montrons la condition suffisante, en supposant (8.1). Soit f une fonction continue et bornée sur \mathbb{R} . Alors

$$\mathbb{E}[f(X)] = \sum_{k \in \mathbb{N}} f(k) \mathbb{P}(X = k) = \sum_{k=0}^N f(k) \mathbb{P}(X = k) + R_N$$

où R_N vérifie

$$|R_N| \leq \|f\|_\infty \mathbb{P}(X > N) \xrightarrow[n \rightarrow \infty]{} 0.$$

Bien sûr, $\mathbb{E}[f(X_n)]$ satisfait une propriété analogue. Soit $\varepsilon > 0$. Choisissons $N > 0$ tel que $|R_N| < \varepsilon / (2\|f\|_\infty)$. Notons que l'on a

$$\mathbb{P}(X_n > N) = 1 - \mathbb{P}(X_n \leq N) = 1 - \sum_{k=0}^N \mathbb{P}(X_n = k) \xrightarrow[n \rightarrow \infty]{} 1 - \sum_{k=0}^N \mathbb{P}(X = k) = \mathbb{P}(X > N),$$

et par conséquent, on il existe n_0 tel que pour tout $n \geq n_0$, on ait $\mathbb{P}(X_n > N) < \varepsilon / (2\|f\|_\infty)$. Pour un tel choix de N et n_0 , on a alors

$$|\mathbb{E}[f(X_n)] - \mathbb{E}[f(X)]| \leq \left| \sum_{k=0}^N f(k)\mathbb{P}(X_n = k) - \sum_{k=0}^N f(k)\mathbb{P}(X = k) \right| + \varepsilon,$$

et donc la limite supérieure du membre de gauche est majorée par ε . On conclut. \square

Par exemple, si $\text{Poi}(\theta)$ est la loi de Poisson de paramètre θ , et si la suite numérique (θ_n) converge vers $\theta > 0$, alors $\text{Poi}(\theta_n) \xrightarrow{(e)} \text{Poi}(\theta)$.

Lemme de Scheffé et convergence ponctuelle de densités

La proposition suivante est une sorte d'analogie « continu » de la proposition précédente.

Proposition 8.3. *Soit (f_n) une suite de densités de probabilités sur \mathbb{R}^d . On suppose que $f_n(x) \rightarrow f(x)$ pour presque tout $x \in \mathbb{R}^d$, où f est une densité de probabilités. Alors*

$$f_n(x)dx \xrightarrow[n \rightarrow \infty]{(e)} f(x)dx.$$

Démonstration. Sous les hypothèses de la proposition, on a que $f_n \rightarrow f$ dans $L^1(\mathbb{R}^d)$. En effet, ceci s'obtient facilement en appliquant le lemme de Fatou à la suite de fonctions positives

$$h_n = f_n + f - |f_n - f| = 2f \wedge f_n.$$

On obtient ainsi que, sous nos hypothèses,

$$\liminf \int_{\mathbb{R}^d} h_n = 2 \int_{\mathbb{R}^d} f - \limsup \int_{\mathbb{R}^d} |f_n - f| \geq 2 \int_{\mathbb{R}^d} \liminf f_n \wedge f = 2 \int_{\mathbb{R}^d} f.$$

En soustrayant on voit que $\limsup \int_{\mathbb{R}^d} |f_n - f| = 0$.

Ensuite, soit g une fonction continue bornée sur \mathbb{R}^d . On a

$$\left| \int_{\mathbb{R}^d} g(x)f_n(x)dx - \int_{\mathbb{R}^d} g(x)f(x)dx \right| \leq \|g\|_\infty \|f - f_n\|_1 \xrightarrow[n \rightarrow \infty]{} 0.$$

\square

Ainsi, on voit par exemple que les lois exponentielles de paramètre θ_n convergent étroitement vers la loi exponentielle de paramètre $\theta > 0$ dès lors que $\theta_n \rightarrow \theta$.

Remarque. La convergence étroite est une notion de convergence qui provient d'une topologie (dite topologie étroite) sur l'ensemble des mesures de probabilités sur \mathbb{R}^d . Si l'on voit ce dernier comme un sous-espace du dual de l'espace $(\mathcal{C}_b(\mathbb{R}^d), \|\cdot\|_\infty)$, une mesure de probabilités μ étant clairement associée à la forme linéaire continue sur $\mathcal{C}_b(\mathbb{R}^d)$

$$f \mapsto \int_{\mathbb{R}^d} f(x)\mu(dx),$$

alors la topologie étroite est la restriction à ce sous-espace de la topologie dite faible-*

Terminons avec un exemple « mixte », illustrant comment les lois continues peuvent être approchées étroitement par des lois discrètes.

Exemple d'approximation de la mesure de Lebesgue

Pour illustrer la notion de convergence étroite, notons que si

$$\mu_n(dx) = \frac{1}{n+1} \sum_{k=0}^n \delta_{k/n},$$

alors μ_n converge étroitement vers la mesure de Lebesgue sur $[0, 1]$. En effet, si f est une fonction continue bornée, l'intégrale $\int_{\mathbb{R}} f(x) \mu_n(dx)$ est une somme de Riemann de f , qui converge vers $\int_0^1 f(x) dx$. Ainsi, une variable aléatoire uniforme sur $\{0, 1/n, 2/n, \dots, n/n\}$ converge en loi vers une variable aléatoire uniforme sur $[0, 1]$.

8.1.2 Liens avec les autres notions de convergence.

Le lien avec les notions de convergences déjà étudiées est donné par les résultats suivants.

Convergence en probabilité

Proposition 8.4. *Soit (X_n) une suite de variables aléatoires convergeant en probabilité vers X . Alors X_n converge en loi vers X .*

Démonstration. Supposons que (X_n) converge en probabilité vers X . Supposons par l'absurde que X_n ne converge pas en loi vers X , et donc qu'il existe une fonction f continue bornée, et un $\varepsilon > 0$, tels que

$$|\mathbb{E}[f(X_n)] - \mathbb{E}[f(X)]| > \varepsilon$$

pour tout n dans un ensemble $A = \{n_1, n_2, \dots\}$ infini, avec $n_1 < n_2 < \dots$

Or on sait que l'on peut extraire une suite extraite de (X_{n_k}) qui converge p.s. vers X . Comme f est continue et bornée, le théorème de convergence dominée montre alors que $\mathbb{E}[f(X_{n_k})] \rightarrow \mathbb{E}[f(X)]$ le long de cette sous-suite, ce qui est évidemment absurde. \square

Il existe une situation où la réciproque est vraie.

Proposition 8.5. *Si la suite de variables aléatoires (X_n) converge en loi vers une variable aléatoire constante p.s. (c'est-à-dire que sa loi est une masse de Dirac), alors elle converge en probabilité.*

Démonstration. Supposons que $X_n \xrightarrow{\text{loi}} c$ où $c \in \mathbb{R}^d$ est une constante. Fixons $\varepsilon > 0$. Posons $f(x) = \min(|(x - c)/\varepsilon|, 1)$, de sorte que f est continue, bornée, nulle en c , et vérifiant $\mathbf{1}_{\{|x-c|>\varepsilon\}} \leq f(x)$ pour tout x . Alors

$$\mathbb{P}(|X_n - c| > \varepsilon) = \mathbb{E}[\mathbf{1}_{\{|X_n - c| > \varepsilon\}}] \leq \mathbb{E}[f(X_n)] \xrightarrow{n \rightarrow \infty} \mathbb{E}[f(c)] = 0. \quad \square$$

Remarque. Cette dernière propriété peut paraître un peu surprenante au premier abord : en effet, la convergence en loi ne dépend pas de l'espace de probabilités sur lequel on se place, au contraire de la convergence en probabilité. La subtilité est que la probabilité $\mathbb{P}(|X_n - X| > \varepsilon)$ ne dépend, dans le cas où X est une constante, que de la loi de X_n !

Convergence en variation totale

Comme il a été vu au cours d'intégration du premier semestre, il y a, en plus de la topologie étroite, une autre topologie naturelle sur les mesures de probabilités sur \mathbb{R}^d . En effet, on peut voir ces dernières comme un sous-ensemble convexe fermé de l'espace de Banach des mesures signées sur \mathbb{R}^d , muni de la norme de variation totale $\|\cdot\|$. Plus précisément, supposons que ν soit une mesure signée sur \mathbb{R}^d , de masse totale $\nu(\mathbb{R}^d) = 0$. Alors pour tout $A \in \mathcal{B}(\mathbb{R}^d)$, on a $\nu(A) + \nu(A^c) = \nu(\mathbb{R}^d) = 0$, et donc

$$2|\nu(A)| = |\nu(A) - \nu(A^c)| \leq |\nu(A)| + |\nu(A^c)| \leq \|\nu\|,$$

par définition de la variation totale. Mais d'autre part, rappelons que la décomposition de Jordan de ν exprime qu'il existe $B \in \mathcal{B}(\mathbb{R}^d)$ tel que $\nu(\cdot \cap B)$ et $-\nu(\cdot \cap B^c)$ sont deux mesures positives finies, et que $\|\nu\| = \nu(B) - \nu(B^c) = 2\nu(B)$. De cela, on déduit que

$$\|\nu\| = 2 \sup_{A \in \mathcal{B}(\mathbb{R}^d)} |\nu(A)|.$$

Ainsi, la suite (μ_n) de mesures de probabilités sur \mathbb{R}^d converge vers μ en norme de variation totale si et seulement si $|\mu_n(A) - \mu(A)|$ converge vers 0 uniformément en $A \in \mathcal{B}(\mathbb{R}^d)$. Ceci illustre le fait que la convergence en norme de variation totale est beaucoup plus « rigide » que la convergence étroite.

Dans le dernier exemple du paragraphe précédent, on a clairement que $\|\mu_n - \mu\|$ ne converge pas vers 0, puisque par exemple $\mu_n(\mathbb{Q}) = 1$ et $\mu(\mathbb{Q}) = 0$.

8.1.3 Caractérisations de la convergence en loi

La définition de la convergence en loi ne fournit pas vraiment de critère pratique pour montrer qu'une suite de variables aléatoires (X_n) donnée converge en loi. Nous allons donc donner plusieurs formulations équivalentes.

Théorème 8.6. *Soit (X_n) et X des variables aléatoires à valeurs dans \mathbb{R}^d . Les propositions suivantes sont équivalentes.*

1. La suite (X_n) converge en loi vers X
2. pour toute fonction f bornée et 1-lipschitzienne, on a $\lim_n \mathbb{E}[f(X_n)] = \mathbb{E}[f(X)]$
3. pour tout ouvert O de \mathbb{R}^d , on a $\liminf_n \mathbb{P}(X_n \in O) \geq \mathbb{P}(X \in O)$
4. pour tout fermé F de \mathbb{R}^d , on a $\limsup_n \mathbb{P}(X_n \in F) \leq \mathbb{P}(X \in F)$
5. pour tout $A \in \mathcal{B}(\mathbb{R}^d)$ tel que $\mathbb{P}(X \in \partial A) = 0$, on a $\lim_n \mathbb{P}(X_n \in A) = \mathbb{P}(X \in A)$
6. pour toute fonction $f: \mathbb{R}^d \rightarrow \mathbb{R}$ mesurable bornée, continue \mathbb{P}_X -presque partout, on a $\lim_n \mathbb{E}[f(X_n)] = \mathbb{E}[f(X)]$.

Notons que ce théorème admet une reformulation en termes de convergence étroite de mesures de probabilités : on remplacera $\mathbb{E}[f(X)]$ par $\int f d\mu$, $\mathbb{P}(X \in A)$ par $\mu(A)$, \mathbb{P}_X par μ , etc.

Démonstration. Les implications 1. \implies 2. et 6. \implies 1. sont évidentes, de même que l'équivalence entre 3. et 4. par un simple passage au complémentaire.

Montrons que 2. \implies 4. Soit donc F un fermé de E . Pour tout $K > 0$, on considère la fonction

$$f_{F,K}(x) = (1 - Kd(x, F))_+,$$

qui est lipschitzienne et vérifie $\mathbf{1}_F \leq f_{F,K} \leq 1$. Ainsi, pour tout $n \geq 1$ on a $\mathbb{P}(X_n \in F) \leq \mathbb{E}[f_{F,K}(X_n)]$, et comme on a supposé 2., on en déduit que

$$\limsup \mathbb{P}(X_n \in F) \leq \mathbb{E}[f_{F,K}(X)].$$

Comme $f_{F,K}$ converge vers $\mathbf{1}_F$ ponctuellement et est bornée par 1, on en déduit par convergence dominée que $\limsup_n \mathbb{P}(X_n \in F) \leq \mathbb{P}(X \in F)$.

Montrons que 3. et 4. impliquent 5. Soit donc $A \in \mathcal{B}(\mathbb{R}^d)$. On applique 3. et 4. aux ensembles A° et \bar{A} (intérieur et adhérence de A), et on trouve

$$\mathbb{P}(X \in A^\circ) \leq \liminf \mathbb{P}(X_n \in A^\circ) \leq \limsup \mathbb{P}(X_n \in \bar{A}) \leq \mathbb{P}(X \in \bar{A}).$$

Mais si l'on a $\mathbb{P}(X \in \partial A) = 0$, alors $\mathbb{P}(X \in A^\circ) = \mathbb{P}(X \in \bar{A}) = \mathbb{P}(X \in A)$, et l'on obtient ce que l'on voulait.

Montrons enfin que 5. \implies 6. Soit donc f une fonction continue \mathbb{P}_X -presque partout et bornée. Sans perte de généralité, on peut supposer que f est positive (on peut en effet écrire $f = f_+ - f_-$ et raisonner sur chaque terme). Soit D l'ensemble des points de discontinuité de f . Notons d'abord que pour toute mesure de probabilités ν sur \mathbb{R}^d , on a

$$\int_{\mathbb{R}^d} f(x) \nu(dx) = \int_0^\infty \nu(\{f \geq y\}) dy. \quad (8.2)$$

C'est en effet une conséquence immédiate du théorème de Fubini, en écrivant l'intégrale

$$\int_{\mathbb{R}^d} \nu(dx) \int_0^\infty \mathbf{1}_{[0, f(x)]}(y) dy = \int_0^\infty dy \int_{\mathbb{R}^d} \nu(dx) \mathbf{1}_{\{f(x) \geq y\}}.$$

Par ailleurs, pour tout $y \geq 0$, notons $A_y = \{x: f(x) \geq y\}$. Soit $x \in \bar{A}_y$, de sorte que x est limite d'une suite x_n telle que $f(x_n) \geq y$. Si $x \notin D$, c'est-à-dire si x est point de continuité de f , alors on a aussi $f(x) \geq y$. Donc $\bar{A}_y \subseteq A_y \cup D$. Par ailleurs, si $f(x) > y$ et $x \notin D$, alors on a également $f(x') > y$ pour x' dans un voisinage de x . Donc $\{f > y\} \setminus D \subseteq A_y^\circ$. Finalement, on en déduit que $\partial A_y \subseteq \{f = y\} \cup D$.

Par ailleurs, l'ensemble $\{y \geq 0: \mathbb{P}_X(\{f = y\}) > 0\}$ est au plus dénombrable. Il est en effet la réunion des ensembles $\{y \geq 0: \mathbb{P}_X(\{f = y\}) \geq 1/r\}$, $r \geq 1$, qui sont respectivement de cardinal au plus r , puisque les ensembles $\{f = y\}$ sont deux-à-deux disjoints. Par 6., on en déduit que pour Lebesgue-presque tout $y \geq 0$, on a $\mathbb{P}(X_n \in A_y) \rightarrow \mathbb{P}(X \in A_y)$ quand $n \rightarrow \infty$. Donc par convergence dominée, en utilisant (?) et le fait que f est bornée, on a

$$\begin{aligned} \mathbb{E}[f(X_n)] &= \int_{\mathbb{R}^d} f(x) \mathbb{P}_{X_n}(dx) &= \int_0^\infty \mathbb{P}(X_n \in A_y) dy \\ & &= \int_0^{\|f\|_\infty} \mathbb{P}(X_n \in A_y) dy \\ & \xrightarrow{n \rightarrow \infty} \int_0^{\|f\|_\infty} \mathbb{P}(X \in A_y) dy &= \mathbb{E}[f(X)], \end{aligned}$$

comme voulu. □

Remarque. Ce théorème est souvent appelé « théorème du porte-manteau », ce qui peut paraître surprenant. Il semble que cela soit dû au nombre et à la variété des énoncés équivalents qui apparaissent dans son énoncé, comme autant d'habits appartenant à la même notion. Pour rendre les choses encore un peu plus confuses, le mathématicien Patrick Billingsley, dans l'édition de 1999 de son livre *Convergence of probability measures*, s'est permis la facétie d'attribuer ce résultat à un mathématicien imaginaire du nom de Jean-Pierre Portmanteau, citant un article tout aussi imaginaire de 1915... En réalité, l'énoncé semble remonter à Alexandrov dans les années 1940.

On a également la possibilité de restreindre la classe des fonctions-test. Notons $\mathcal{C}_c(\mathbb{R}^d)$ l'ensemble des fonctions continues à support compact définies sur \mathbb{R}^d .

Proposition 8.7. *Soit H un sous-ensemble de fonctions mesurables bornées définies sur \mathbb{R}^d , et dont l'adhérence pour la norme uniforme contient $\mathcal{C}_c(\mathbb{R}^d)$. Soit (X_n) et X des variables aléatoires à valeurs dans \mathbb{R}^d . Si l'on a que $\mathbb{E}[f(X_n)] \rightarrow \mathbb{E}[f(X)]$ pour tout $f \in H$, alors $X_n \xrightarrow{\text{loi}} X$.*

Démonstration. Supposons dans un premier temps que les hypothèses du théorème sont vérifiées avec $H = \mathcal{C}_c(\mathbb{R}^d)$. Pour $r > 0$, soit

$$\chi_r(x) = 0 \vee (r + 1 - |x|) \wedge 1,$$

de sorte que χ_r est positive, continue, bornée par 1, égale à 1 sur $B_{\mathbb{R}^d}(0, r)$ et nulle hors de $B_{\mathbb{R}^d}(0, r + 1)$. Si $f \in \mathcal{C}_b(\mathbb{R}^d)$ alors $f\chi_r \in \mathcal{C}_c(\mathbb{R}^d)$, et donc $\mathbb{E}[(f\chi_r)(X_n)] \rightarrow \mathbb{E}[(f\chi_r)(X)]$ par hypothèse. On a alors

$$\begin{aligned} |\mathbb{E}[f(X_n)] - \mathbb{E}[f(X)]| &\leq |\mathbb{E}[(f\chi_r)(X_n)] - \mathbb{E}[(f\chi_r)(X)]| \\ &\quad + \|f\|_\infty (\mathbb{E}[(1 - \chi_r)(X_n)] + \mathbb{E}[(1 - \chi_r)(X)]) \end{aligned}$$

Pour conclure, notons que $\mathbb{E}[\chi_r(X_n)] \rightarrow \mathbb{E}[\chi_r(X)]$, et donc

$$\limsup_{n \rightarrow \infty} |\mathbb{E}[f(X_n)] - \mathbb{E}[f(X)]| \leq 2\|f\|_\infty \mathbb{E}[(1 - \chi_r)(X)] \leq 2\|f\|_\infty \mathbb{P}(|X| > r).$$

La quantité de droite converge vers 0 lorsque $r \rightarrow \infty$, et on conclut dans ce cas.

Dans le cas général où H est un ensemble de fonctions mesurables dense dans $\mathcal{C}_c(\mathbb{R}^d)$, donnons-nous $f \in \mathcal{C}_c(\mathbb{R}^d)$, et $g \in H$ telle que $\|f - g\|_\infty \leq \varepsilon/2$. Alors, comme on a $|\mathbb{E}[f(X)] - \mathbb{E}[g(X)]| \leq \|f - g\|_\infty$, et de même avec X_n à la place de X , on déduit que

$$\limsup_{n \rightarrow \infty} |\mathbb{E}[f(X_n)] - \mathbb{E}[f(X)]| \leq \varepsilon + \limsup_{n \rightarrow \infty} |\mathbb{E}[g(X_n)] - \mathbb{E}[g(X)]| = \varepsilon$$

par hypothèse. On en conclut que $\mathbb{E}[f(X_n)] \rightarrow \mathbb{E}[f(X)]$ pour tout $f \in \mathcal{C}_c(\mathbb{R}^d)$, et on conclut par la première partie de la preuve. \square

On déduit des résultats précédents un critère de convergence en loi de variables aléatoires réelles faisant intervenir les fonctions de répartition.

Corollaire 8.8. *La suite (X_n) de variables aléatoires réelles converge vers la variable aléatoire réelle X si et seulement si $F_{X_n}(x) \rightarrow F_X(x)$ pour tout x qui est un point de continuité de F_X , c'est-à-dire que $\mathbb{P}(X = x) = 0$.*

Démonstration. Supposons que $X_n \xrightarrow{\text{loi}} X$. Soit x un point de continuité de F_X . On applique le point 5. du théorème 8.6 à l'ensemble $A =]-\infty, x]$, dont la frontière $\partial A = \{x\}$ n'est pas chargée par \mathbb{P}_X puisque $\mathbb{P}_X(\{x\}) = \mathbb{P}(X = x)$. On obtient bien que

$$\lim_{n \rightarrow \infty} \mathbb{P}(X_n \leq x) = \mathbb{P}(X \leq x).$$

Réciproquement, supposons que $F_{X_n} \rightarrow F_X$ en tout point de continuité de F_X . Notons D l'ensemble des points de discontinuité de F_X . Alors D est au plus dénombrable. De plus, si $a, b \notin D$ et $a < b$, on a que

$$\mathbb{P}(X_n \in]a, b]) = F_{X_n}(b) - F_{X_n}(a) \longrightarrow F_X(b) - F_X(a) = \mathbb{P}(X \in]a, b]).$$

Soit H l'espace vectoriel engendré par les fonctions $\mathbf{1}_{]a, b]}$ avec $a, b \in \mathbb{R} \setminus D$. Alors la convergence ci-dessus s'étend, par linéarité de l'espérance, à $\mathbb{E}[f(X_n)] \longrightarrow \mathbb{E}[f(X)]$ pour tout $f \in H$. Comme H est dense dans $\mathcal{C}_c(\mathbb{R}^d)$ pour la norme uniforme, on conclut par la proposition 8.7. \square

Enfin, nous montrons que la convergence en loi peut être formulée en termes des fonctions caractéristiques.

Théorème de Lévy. *La suite de mesures de probabilités (μ_n) sur \mathbb{R}^d converge étroitement vers la mesure de probabilités μ si et seulement si pour tout $\xi \in \mathbb{R}^d$, on a*

$$\hat{\mu}_n(\xi) \xrightarrow[n \rightarrow \infty]{} \hat{\mu}(\xi).$$

Autrement dit, la suite de variables aléatoires (X_n) converge en loi vers X si et seulement si, pour tout $\xi \in \mathbb{R}^d$, on a

$$\varphi_{X_n}(\xi) \xrightarrow[n \rightarrow \infty]{} \varphi_X(\xi).$$

Démonstration. La condition nécessaire est facile : si $\mu_n \xrightarrow{(e)} \mu$, alors, comme pour tout $\xi \in \mathbb{R}^d$, la fonction $x \mapsto e^{i\xi x}$ est continue et bornée, on a bien que $\hat{\mu}_n(\xi) = \int e^{i\xi x} \mu_n(dx)$ converge vers $\hat{\mu}(\xi) = \int e^{i\xi x} \mu(dx)$.

Pour le sens réciproque, supposons que $\hat{\mu}_n \rightarrow \hat{\mu}$ ponctuellement. Soit $f \in \mathcal{C}_c(\mathbb{R}^d)$. Si f était la transformée de Fourier d'une fonction φ , l'on pouvait écrire

$$\int_{\mathbb{R}^d} f(x) \mu_n(dx) = \int_{\mathbb{R}^d} \varphi(\xi) \hat{\mu}_n(\xi) d\xi$$

par la formule de réciprocity de la transformation de Fourier, et essayer de passer à la limite dans l'intégrale. L'idée est donc de remplacer f par une fonction proche, qui soit une transformée de Fourier. Soit $\sigma > 0$, et soit g_σ la densité gaussienne de la loi $\mathcal{N}(0, \sigma^2)$. La transformée de Fourier de $g_\sigma * f$ est $(2\pi/\sigma^2)^{d/2} g_{1/\sigma} \hat{f}$, qui est dans L^1 . En effet, le fait que f soit à support compact implique que \hat{f} est bornée, et $g_{1/\sigma}$ est dans L^1 . On note $R\varphi = (2\pi/\sigma^2)^{d/2} g_{1/\sigma} \hat{f}$ cette fonction, où l'on rappelle la notation $R\varphi(x) = \varphi(-x)$. La formule d'inversion de Fourier implique que $\hat{\varphi} = f$, et l'on a donc par la discussion ci-dessus que

$$\int_{\mathbb{R}^d} g_\sigma * f(x) \mu_n(dx) = \int_{\mathbb{R}^d} \varphi(\xi) \hat{\mu}_n(\xi) d\xi.$$

Comme $|\hat{\mu}_n(\xi)| \leq 1$ et $|\varphi(\xi)| \leq (2\pi/\sigma^2)^{d/2} \|\hat{f}\|_\infty g_{1/\sigma}(\xi)$ est intégrable, la convergence dominée montre que ceci converge lorsque $n \rightarrow \infty$ vers

$$\int_{\mathbb{R}^d} \varphi(\xi) \hat{\mu}(\xi) d\xi = \int_{\mathbb{R}^d} g_\sigma * f(x) \mu(dx),$$

où l'on a utilisé à nouveau la formule de réciprocity.

Nous avons donc obtenu que $\int h d\mu_n \rightarrow \int h d\mu$ pour toute fonction h dans l'ensemble $H = \{g_\sigma * f : f \in \mathcal{C}_c(\mathbb{R}^d), \sigma > 0\}$. Comme on a que $g_\sigma * f$ converge vers f uniformément pour tout $f \in \mathcal{C}_c(\mathbb{R}^d)$ par la proposition 1.9, on obtient que H est dense dans $\mathcal{C}_c(\mathbb{R}^d)$ pour la norme uniforme, et on conclut par la proposition 8.7. \square

8.2 Le théorème central limite

Si X_1, X_2, \dots sont des variables aléatoires i.i.d. dans L^1 , la loi des grands nombres stipule que S_n/n converge vers $\mathbb{E}[X_1]$ p.s. Comme on l'a dit au début du chapitre, on cherche à donner un développement à l'ordre supérieur dans cette convergence, en donnant la vitesse à laquelle la quantité $(S_n/n) - \mathbb{E}[X_1]$ converge vers 0. Notons que si les variables aléatoires considérées sont en fait dans L^2 , on a que la variance de S_n est de $n \text{Var}(X_1)$, ce qui indique que la distance à la moyenne de S_n est typiquement de l'ordre de \sqrt{n} . Ceci justifie la renormalisation choisie dans l'énoncé suivant.

Théorème central limite. *Soit (X_1, X_2, \dots) une suite de variables aléatoires réelles dans L^2 , indépendantes et de même loi. On suppose que $\sigma^2 = \text{Var}(X_1) > 0$. Soit $S_n = X_1 + \dots + X_n$. Alors on a*

$$\frac{S_n - n\mathbb{E}[X_1]}{\sigma\sqrt{n}} \xrightarrow[n \rightarrow \infty]{\text{loi}} \mathcal{N}(0, 1).$$

Démonstration. Quitte à remplacer X_n par $X_n - \mathbb{E}[X_1]$, on suppose sans perte de généralité que les variables aléatoires sont centrées. Sous l'hypothèse que $X_1 \in L^2$, la fonction caractéristique φ_X est de classe $\mathcal{C}^2(\mathbb{R}, \mathbb{C})$, et l'on a, par la formule de Taylor-Young,

$$\varphi_X(\xi) = 1 - \frac{\sigma^2}{2} \xi^2 + o(\xi^2).$$

Calculons alors la fonction caractéristique de $S_n/\sigma\sqrt{n}$: comme les $X_j, j \geq 1$ sont indépendantes et de même loi,

$$\varphi_{S_n/\sigma\sqrt{n}}(\xi) = \mathbb{E}[e^{i(X_1 + \dots + X_n)\xi/\sigma\sqrt{n}}] = \prod_{j=1}^n \mathbb{E}[e^{iX_j\xi/\sigma\sqrt{n}}] = \varphi_{X_1}(\xi/\sigma\sqrt{n})^n.$$

Donc, pour tout $\xi \in \mathbb{R}$,

$$\varphi_{S_n/\sigma\sqrt{n}}(\xi) = \left(1 - \frac{\xi^2}{2n} + o\left(\frac{1}{n}\right)\right)^n.$$

Pour n assez grand la quantité entre parenthèses est dans la boule ouverte de rayon 1 centrée en 1, et on peut prendre la détermination principale du logarithme (ne pas oublier que la quantité entre parenthèses est un nombre complexe)

$$\varphi_{S_n/\sigma\sqrt{n}}(\xi) = \exp\left(n \operatorname{Log}\left(1 - \frac{\xi^2}{2n} + o\left(\frac{1}{n}\right)\right)\right) \xrightarrow{n \rightarrow \infty} \exp\left(-\frac{\xi^2}{2}\right).$$

On reconnaît la transformée de Fourier de la densité gaussienne standard g_1 , et on conclut par le théorème de Lévy. \square

Voici comment Galton décrit l'impression qu'exerce sur lui ce théorème.

I know of scarcely anything so apt to impress the imagination as the wonderful form of cosmic order expressed by the "Law of Frequency of Error". The law would have been personified by the Greeks and deified, if they had known of it. It reigns with serenity and in complete self-effacement, amidst the wildest confusion. The huger the mob, and the greater the apparent anarchy, the more perfect is its sway. It is the supreme law of Unreason.

Les Grecs l'auraient déifiée ! Pourquoi cette fascination ? Entre autres, du fait du caractère *universel* de la loi gaussienne qui est révélé par ce théorème. Sous la simple hypothèse de l'existence d'un moment d'ordre 2, c'est toujours la loi gaussienne qui régit les fluctuations de la somme d'une suite de variables aléatoires indépendantes et de même loi. Il y a aussi une forme de « miracle » dans le fait que la densité de la loi qui apparaît ainsi soit explicite, et aussi simple.

Expliquons davantage la signification de ce théorème. Par la caractérisation de la convergence en loi par les fonctions caractéristiques, et comme la loi gaussienne est diffuse et admet donc une fonction caractéristique continue en tout point, on voit que le théorème est équivalent à dire que pour tout $x \in \mathbb{R}$,

$$\mathbb{P}\left(\frac{S_n - n\mathbb{E}[X_1]}{\sigma\sqrt{n}} \leq x\right) \xrightarrow{n \rightarrow \infty} \int_{-\infty}^x \frac{e^{-y^2/2}}{\sqrt{2\pi}} dy,$$

ou encore, que pour tout $a, b \in \mathbb{R}$ avec $a < b$,

$$\mathbb{P}\left(a \leq \frac{S_n - n\mathbb{E}[X_1]}{\sigma\sqrt{n}} \leq b\right) \xrightarrow{n \rightarrow \infty} \int_a^b \frac{e^{-y^2/2}}{\sqrt{2\pi}} dy.$$

Application aux statistiques : estimation paramétrique et intervalles de confiance

Cette reformulation a une importance cruciale en statistiques. Pour fixer les idées, la problématique de base de la statistique inférentielle est, étant donnée une réalisation donnée d'une suite de variables aléatoires i.i.d. $X_1(\omega), X_2(\omega), \dots, X_n(\omega)$, de déterminer avec la meilleure précision possible la loi *inconnue* de ces variables aléatoires, ou certaines fonctions naturelles de cette loi.

Par exemple, si l'on veut connaître l'espérance m de X_1 (en supposant qu'elle existe), il est naturel de l'*estimer* à l'aide de la moyenne empirique des observations

$$m_n(\omega) = \frac{X_1(\omega) + \dots + X_n(\omega)}{n},$$

dont on sait qu'elle converge pour presque tout ω vers $\mathbb{E}[X_1]$. On dit que m_n est un estimateur *consistant* de m .

Sous l'hypothèse que X_1 est dans L^2 , le théorème central limite détermine l'erreur asymptotique que l'on commet en assimilant m_n à m . Plus précisément, pour tout $x > 0$ on a

$$\mathbb{P}\left(|m_n - m| > \frac{\sigma x}{\sqrt{n}}\right) \xrightarrow{n \rightarrow \infty} 2 \int_x^\infty \frac{e^{-y^2/2}}{\sqrt{2\pi}} dy. \quad (8.3)$$

Si l'on choisit $x = 1,96$, alors le membre de droite est inférieur à 0,05. Ceci signifie que pour n assez grand, l'erreur que l'on commet en assimilant m_n à m sera au plus de $1,96 \cdot \sigma / \sqrt{n}$, avec probabilité supérieure à 95%. On dit que l'intervalle

$$\left[m_n(\omega) - \frac{1,96 \cdot \sigma}{\sqrt{n}}, m_n(\omega) + \frac{1,96 \cdot \sigma}{\sqrt{n}} \right] \quad (8.4)$$

est un *intervalle de confiance* (asymptotique, bilatère) pour m à 95%. Noter que l'on a ici deux niveaux d'incertitude, ce qui est une situation typique en statistiques :

- on ne peut évidemment pas prétendre déterminer la valeur exacte de m à l'aide d'un nombre fini d'observations, mais seulement donner un intervalle de valeurs plausibles, et
- on ne peut pas être parfaitement certain que m appartient à cet intervalle de confiance, mais seulement déterminer la probabilité qu'il s'y trouve.

C'est sans doute avec ces applications statistiques à l'esprit que Galton parle du théorème central limite comme de la « loi de la fréquence des erreurs ».

Il y a une confusion fréquemment faite sur le dernier point : comme m n'est pas une variable aléatoire, comment peut-on parler de la probabilité qu'il se trouve dans un intervalle ? Il faut comprendre qu'ici, c'est l'*intervalle* qui est aléatoire (il dépend de $m_n(\omega)$, donc des observations).

Prenons pour illustrer ceci l'exemple d'un sondage sur une opinion binaire (« oui ou non »). On partage la population française, de N individus, en deux parties de tailles N_0 et N_1 selon l'opinion (non/oui), et on note $p = N_1/N$ la proportion des habitants ayant l'opinion « oui ». C'est ce paramètre p , inconnu, qui intéresse le sondeur. Pour l'estimer, il va contacter un nombre n d'individus très petit devant N (typiquement, $n = 1000$ dans la vraie vie), et recueillir leur opinion. Si les individus sont choisis indépendamment et uniformément, on voit que la suite X_1, X_2, \dots, X_n des opinions recueillies est i.i.d. de loi de Bernoulli de paramètre p (si l'on déclare que $X_i = 1$ si le i -ième individu contacté est d'opinion « oui », et $X_i = 0$ sinon).

Pour estimer p à partir des informations recueillies, le sondeur forme la moyenne empirique

$$p_n(\omega) = \frac{X_1(\omega) + \dots + X_n(\omega)}{n},$$

et en vertu des résultats précédents, il forme l'intervalle de confiance (8.4). Cependant, on doit prendre garde au fait que dans cette situation, $\sigma = \sqrt{p(1-p)}$ est un paramètre inconnu également. On peut néanmoins le majorer par $1/2$, si bien que

$$\left[p_n(\omega) - \frac{0,98}{\sqrt{n}}, p_n(\omega) + \frac{0,98}{\sqrt{n}} \right]$$

est un intervalle de confiance asymptotique à 95% pour p .

Pour que ce résultat soit parfaitement valide, il faudrait également estimer l'erreur commise en assimilant le membre de gauche de (8.3) et sa limite lorsque $n \rightarrow \infty$. Ce genre d'estimation est possible, par exemple à travers le théorème de Berry-Esseen, même s'il requiert des hypothèses de moment supplémentaires. Néanmoins, ce résultat peut déjà nous renseigner sur la qualité d'un sondage auprès de 1000 personnes : on a que $0,98 / \sqrt{1000} = 0,03099\dots$, ce qui signifie que l'intervalle de confiance est de rayon 3%. Autrement dit, avec une probabilité de l'ordre de 95% au mieux, le paramètre p égale le résultat du sondage $p_n(\omega)$ à *plus ou moins 3% près*.

8.3 Vecteurs aléatoires gaussiens et théorème central limite multidimensionnel

Il existe une extension du théorème central limite au cas de variables aléatoires à valeurs dans \mathbb{R}^d . Avant de l'énoncer, nous devons comprendre un peu mieux la nature des variables aléatoires gaussiennes à valeurs dans \mathbb{R}^d .

8.3.1 Vecteurs aléatoires gaussiens

Rappelons que pour $m \in \mathbb{R}$ et $\sigma > 0$, la loi $\mathcal{N}(m, \sigma^2)$ est la mesure de probabilités sur \mathbb{R} de densité

$$\frac{1}{(2\pi\sigma^2)^{1/2}} \exp\left(-\frac{(x-m)^2}{2\sigma^2}\right), \quad x \in \mathbb{R}.$$

On étend cette définition au cas où $\sigma = 0$ en disant que la loi $\mathcal{N}(m, 0)$ est δ_m , ce qui est cohérent avec le fait que g_σ soit une approximation de δ_0 pour la convolution, lorsque $\sigma \rightarrow 0$.

En particulier, si X est une v.a. de loi $\mathcal{N}(0, \sigma^2)$, alors $X + m$ a pour loi $\mathcal{N}(m, \sigma^2)$.

Définition 8.9. *Un vecteur aléatoire $X = (X_1, \dots, X_d)$ à valeurs dans \mathbb{R}^d est appelé un vecteur aléatoire gaussien si toute combinaison linéaire $\langle \xi, X \rangle = \xi_1 X_1 + \dots + \xi_d X_d$, avec $\xi \in \mathbb{R}^d$ est une variable aléatoire gaussienne réelle, c'est-à-dire qu'il existe $m_\xi \in \mathbb{R}, \sigma_\xi \geq 0$ tels que $\langle \xi, X \rangle$ a pour loi $\mathcal{N}(m_\xi, \sigma_\xi^2)$.*

Remarque. Attention, cette propriété est plus forte que de stipuler que les variables aléatoires X_1, \dots, X_d sont des variables gaussiennes. Pour s'en convaincre, soit X une v.a. de loi $\mathcal{N}(0, 1)$, et soit ε une v.a. indépendante de X de loi uniforme sur $\{-1, 1\}$. Alors X et εX sont toutes deux de loi $\mathcal{N}(0, 1)$. En revanche, la loi de $X + \varepsilon X$ n'est pas gaussienne, car $\mathbb{P}(X + \varepsilon X = 0) = \mathbb{P}(\varepsilon = -1) = 1/2$. Donc $(X, \varepsilon X)$ n'est pas un vecteur aléatoire gaussien.

Soit X un vecteur aléatoire gaussien à valeurs dans \mathbb{R}^d . Notons $m = (m_1, \dots, m_d)$ son espérance, et $\Sigma = (\text{Cov}(X_i, X_j), 1 \leq i, j \leq d)$ sa matrice de variance-covariance. Rappelons que Σ est une matrice symétrique positive. Remarquons alors que pour tout $\xi \in \mathbb{R}^d$, l'on a que

$$\mathbb{E}[\langle \xi, X \rangle] = \langle \xi, m \rangle$$

et que

$$\text{Var}(\langle \xi, X \rangle) = \langle \Sigma \xi, \xi \rangle.$$

De ce fait, on voit que la loi de $\langle \xi, X \rangle$, qui est supposée être gaussienne par définition, est $\mathcal{N}(\langle \xi, m \rangle, \langle \Sigma \xi, \xi \rangle)$, et est donc entièrement déterminée par m et Σ . De plus, on a que la fonction caractéristique de X est donnée par

$$\varphi_X(\xi) = \mathbb{E}[e^{i\langle \xi, X \rangle}] = \exp\left(i\langle \xi, m \rangle - \frac{\langle \Sigma \xi, \xi \rangle}{2}\right),$$

puisque c'est la transformée de Fourier de la loi $\mathcal{N}(\langle \xi, m \rangle, \langle \Sigma \xi, \xi \rangle)$ évaluée en 1. À son tour, la fonction φ_X ne dépend que de m, Σ , et par conséquent, pour un couple (m, Σ) donné, la loi d'un vecteur gaussien d'espérance m et de matrice de variance-covariance Σ , si elle existe, est unique.

Remarquons que réciproquement, si la fonction caractéristique de X est donnée par (8.9), alors on a bien que X est un vecteur aléatoire gaussien d'espérance m et de matrice de variance-covariance $\Sigma_X = \Sigma$. puisque pour tout $\xi \in \mathbb{R}^d$, on a alors, en changeant ξ en $t\xi$ dans (8.9) avec $t \in \mathbb{R}$, on a

$$\mathbb{E}[e^{it\langle \xi, X \rangle}] = \exp\left(it\langle \xi, m \rangle - \frac{\langle \Sigma \xi, \xi \rangle t^2}{2}\right),$$

c'est-à-dire que $\langle \xi, m \rangle$ a pour loi $\mathcal{N}(\langle \xi, m \rangle, \langle \Sigma \xi, \xi \rangle)$, d'où l'on tire que pour tout $\xi \in \mathbb{R}^d$,

$$\mathbb{E}[\langle \xi, X \rangle] = \langle \xi, m \rangle, \text{Var}(\langle \xi, X \rangle) = \langle \Sigma \xi, \xi \rangle,$$

et donc $m = \mathbb{E}[X]$, et $\Sigma = \Sigma_X$.

Nous notons $\mathcal{N}(m, \Sigma)$ la loi dont la transformée de Fourier est (8.9), sous réserve qu'elle existe. Nous allons montrer que c'est bien le cas pour tout $m \in \mathbb{R}^d$ et tout Σ symétrique positive. Pour ce faire, notons que les lois gaussiennes sont préservées par les applications linéaires, comme l'indique le résultat suivant.

Lemme 8.10. *Soit X un vecteur aléatoire gaussien de loi $\mathcal{N}(m, \Sigma)$ sur \mathbb{R}^d . Soit $A \in \mathfrak{M}_{k,d}(\mathbb{R})$. Alors AX est un vecteur aléatoire gaussien à valeurs dans \mathbb{R}^k , de loi $\mathcal{N}(Am; A\Sigma A^T)$.*

À ce stade, il faut encore comprendre ce lemme sous la forme « si la loi $\mathcal{N}(m, \Sigma)$ existe, alors la loi $\mathcal{N}(Am, A\Sigma A^T)$ existe également. Nous avons énoncé ce résultat en anticipant le résultat de la proposition 8.11.

Démonstration. Soit $\xi \in \mathbb{R}^k$. Alors on a que $\langle \xi, AX \rangle = \langle A^T \xi, X \rangle$ et donc

$$\begin{aligned} \mathbb{E}[e^{i\langle \xi, AX \rangle}] &= \mathbb{E}[e^{i\langle A^T \xi, X \rangle}] \\ &= \exp\left(i\langle A^T \xi, m \rangle - \frac{\langle \Sigma A^T \xi, A^T \xi \rangle}{2}\right) \\ &= \exp\left(i\langle \xi, Am \rangle - \frac{\langle A\Sigma A^T \xi, \xi \rangle}{2}\right), \end{aligned}$$

d'où le résultat. \square

Proposition 8.11. *Soit $m \in \mathbb{R}^d$, et $\Sigma \in \mathfrak{M}_d(\mathbb{R})$ une matrice symétrique positive. Alors il existe un vecteur gaussien d'espérance m et de matrice de variance-covariance Σ .*

Démonstration. Traitons d'abord le cas où $m = 0$ et où $\Sigma = I_d$. Il suffit pour cela de prendre un vecteur aléatoire $X = (X_1, \dots, X_d)$ dont les composantes sont des variables aléatoires i.i.d. de loi $\mathcal{N}(0, 1)$: en effet on a pour tout $\xi \in \mathbb{R}^d$

$$\mathbb{E}[e^{i\langle \xi, X \rangle}] = \prod_{j=1}^d \mathbb{E}[e^{i\xi_j X_j}] = \prod_{j=1}^d e^{-\frac{\xi_j^2}{2}} = e^{-\frac{|\xi|^2}{2}}$$

(on aurait également pu invoquer des résultats déjà connus de la transformée de Fourier de densités gaussiennes sur \mathbb{R}^d).

Pour traiter le cas général, on se donne (Y_1, \dots, Y_d) de loi $\mathcal{N}(0, I_d)$. Comme Σ est symétrique positive, elle admet une unique racine carrée symétrique positive $\sqrt{\Sigma}$ (on diagonalise $\Sigma = PDP^{-1}$ avec D à diagonale positive et $P \in O(d)$, et on pose $\sqrt{\Sigma} = P\sqrt{D}P^{-1}$). Posons $X = m + \sqrt{\Sigma}Y$. Par le lemme 8.10, cette variable aléatoire a pour loi $\mathcal{N}(m, \sqrt{\Sigma}I_d\sqrt{\Sigma}^T) = \mathcal{N}(m, \Sigma)$. \square

Il est très utile en pratique de se souvenir qu'une variable aléatoire ayant cette loi peut se définir par une simple transformation linéaire à partir d'une suite de variables i.i.d. (Y_1, \dots, Y_d) de loi $\mathcal{N}(0, 1)$, comme on l'a fait dans la preuve précédente.

Comme autre application simple du lemme 8.10, notons une propriété fondamentale des lois gaussiennes dans le cas où Σ est une matrice scalaire.

Proposition 8.12. *Pour tout $\sigma \geq 0$, la loi $\mathcal{N}(0, \sigma^2 I_d)$ est invariante par l'action du groupe orthogonal de \mathbb{R}^d : si X a pour loi $\mathcal{N}(0, \sigma^2 I_d)$, alors pour toute matrice orthogonale P , PX a même loi que X .*

Exercice 8.1. Montrer que cette propriété d'invariance caractérise les lois gaussiennes de matrice de variance-covariance scalaire et d'espérance nulle parmi toutes les lois gaussiennes.

Une autre propriété cruciale est que, pour les vecteurs aléatoires gaussiens, indépendance et décorrélation sont équivalentes.

Proposition 8.13. *Soit $(X_1, \dots, X_d, X'_1, \dots, X'_{d'})$ un vecteur aléatoire gaussien. On suppose que pour tout $i, j \in \{1, 2, \dots, d\} \times \{1, 2, \dots, d'\}$, on a $\text{Cov}(X_i, X'_j) = 0$. Alors les vecteurs aléatoires gaussiens $X = (X_1, \dots, X_d)$ et $X' = (X'_1, \dots, X'_{d'})$ sont indépendants.*

Démonstration. Nos hypothèses impliquent que la matrice de variance-covariance $\Sigma_{(X, X')}$ de $(X_1, \dots, X_d, X'_1, \dots, X'_{d'})$ est diagonale par blocs de tailles d et d' . Si ξ, ξ' sont respectivement dans \mathbb{R}^d et $\mathbb{R}^{d'}$, et (ξ, ξ') est le vecteur de $\mathbb{R}^{d+d'}$ obtenu en les concaténant, alors en notant $m = \mathbb{E}[X]$, $m' = \mathbb{E}[X']$, $\Sigma = \Sigma_X$ et $\Sigma' = \Sigma_{X'}$, la fonction caractéristique de (X, X') est donc donnée par

$$\varphi_{(X, X')}((\xi, \xi')) = \exp\left(i\langle \xi, m \rangle + i\langle \xi', m' \rangle - \frac{\langle \Sigma \xi, \xi \rangle + \langle \Sigma' \xi', \xi' \rangle}{2}\right)$$

ce qui se factorise en $\varphi_X(\xi) \varphi_{X'}(\xi')$. On applique alors le critère d'indépendance à l'aide des fonctions caractéristiques. \square

Terminons ces généralités sur les vecteurs aléatoires gaussiens en mentionnant que, contrairement au cas de la dimension 1, un vecteur aléatoire de loi $\mathcal{N}(m, \Sigma)$ n'admet pas nécessairement de densité par rapport à la mesure de Lebesgue.

Proposition 8.14. *Soit $m \in \mathbb{R}^d$ et Σ une matrice symétrique positive. Si Σ est définie positive, alors la loi $\mathcal{N}(m, \Sigma)$ admet une densité par rapport à la mesure de Lebesgue sur \mathbb{R}^d , donnée par*

$$\frac{1}{\sqrt{\det(2\pi\Sigma)}} \exp\left(-\frac{\langle \Sigma^{-1}(x-m), (x-m) \rangle}{2}\right), \quad x \in \mathbb{R}^d.$$

Si le rang de Σ est strictement inférieur à d , alors la loi $\mathcal{N}(m, \Sigma)$ est portée par le sous-espace affine $m + \Sigma\mathbb{R}^d$, de dimension $\text{rg}(\Sigma)$.

Démonstration. Supposons Σ inversible. Rappelons qu'un vecteur aléatoire $X = m + \sqrt{\Sigma}Y$ a la loi voulue, si $Y = (Y_1, \dots, Y_d)$ est de loi $\mathcal{N}(0, I_d)$, dont la densité est la densité gaussienne $g_1(x) = (2\pi)^{-d/2} \exp(-|x|^2/2)$ sur \mathbb{R}^d . Par la formule du changement de variables (dans le cas d'un isomorphisme linéaire) on a donc, pour toute fonction f mesurable positive,

$$\begin{aligned} \mathbb{E}[f(X)] &= \frac{1}{(2\pi)^{d/2}} \int_{\mathbb{R}^d} f(m + \sqrt{\Sigma}y) \exp\left(-\frac{|y|^2}{2}\right) dy \\ &= \frac{1}{(2\pi)^{d/2}} \int_{\mathbb{R}^d} f(x) \exp\left(-\frac{|\sqrt{\Sigma}^{-1}(x-m)|^2}{2}\right) \frac{dx}{\det(\sqrt{\Sigma})}, \end{aligned}$$

et on conclut par un réagencement des termes.

La seconde partie de la proposition consiste à appliquer le même raisonnement, en posant $X = m + \sqrt{\Sigma}Y$, et en constatant que $\sqrt{\Sigma}\mathbb{R}^d = \Sigma\mathbb{R}^d$, ce que l'on constate aisément en diagonalisant Σ . \square

8.3.2 Théorème central limite : le cas de \mathbb{R}^d

Théorème 8.15. *Soit X_1, X_2, \dots une suite i.i.d. de variables aléatoires à valeurs dans \mathbb{R}^d , dont les coordonnées sont toutes dans L^2 . On note $m = \mathbb{E}[X_1]$ et $\Sigma = \Sigma_X$. Alors, si $S_n = X_1 + \dots + X_n$,*

$$\frac{S_n - nm}{\sqrt{n}} \xrightarrow[n \rightarrow \infty]{\text{loi}} \mathcal{N}(0, \Sigma).$$

Démonstration. Soit $\xi \in \mathbb{R}^d$. Alors les variables aléatoires $\langle \xi, X_j \rangle, j \geq 1$ sont i.i.d. De plus, elles sont clairement dans L^2 , de moyenne $\langle \xi, m \rangle$ et de matrice de variance-covariance $\langle \Sigma \xi, \xi \rangle$. Le théorème central limite appliqué à ces variables réelles donne que

$$\frac{\langle \xi, S_n - nm \rangle}{\sqrt{n}} \xrightarrow[n \rightarrow \infty]{\text{loi}} \mathcal{N}(0, \langle \Sigma \xi, \xi \rangle).$$

Par le théorème de Lévy (dans le sens facile) ceci implique que pour tout $\xi \in \mathbb{R}^d$,

$$\mathbb{E} \left[\exp \left(i \frac{\langle \xi, S_n - nm \rangle}{\sqrt{n}} \right) \right] \xrightarrow{n \rightarrow \infty} \exp \left(- \frac{\langle \Sigma \xi, \xi \rangle}{2} \right).$$

Mais on reconnaît là la convergence de la fonction caractéristique de $(S_n - nm) / \sqrt{n}$ vers la transformée de Fourier de $\mathcal{N}(0, \Sigma)$, et en utilisant le théorème de Lévy à nouveau, on conclut. \square

Remarque 8.16. Nous voyons ici une autre propriété fascinante du théorème central limite : si la matrice de variance-covariance de X_1 est scalaire, alors la loi limite de $(S_n - nm) / \sqrt{n}$ est invariante par l'action du groupe orthogonal de \mathbb{R}^d . C'est une propriété de symétrie extrêmement forte !

Exercice 8.2. Soit X_1, X_2, \dots des variables aléatoires i.i.d. uniformes dans $\{-1, 1\}$, et $S_n = X_1 + \dots + X_n$. Montrer que la probabilité $\mathbb{P}(S_n > 0, S_{2n} < 0)$ a une limite lorsque $n \rightarrow \infty$, et la calculer.

8.3.3 Une application : le test d'adéquation du χ^2

Soit $k \geq 1$ un entier fixé. On se donne une suite Y_1, Y_2, \dots i.i.d. de variables aléatoires à valeurs dans $\{1, 2, \dots, k\}$.

On se place du point de vue d'un statisticien qui ignore la loi de Y_1 . Cette dernière est un vecteur de probabilités $\mathbf{p} = (p_1, p_2, \dots, p_k)$, qui est donc inconnu de l'observateur. Ce dernier voudrait *tester* l'hypothèse que $\mathbf{p} = \mathbf{p}^0$, où \mathbf{p}^0 est un vecteur de probabilités fixé. Cette hypothèse est appelée *l'hypothèse nulle*, et est souvent notée H_0 . Le point de vue du test statistique est en quelque sorte de prendre le point de vue d'un « procès », qui met l'hypothèse H_0 à l'épreuve. On va donc supposer H_0 , et essayer de déduire de nos observations Y_1, Y_2, \dots si cette hypothèse est vraisemblable. Dans le cas contraire, on rejettera H_0 . Une situation très concrète consiste à disposer d'un dé, et de tester par une série de lancers indépendants s'il est correctement équilibré.

Clairement, si on dispose de toutes les observations Y_1, Y_2, \dots , la loi des grands nombres nous permet de retrouver exactement p_i comme la limite de $N_n^{(i)} / n$, où $N_n^{(i)} = \#\{j \leq n : Y_j = i\}$, mais cette situation n'est pas réaliste.

Un test asymptotique couramment utilisé est le test d'adéquation du χ^2 , et qui consiste à introduire la « statistique du χ^2 » suivante :

$$\chi_n = \sum_{i=1}^k \frac{(N_n^{(i)} - np_i^0)^2}{np_i^0}.$$

Proposition 8.17. *Sous l'hypothèse H_0 que la suite Y_1, Y_2, \dots est i.i.d. de loi commune \mathbf{p}^0 , la variable aléatoire χ_n converge en loi lorsque $n \rightarrow \infty$ vers une loi gamma de paramètres $(1/2, (k-1)/2)$. Cette dernière loi est encore appelée loi du χ^2 à $k-1$ degrés de liberté, et est la loi de $|X|^2$ où X est un vecteur gaussien de loi $\mathcal{N}(0, I_{k-1})$.*

Démonstration. Notons $X_j = ((\mathbf{1}_{\{Y_j=i\}} - p_i) / \sqrt{p_i}, 1 \leq i \leq k)$, ce qui définit pour chaque j une variable aléatoire à valeurs dans \mathbb{R}^k . La suite $(X_j, j \geq 1)$ est clairement i.i.d., bornée et centrée, et l'on a que

$$\chi_n = \left| \frac{\sum_{j=1}^n X_j}{\sqrt{n}} \right|^2.$$

Le théorème central limite multidimensionnel, en plus du fait que la fonction $x \mapsto |x|^2$ est continue de \mathbb{R}^k dans \mathbb{R} , montre que cette variable aléatoire converge vers $|G|$, où G suit la loi $\mathcal{N}(0, \Sigma)$, où $\Sigma = \Sigma_{X_1}$ est la matrice de variance-covariance de X_1 . Son calcul est facile :

$$\Sigma = (\delta_{i,j} - \sqrt{p_i p_j})_{1 \leq i, j \leq k} = I_k - \sqrt{\mathbf{p}} \sqrt{\mathbf{p}}^*,$$

où l'on a noté $\sqrt{\mathbf{p}} = (\sqrt{p_i}, 1 \leq i \leq k)$. Comme ce vecteur est de norme euclidienne 1, on reconnaît que Σ est la matrice de projection orthogonale sur l'hyperplan $\sqrt{\mathbf{p}}^\perp$ dans \mathbb{R}^k . Cette matrice est sa propre racine carrée, $\Sigma^2 = \Sigma$, puisque c'est un projecteur, et par conséquent la variable aléatoire G a même loi que $\Sigma G'$, où G' est de loi $\mathcal{N}(0, I_k)$. Par ailleurs, si A est une matrice orthogonale envoyant $\sqrt{\mathbf{p}}$ sur le dernier vecteur de la base canonique de \mathbb{R}^k , alors $A \Sigma G'$ et AG sont de même loi $\mathcal{N}(0, P)$ où P est la matrice de la projection $(x_1, \dots, x_k) \mapsto (x_1, \dots, x_{k-1}, 0)$, et donc AG a même loi que $(G_1, G_2, \dots, G_{k-1}, 0)$, où les G_i sont indépendantes de loi $\mathcal{N}(0, 1)$. Comme par ailleurs la norme de AG égale celle de G , on déduit bien que $|G|^2$ a la même loi que $G_1^2 + G_2^2 + \dots + G_{k-1}^2$. On laisse en exercice le soin de vérifier que la loi de cette variable aléatoire est celle annoncée. \square

Soit alors $\alpha \in (0, 1)$. On se donne $x_\alpha > 0$ tel que $\mathbb{P}(\chi \geq x_\alpha) < \alpha$, où χ suit une loi du χ^2 à $k-1$ degrés de liberté. Le résultat précédent montre que sous l'hypothèse H_0 , si n est assez grand, on a que $\mathbb{P}(\chi_n \geq x_\alpha) < \alpha$. Le test du χ^2 consiste donc à

rejeter l'hypothèse H_0 si $\chi_n \geq x_\alpha$,

et à la conserver sinon. Le *niveau* (asymptotique) du test, c'est-à-dire la probabilité d'un faux positif (rejeter H_0 alors qu'elle est vraie) est d'au plus α . En pratique, la valeur $\alpha = 5\%$ est souvent utilisée. Pour un k fixé, il est aisé de trouver de telles valeurs de x_α à l'aide d'outils numériques.

À noter que si H_0 n'est pas vérifiée, c'est-à-dire que $\mathbf{p} \neq \mathbf{p}^0$, alors la statistique du χ^2 diverge par la loi des grands nombres : $\chi_n \rightarrow \infty$ presque sûrement. De ce fait, l'hypothèse H_0 sera bien rejetée avec une probabilité tendant vers 1 si n est assez grand. Cette probabilité est appelée la *puissance* du test.

8.4 L'inégalité de Hoeffding

Nous avons vu comment le théorème central limite permet d'obtenir de intervalles de confiance asymptotiques pour des problèmes d'estimation statistique. Un problème important est que ces intervalles de confiance ne sont en effet qu'asymptotiques, et en toute rigueur on ne peut pas les appliquer en pratique à n fixé sans connaissance supplémentaire sur l'erreur commise. Il est donc important de savoir donner des intervalles de confiance non-asymptotiques.

Notons que le problème revient *in fine* à donner une borne la meilleure possible pour des quantités de la forme

$$\mathbb{P}(|S_n - \mathbb{E}[S_n]| \geq x)$$

où $S_n = X_1 + \dots + X_n$ est une somme de v.a.i.i.d et $x \geq 0$ est typiquement de la forme $c\sqrt{n}$. Bien sûr, la première borne exacte qui vient à l'esprit est celle de Bienaymé-Chebychev, qui donne

$$\mathbb{P}(|S_n - \mathbb{E}[S_n]| \geq x) \leq n \frac{\text{Var}(X_1)}{x^2}.$$

Pour $x = c\sqrt{n}$, on obtient une borne $\text{Var}(X_1)/c^2$. On voit qu'il s'agit d'une borne relativement mauvaise si on la compare avec la borne asymptotique donnée par le théorème central limite :

$$\lim_{n \rightarrow \infty} \mathbb{P}(|S_n - \mathbb{E}[S_n]| \geq c\sqrt{n}) = 2 \int_c^\infty g_\sigma(x) dx$$

où $\sigma^2 = \text{Var}(X_1)$ et g_σ est la densité de la loi gaussienne $\mathcal{N}(0, \sigma^2)$. En effet, un calcul élémentaire donne (exercice)

$$2 \int_c^\infty g_\sigma(x) dx \leq 2 \frac{\sigma e^{-c^2/2\sigma^2}}{c\sqrt{2\pi}},$$

qui est une borne bien meilleure que la borne de Bienaymé-Chebychev pour les grandes valeurs de c .

Le théorème suivant montre que, si l'on s'intéresse à des variables aléatoires bornées, il existe une borne non-asymptotique qui permet de faire presque aussi bien que la borne asymptotique.

Théorème 8.18. *Soit X_1, \dots, X_n des variables aléatoires indépendantes, respectivement à valeurs dans des intervalles compacts $[a_1, b_1], \dots, [a_n, b_n]$. Notons $S_n = X_1 + \dots + X_n$. Alors on a, pour tout $x \geq 0$,*

$$\mathbb{P}(S_n - \mathbb{E}[S_n] \geq x) \leq \exp\left(-\frac{2x^2}{\sum_{i=1}^n (b_i - a_i)^2}\right).$$

et de plus, toujours pour $x \geq 0$,

$$\mathbb{P}(|S_n - \mathbb{E}[S_n]| \geq x) \leq 2 \exp\left(-\frac{2x^2}{\sum_{i=1}^n (b_i - a_i)^2}\right).$$

Démonstration. Soit Y une variable aléatoire bornée, disons que $Y \in [a, b]$ avec probabilité 1, et d'espérance nulle. Considérons la transformée de log-Laplace de Y , définie par

$$\Phi(\lambda) = \ln(\mathbb{E}[e^{\lambda Y}]), \quad \lambda \in \mathbb{R}.$$

La fonction Φ est alors de classe $\mathcal{C}^\infty(\mathbb{R})$, et l'on a

$$\Phi'(\lambda) = \frac{\mathbb{E}[Ye^{\lambda Y}]}{\mathbb{E}[e^{\lambda Y}]}, \quad \Phi''(\lambda) = \frac{\mathbb{E}[Y^2 e^{\lambda Y}]\mathbb{E}[e^{\lambda Y}] - \mathbb{E}[Ye^{\lambda Y}]^2}{\mathbb{E}[e^{\lambda Y}]^2}.$$

On peut réécrire $\Phi''(\lambda)$ sous la forme

$$\Phi''(\lambda) = \mathbb{E} \left[Y^2 \frac{e^{\lambda Y}}{\mathbb{E}[e^{\lambda Y}]} \right] - \mathbb{E} \left[Y \frac{e^{\lambda Y}}{\mathbb{E}[e^{\lambda Y}]} \right]^2 = \text{Var}_\lambda(Y),$$

où nous avons noté Var_λ la variance pour la nouvelle mesure de probabilités \mathbb{P}_λ définie par

$$\mathbb{P}_\lambda(A) = \mathbb{E} \left[\mathbf{1}_A \frac{e^{\lambda Y}}{\mathbb{E}[e^{\lambda Y}]} \right],$$

c'est-à-dire que \mathbb{P}_λ est absolument continue par rapport à \mathbb{P} , et de dérivée de Radon-Nikodym donnée par $e^{\lambda Y} / \mathbb{E}[e^{\lambda Y}]$. Nous en déduisons d'une part que $\Phi''(\lambda) \geq 0$ pour tout λ , c'est-à-dire que Φ est convexe, mais également que

$$\Phi''(\lambda) = \text{Var}_\lambda(Y) = \inf \{ \mathbb{E}_\lambda[(Y - c)^2] : c \in \mathbb{R} \} \leq \frac{(b - a)^2}{4},$$

où l'on a noté \mathbb{E}_λ l'espérance associée à \mathbb{P}_λ , et où l'on a remarqué que la variance de Y est la distance de Y au sous-espace des fonctions constantes dans $L^2(\Omega, \mathcal{F}, \mathbb{P}_\lambda)$, puisque $\mathbb{E}_\lambda[Y]$ est le projeté orthogonal de Y sur ce sous-espace, et où l'on a pris $c = (b - a)/2$ pour obtenir la dernière inégalité.

De cela, on déduit, en intégrant deux fois entre 0 et λ , et en constatant que $\Phi(0) = \Phi'(0) = 0$,

$$\Phi(\lambda) \leq \frac{(b - a)^2}{8} \lambda^2, \quad \lambda \in \mathbb{R}. \quad (8.5)$$

Montrons maintenant le théorème. On fixe $x, \lambda \geq 0$. En appliquant l'inégalité de Markov, puis en utilisant l'indépendance des variables aléatoires X_1, \dots, X_n , on a

$$\begin{aligned} \mathbb{P}(S_n - \mathbb{E}[S_n] \geq x) &= \mathbb{P}(\exp(\lambda(S_n - \mathbb{E}[S_n])) \geq \exp(\lambda x)) \\ &\leq e^{-\lambda x} \mathbb{E} \left[\exp \left(\lambda \sum_{i=1}^n (X_i - \mathbb{E}[X_i]) \right) \right] \\ &\leq e^{-\lambda x} \prod_{i=1}^n \mathbb{E}[\exp(\lambda(X_i - \mathbb{E}[X_i]))] \\ &\leq e^{-\lambda x} \exp \left(\sum_{i=1}^n \frac{(b_i - a_i)^2}{8} \lambda^2 \right). \end{aligned}$$

À la dernière étape, on a utilisé l'inégalité (8.5) pour les variables aléatoires $X_i - \mathbb{E}[X_i]$, qui sont à valeurs dans l'intervalle $[a_i - \mathbb{E}[X_i], b_i - \mathbb{E}[X_i]]$, de diamètre $b_i - a_i$. Finalement, on obtient la première inégalité de l'énoncé en choisissant $\lambda = x / (2 \sum_{i=1}^n (b_i - a_i)^2)$, ce qui revient à minimiser le majorant dans l'inégalité précédente.

La seconde inégalité s'obtient en appliquant la première inégalité aux variables aléatoires $-X_1, \dots, -X_n$, ce qui donne la même borne que la première pour $\mathbb{P}(-S_n + \mathbb{E}[S_n] \geq x)$, et en utilisant le fait que $\{|S_n - \mathbb{E}[S_n]| \geq x\}$ est la réunion des événements $\{S_n - \mathbb{E}[S_n] \geq x\}$ et $\{-S_n + \mathbb{E}[S_n] \geq x\}$. \square

En particulier, si les variables aléatoires (X_1, \dots, X_n) sont i.i.d. à valeurs dans $[a, b]$, on obtient que

$$\mathbb{P}(|S_n - \mathbb{E}[S_n]| \geq c\sqrt{n}) \leq 2 \exp\left(-\frac{2c^2}{(b-a)^2}\right).$$

En reprenant l'exemple des sondages, où X_1 est une variable aléatoire de Bernoulli de paramètre p inconnu, on a $a = 0$ et $b = 1$, et en choisissant $c = \sqrt{\ln(2/\delta)}/2$ de sorte que le majorant précédent soit inférieur à δ , on voit que

$$\left[\frac{S_n}{n} - \sqrt{\frac{\ln(2/\delta)}{2n}}, \frac{S_n}{n} + \sqrt{\frac{\ln(2/\delta)}{2n}} \right]$$

est un intervalle de confiance (exact) pour p au niveau δ , c'est-à-dire que la probabilité que cet intervalle ne contienne pas p est inférieure à δ . Pour $\delta = 5\%$, notons que $\sqrt{\ln(2/\delta)}/2 = 1,358\dots$, à comparer avec la valeur numérique 0,98 que l'on avait obtenue par le théorème central limite. Il est normal d'obtenir un intervalle plus grand, c'est-à-dire moins bon que la valeur asymptotique « idéale ».

Chapitre 9

Réurrence et transience pour la marche aléatoire simple sur \mathbb{Z}^d

Un problème important en probabilités est de déterminer le comportement en temps long de processus aléatoires. L'exemple le plus fondamental est celui des marches aléatoires dans \mathbb{Z}^d . Soit μ une loi de probabilités sur \mathbb{Z}^d , et X_1, X_2, \dots une suite i.i.d. de loi commune μ . Notons $S_n = X_1 + \dots + X_n$, pour $n \geq 1$, et $S_0 = 0$.

Considérons l'événement $R = \limsup_{n \rightarrow \infty} \{S_n = 0\}$ sur lequel $(S_n, n \geq 0)$ prend une infinité de fois la valeur 0. Si $\mathbb{P}(R) = 1$, on dit que la marche aléatoire est *récurrente*. Si $\mathbb{P}(R) = 0$, on dit qu'elle est *transiente*.

Proposition 9.1. *On a que $\mathbb{P}(R) \in \{0, 1\}$, c'est-à-dire qu'une marche aléatoire est ou bien récurrente, ou bien transiente. Cette dernière propriété ne dépend que de μ .*

Il est à noter que cette proposition n'est pas une conséquence directe de la loi du 0-1 de Kolmogorov. On peut l'obtenir par un autre type de loi du 0-1, dite loi de Hewitt et Savage, mais nous allons donner une preuve directe de la proposition qui ne fait pas appel à cette loi.

Pour tout $x \in \mathbb{Z}^d$, notons

$$g(x) = \sum_{n \geq 0} \mathbb{P}(S_n = x) = \sum_{n \geq 0} \mu^{*n}(x) = \mathbb{E} \left[\sum_{n \geq 0} \mathbf{1}_{\{S_n = x\}} \right].$$

Ainsi, $g(x)$ est le nombre moyen de fois où S_n vaut x . Notons également $H_0 = \inf \{n > 0 : S_n = 0\}$ le premier temps de retour en 0. Nous allons voir que

$$g(0) = \frac{1}{1 - \mathbb{P}(H_0 < \infty)}. \quad (9.1)$$

Pour cela, notons $H_0^{(1)} = H_0$, et par récurrence, pour $k \geq 2$,

$$H_0^{(k)} = \inf \{n > H_0^{(k-1)} : S_n = 0\}$$

de sorte que $H_0^{(k)}$ est l'instant du k -ième retour en 0 pour $(S_n, n \geq 0)$. Notons que l'on note $H_0^{(k)} = \infty$ si l'ensemble sur lequel on prend la borne inférieure est vide, et en particulier si $H_0^{(k-1)} = \infty$.

Lemme 9.2. *Pour tout $k \geq 1$, on a que $\mathbb{P}(H_0^{(k)} < \infty) = \mathbb{P}(H_0 < \infty)^k$.*

Démonstration. La preuve est par récurrence sur k . Pour $k = 1$ c'est évident par définition. Si c'est vrai au rang k , on a

$$\begin{aligned} \mathbb{P}(H_0^{(k+1)} < \infty) &= \mathbb{P}(H_0^{(k)} < \infty, H_0^{(k+1)} < \infty) \\ &= \sum_{r>0} \mathbb{P}(H_0^{(k)} = r, H_0^{(k+1)} < \infty). \end{aligned}$$

Or sur l'événement $\{H_0^{(k)} = r\}$, on a que $S_n = X_{r+1} + X_{r+2} + \dots + X_n = S_{n-r}^{(r)}$ pour tout $n \geq r$, où l'on a noté $S_0^{(r)} = 0$ et

$$S_n^{(r)} = X_{r+1} + \dots + X_{r+n}, \quad n \geq 1,$$

de sorte que $(S_n^{(r)}; n \geq 0)$ est une marche aléatoire de même loi que $(S_n, n \geq 0)$, tout en étant indépendante de $\sigma(X_1, \dots, X_r)$ par le lemme de regroupement par paquets. Toujours en restriction à l'événement $\{H_0^{(k)} = r\}$, on a alors que $H_0^{(k+1)} < \infty$ si et seulement si $S_n^{(r)}$ s'annule pour au moins un $n > 0$, et donc

$$\{H_0^{(k)} = r, H_0^{(k+1)} < \infty\} = \{H_0^{(k)} = r\} \cap \{\exists n > 0: S_n^{(r)} = 0\}.$$

L'événement à droite de l'intersection est mesurable par rapport à $\sigma(S_n^{(r)}, n \geq 0)$ et est donc indépendant de $\sigma(X_1, \dots, X_r)$, tandis que celui de gauche est dans $\sigma(X_1, \dots, X_r)$ puisqu'on peut le réécrire comme

$$\{H_0^{(k)} = r\} = \left\{ \sum_{n=1}^{r-1} \mathbf{1}_{\{S_n=0\}} = k-1 \right\} \cap \{S_r = 0\}.$$

On conclut que les deux événements ci-dessus sont indépendants, et par conséquent

$$\begin{aligned} \mathbb{P}(H_0^{(k+1)} < \infty) &= \sum_{r>0} \mathbb{P}(H_0^{(k)} = r) \mathbb{P}(\exists n > 0: S_n^{(r)} = 0) \\ &= \mathbb{P}(\exists n > 0: S_n = 0) \sum_{r>0} \mathbb{P}(H_0^{(k)} = r) \\ &= \mathbb{P}(H_0 < \infty) \mathbb{P}(H_0^{(k)} < \infty), \end{aligned}$$

où l'on a utilisé le fait que $(S_n^{(r)}, n \geq 0)$ et $(S_n, n \geq 0)$ ont la même loi. On conclut par l'hypothèse de récurrence. \square

Montrons à présent l'égalité (9.1). Pour cela, on constate que

$$\sum_{n \geq 1} \mathbf{1}_{\{S_n=0\}} = \sum_{k \geq 1} \mathbf{1}_{\{H_0^{(k)} < \infty\}},$$

de sorte que

$$\begin{aligned} g(0) &= 1 + \mathbb{E} \left[\sum_{k \geq 1} \mathbf{1}_{\{H_0^{(k)} < \infty\}} \right] \\ &= \sum_{k \geq 0} \mathbb{P}(H_0 < \infty)^k = \frac{1}{1 - \mathbb{P}(H_0 < \infty)}, \end{aligned}$$

comme voulu.

Nous pouvons maintenant donner la preuve de la proposition 9.1.

- Si $\mathbb{P}(H_0 < \infty) < 1$, alors $g(0) < \infty$ par (9.1). Donc $\sum_{n \geq 0} \mathbf{1}_{\{S_n=0\}} < \infty$ avec probabilité 1. On a ainsi que $\mathbb{P}(R) = 0$, et la marche aléatoire est transiente.
- Si $\mathbb{P}(H_0 < \infty) = 1$, le lemme 9.2 montre que $\mathbb{P}(H_0^{(k)} < \infty) = 1$ pour tout $k \geq 1$. Comme $\{H_0^{(k)} < \infty\}$ est l'événement que $(S_n, n \geq 0)$ prend au moins $k + 1$ fois la valeur 0, ces événements décroissent vers l'événement R lorsque $k \rightarrow \infty$. On obtient donc que $\mathbb{P}(R) = 1$, et la marche est récurrente.

Notons au passage que nous avons obtenu le résultat suivant.

Proposition 9.3. *La marche aléatoire $(S_n, n \geq 0)$ est récurrente si et seulement si $\mathbb{P}(H_0 < \infty) = 1$, et ce si et seulement si $g(0) = \infty$.*

Nous allons maintenant donner un résultat important dans le cas particulier de la marche aléatoire simple sur \mathbb{Z}^d , que nous définissons maintenant. Notons (e_1, \dots, e_d) la base canonique de \mathbb{R}^d . La marche aléatoire simple sur \mathbb{Z}^d est la marche aléatoire correspondant au choix de μ soit donné par

$$\mu(\{x\}) = \frac{1}{2d}, \quad x \in \{e_1, \dots, e_d, -e_1, \dots, -e_d\},$$

et $\mu(x) = 0$ si $x \notin \{e_1, \dots, e_d, -e_1, \dots, -e_d\}$. Ainsi, à chaque pas, la marche aléatoire choisit uniformément au hasard l'un de ses $2d$ voisins et s'y déplace.

Théorème 9.4. *La marche aléatoire simple sur \mathbb{Z}^d est récurrente si $d \in \{1, 2\}$, et transiente si $d \geq 3$.*

Remarque. Dans le cas où $d = 1$, nous avons déjà montré ce résultat, qui découle immédiatement du corollaire 6.17, qui découlait de la loi du 0-1 de Kolmogorov. Néanmoins, nous allons redonner une preuve également dans ce cas à l'aide des outils développés dans ce chapitre, car cette preuve resservira pour le cas $d = 2$.

Démonstration. Commençons donc par le cas où $d = 1$. Dans ce cas, on a immédiatement que $\mathbb{P}(S_n = 0) = 0$ si n est impair, et $\mathbb{P}(S_{2n} = 0) = \mathbb{P}(X_1 + \dots + X_{2n} = 0)$ où X_1, X_2, \dots sont i.i.d. avec $\mathbb{P}(X_1 = 1) = \mathbb{P}(X_1 = -1) = 1/2$. L'événement ci-dessus correspond au fait que parmi les valeurs (X_1, \dots, X_{2n}) , exactement n sont égales à -1 , et les n autres valent 1. Comme il y a $\binom{2n}{n}$ choix des indices correspondants, on obtient que

$$\mathbb{P}(S_{2n} = 0) = \frac{1}{2^{2n}} \binom{2n}{n} \underset{n \rightarrow \infty}{\sim} \frac{1}{\sqrt{\pi n}},$$

grâce à la formule de Stirling. On en déduit immédiatement que

$$g(0) = \sum_{n \geq 0} \mathbb{P}(S_n = 0) = \infty,$$

et donc que la marche aléatoire est récurrente.

Dans le cas où $d=2$, on a toujours $\mathbb{P}(S_n=0)=0$ si n est impair. Pour calculer $\mathbb{P}(S_{2n}=0)$, on a recours à une astuce. Si X_1 est une variable aléatoire de loi μ , et si R est la rotation de centre 0 et d'angle $\pi/4$ de \mathbb{R}^2 , alors $R(X_1)$ est une variable aléatoire égale à chacune des quatre valeurs $\sqrt{2}/2(\pm 1, \pm 1)$ avec même probabilité. Ainsi, $R(X_1)$ a même loi que $\sqrt{2}/2(X'_1, X''_1)$ où X'_1, X''_1 sont des variables aléatoires indépendantes, réelles, de même loi, telles que $\mathbb{P}(X'_1=1)=\mathbb{P}(X'_1=-1)=1/2$. Finalement, on voit que $(R(S_n), n \geq 0)$ a même loi que $(\sqrt{2}/2(S'_n, S''_n), n \geq 0)$, où $(S'_n, n \geq 0)$ et $(S''_n, n \geq 0)$ sont deux marches aléatoires simples indépendantes à valeurs dans \mathbb{Z} . De ce fait, on a que

$$\mathbb{P}(S_{2n}=0) = \mathbb{P}(S'_{2n}=0)\mathbb{P}(S''_{2n}=0) \underset{n \rightarrow \infty}{\sim} \frac{1}{\pi n},$$

en utilisant le cas précédent. On voit une nouvelle fois que $g(0) = \infty$, et donc la marche est récurrente.

L'astuce précédente de fonctionne plus en dimension supérieure à 3. On a donc recours au lemme suivant, valable pour toutes les marches aléatoires sur \mathbb{Z}^d .

Lemme 9.5. *La marche aléatoire $(S_n, n \geq 0)$ est récurrente si et seulement si*

$$\lim_{t \uparrow 1} \int_{[-\pi, \pi]^d} \Re \left(\frac{1}{1 - t\varphi_\mu(\xi)} \right) d\xi = \infty.$$

Démonstration. En vertu de la discussion précédente, il suffit de montrer que la limite considérée est égale à $cg(0)$ où c est une constante strictement positive finie. Posons

$$g_t(0) = \sum_{n \geq 0} t^n \mathbb{P}(S_n=0)$$

de sorte que $g_t(0)$ converge vers $g_1(0) = g(0)$ lorsque $t \uparrow 1$ par convergence monotone. On remarque alors que la fonction caractéristique de S_n est $\hat{\mu}^n$ puis que

$$\mathbb{P}(S_n=0) = \frac{1}{(2\pi)^d} \int_{[-\pi, \pi]^d} \varphi_\mu(\xi)^n d\xi. \quad (9.2)$$

Ceci est un fait général : si ν est une loi de probabilités sur \mathbb{Z}^d , alors

$$\varphi_\nu(\xi) = \sum_{x \in \mathbb{Z}^d} e^{i\langle \xi, x \rangle} \nu(\{x\})$$

et donc, par convergence dominée,

$$\int_{[-\pi, \pi]^d} \varphi_\nu(\xi) d\xi = \sum_{x \in \mathbb{Z}^d} \int_{[-\pi, \pi]^d} e^{i\langle \xi, x \rangle} \nu(\{x\}) d\xi.$$

On voit que cette dernière intégrale est nulle dès que l'une des coordonnées de x est non nulle (car ces coordonnées sont entières), et elle vaut $(2\pi)^d \nu(\{0\})$ sinon. En appliquant ceci à μ^{*n} , de sorte que $\mu^{*n}(\{0\}) = \mathbb{P}(S_n=0)$, on voit que l'on a (9.2). Pour conclure, on somme ces égalités : pour tout $t \in [0, 1[$

$$\sum_{n \geq 0} t^n \mathbb{P}(S_n=0) = \frac{1}{(2\pi)^d} \int_{[-\pi, \pi]^d} \sum_{n \geq 0} (t\varphi_\mu)^n(\xi) d\xi = \frac{1}{(2\pi)^d} \int_{[-\pi, \pi]^d} \frac{1}{1 - t\varphi_\mu(\xi)} d\xi,$$

où la première égalité est justifiée par convergence dominée (c'est la raison pour laquelle on prend $t < 1$). On obtient le résultat en prenant la partie réelle et en faisant tendre t vers 1. \square

Dans le cas particulier de la marche aléatoire simple, on a

$$\varphi_\mu(\xi) = \frac{1}{d} \sum_{k=1}^d \cos(\xi_k).$$

Ceci est une fonction réelle à valeurs dans $[-1, 1]$ égale à 1 seulement aux points de $2\pi\mathbb{Z}^d$, et l'on voit que $(1 - \varphi_\mu(\xi))^{-1}$ est une fonction continue sur $[-\pi, \pi]^d \setminus \{0\}$, équivalente à $(|\xi|^2/2d)^{-1}$ lorsque $\xi \rightarrow 0$. Comme $d \geq 3$, cette fonction est sommable sur $[-\pi, \pi]^d$, et la convergence dominée implique que (notons que dans le cas présent, la partie réelle dans l'intégrale n'est pas nécessaire)

$$\lim_{t \uparrow 1} \int_{[-\pi, \pi]^d} \frac{1}{1 - t\varphi_\mu(\xi)} d\xi = \int_{[-\pi, \pi]^d} \frac{1}{1 - \frac{1}{d} \sum_{k=1}^d \cos(\xi_k)} < \infty.$$

Donc la marche aléatoire est transiente. \square

Avec un peu plus de travail, le critère donné dans le lemme 9.5 permet de montrer le caractère récurrent ou transient de marches aléatoires dans \mathbb{Z}^d beaucoup plus générales. On montre également qu'il est équivalent au critère beaucoup plus naturel

$$\int_{[-\pi, \pi]^d} \Re\left(\frac{1}{1 - \varphi_\mu(\xi)}\right) d\xi = \infty,$$

ce qui correspond formellement à intervertir limite et intégrale dans l'énoncé, mais ceci nécessite beaucoup plus de travail.

Théorème 9.6. *Soit $(S_n, n \geq 0)$ une marche aléatoire sur \mathbb{Z} . On suppose que $\sum_{k \in \mathbb{Z}} |k| \mu(k) < \infty$.*

Alors $(S_n, n \geq 0)$ est récurrente si et seulement si $\sum_{k \in \mathbb{Z}} k \mu(k) = 0$.

Démonstration. Notons $g_N(x) = \sum_{n=0}^N \mathbb{P}(S_n = x)$. Alors on note que $g_N(0) \geq g_N(x)$ pour tout $x \in \mathbb{Z}$. En effet, si l'on note $T_x = \inf \{n \geq 0 : S_n = x\}$, on a

$$\begin{aligned} \sum_{n=0}^N \mathbb{P}(S_n = x) &= \mathbb{E} \left[\sum_{n=T_x}^N \mathbf{1}_{\{S_n = x\}} \right] \\ &= \sum_{k \geq 0} \mathbb{E} \left[\sum_{n=k}^N \mathbf{1}_{\{S_n = x, T_x = k\}} \right] \\ &= \sum_{k \geq 0} \mathbb{E} \left[\sum_{n=k}^N \mathbf{1}_{\{S_{n-k}^{(k)} = 0\}} \right] \mathbb{P}(T_x = k) \\ &\leq \sum_{k \geq 0} \mathbb{P}(T_x = k) \mathbb{E} \left[\sum_{n=0}^N \mathbf{1}_{\{S_n^{(k)} = 0\}} \right] \\ &\leq g_N(0) \end{aligned}$$

où l'on a noté $S_n^{(k)} = X_{k+1} + \dots + X_{k+n}$, comme au début du chapitre, et où l'on a utilisé l'indépendance de $(S_n^{(k)}, n \geq 0)$ et de $\{T_x = k\}$, qui est une conséquence du regroupement par paquets.

On en déduit que pour tout $M \geq 0$, on a

$$g_N(0) \geq \frac{1}{2M+1} \sum_{|x| \leq M} g_N(x).$$

Or

$$\sum_{|x| \leq M} g_N(x) = \sum_{n=0}^N \sum_{|x| \leq M} \mathbb{P}(S_n = x) = \sum_{n=0}^N \mathbb{P}(|S_n| \leq M) \geq \sum_{n=0}^N \mathbb{P}(|S_n/n| \leq M/N)$$

La loi des grands nombres montre que $\mathbb{P}(|S_n/n| \leq a)$ converge vers 1 pour tout $a > 0$. En prenant $M = aN$, on en déduit par le lemme de Cesaro que

$$g(0) = \lim_{N \rightarrow \infty} g_N(0) \geq \lim_{N \rightarrow \infty} \frac{1}{2aN+1} \sum_{n=0}^N \mathbb{P}(|S_n/n| \leq a) = \frac{1}{2a}.$$

Comme $a > 0$ est arbitraire, on conclut que $g(0) = \infty$, et donc que $(S_n, n \geq 0)$ est récurrente. \square

Corollaire 9.7. *Sous les hypothèses du théorème précédent, si μ est d'espérance nulle, et $\mu \neq \delta_0$, alors p.s. on a*

$$\limsup_{n \rightarrow \infty} S_n = \infty \quad \text{et} \quad \liminf_{n \rightarrow \infty} S_n = -\infty.$$

Chapitre 10

Processus de branchement

On cherche à modéliser l'évolution d'une population asexuée dont les individus se reproduisent indépendamment les uns des autres, et suivant la même loi. On se donne ainsi une mesure de probabilités μ sur \mathbb{N} , et l'on appellera processus de branchement toute suite (Z_0, Z_1, \dots) de variables aléatoires à valeurs dans \mathbb{N} telles que pour tout $n \in \mathbb{N}$, et pour tout $z_0, z_1, \dots, z_{n-1}, x, y \in \mathbb{N}$, on a

$$\mathbb{P}(Z_{n+1} = y \mid Z_0 = z_0, Z_1 = z_1, \dots, Z_{n-1} = z_{n-1}, Z_n = x) = \mu^{*x}(y) \quad (10.1)$$

dès lors que cette probabilité est bien définie (c'est-à-dire que l'événement par lequel on conditionne est de probabilité strictement positive), et où l'on a noté μ^{*x} la convolée de μ avec elle-même x fois. On notera que μ^{*x} est la loi de $\xi_1 + \dots + \xi_x$, où les variables aléatoires ξ_1, \dots, ξ_x sont indépendantes de loi μ . Ainsi, la suite (Z_0, Z_1, \dots) décrit bien le modèle indiqué ci-dessus : à la génération n , chacun des x individus présents donne naissance à une famille d'individus dont la taille est aléatoire de loi μ , de façon indépendante entre tous les individus.

On note également que les formules ci-dessus ne disent rien *a priori* sur la loi de la taille de la population initiale Z_0 , qui peut donc être une variable aléatoire quelconque.

On peut se demander s'il existe un processus de branchement de loi μ . En voici une construction explicite, utilisant une famille $(\xi_{n,i} : n \geq 1, i \geq 1)$ de variables aléatoires indépendantes et de même loi μ , ainsi qu'une variable aléatoire Z_0 indépendante de la famille précédente. Si l'on connaît la variable aléatoire Z_n , on pose alors

$$Z_{n+1} = \sum_{i=1}^{Z_n} \xi_{n+1,i}.$$

Par récurrence, ceci permet de définir la suite (Z_0, Z_1, \dots) , de sorte que Z_0, Z_1, \dots, Z_n soit mesurable par rapport à la tribu engendrée par Z_0 et les $\xi_{m,i}$ avec $m \in \{1, 2, \dots, n\}$ et $i \geq 1$. En particulier, $(\xi_{m,i} : m > n, i \geq 1)$ est indépendante de (Z_0, \dots, Z_n) . On a donc, en notant $B = \{Z_n = x, Z_{n-1} = z_{n-1}, \dots, Z_0 = z_0\}$, qui est mesurable par rapport à $\sigma\{Z_0, \dots, Z_n\}$,

$$\begin{aligned} \mathbb{P}(Z_{n+1} = y, B) &= \mathbb{P}\left(\sum_{i=1}^x \xi_{n+1,i} = y, B\right) \\ &= \mathbb{P}\left(\sum_{i=1}^x \xi_{n+1,i} = y\right) \mathbb{P}(B) \\ &= \mu^{*x}(y) \mathbb{P}(B), \end{aligned}$$

d'où (10.1) en divisant par $\mathbb{P}(B)$ lorsque cette quantité est non nulle.

Notons que, d'après la définition, ou par la construction précédente, presque sûrement, s'il existe $n \in \mathbb{N}$ tel que $Z_n = 0$, alors $Z_{n+k} = 0$ pour tout $k \geq 0$. On a donc seulement deux possibilités : ou bien $Z_n > 0$ pour tout $n \geq 0$, ou bien la suite (Z_n) est stationnaire en 0.

Notons $A = \{Z_n > 0 \text{ pour tout } n \geq 0\}$ l'événement de survie de la population. Quand a-t-on $\mathbb{P}(A) > 0$? De quelle façon cette probabilité dépend-elle de Z_0 ?

Dans un premier temps nous allons supposer que $Z_0 = 1$. Nous allons supposer une fois pour toutes que $\mu \neq \delta_1$ pour éviter le cas trivial où $Z_n = 1$ pour tout n . Notons $m = \sum_{k \geq 0} k\mu(k)$ la moyenne de μ , qui est un nombre dans $[0, \infty]$.

Théorème 10.1. *Si $m \leq 1$ alors le processus s'éteint presque sûrement : $\mathbb{P}(A) = 0$.*

Si $m > 1$, alors le processus survit avec probabilité positive : $\mathbb{P}(A) > 0$.

Ce théorème a été démontré par Bienaymé au milieu du XIX^{ème} siècle. Néanmoins, sa preuve ne fut pas publiée, et on n'en retrouve les détails que par reflets dans les ouvrages de son époque. La question fut reposée par Galton vingt ans plus tard, et résolue par Galton et Watson peu après (avec, dit-on, une erreur).

Pour le montrer, nous allons avoir recours à la fonction génératrice de la loi μ :

$$g(s) = \sum_{k \geq 0} \mu(k) s^k.$$

Comme on le sait, g est une fonction convexe, croissante sur $[0, 1]$, dérivable sur $]0, 1[$, et sa dérivée admet en 1 la limite à gauche $g'(1-) = m$.

Lemme 10.2. *Soit (Z_0, Z_1, \dots) un processus de branchement avec $Z_0 = 1$ p.s. Alors la fonction génératrice de Z_n est donnée par la composée n fois de g :*

$$g_{Z_n} = g \circ g \circ \dots \circ g \quad (n \text{ fois}).$$

Démonstration. Nous montrons cette propriété par récurrence. La propriété est claire pour $n = 0$ et $n = 1$, avec la convention que g composée 0 fois avec elle-même est l'identité. Supposons la propriété vraie au rang n . On a alors, pour tout $s \in [0, 1]$,

$$\begin{aligned} g_{Z_{n+1}}(s) &= \mathbb{E}[s^{Z_{n+1}}] = \sum_{x, y \in \mathbb{N}} s^y \mathbb{P}(Z_{n+1} = y, Z_n = x) \\ &= \sum_{x, y \in \mathbb{N}} s^y \mathbb{P}(Z_n = x) \mu^{*x}(y). \end{aligned}$$

La dernière égalité est une application aisée de (10.1). Rappelons que $\mu^{*x}(y)$ est la probabilité que $\xi_1 + \dots + \xi_x = y$ où ξ_1, \dots, ξ_x sont indépendantes de loi μ . En particulier, par indépendance,

$$\sum_{y \in \mathbb{N}} s^y \mu^{*x}(y) = \mathbb{E}[s^{\xi_1 + \dots + \xi_x}] = \mathbb{E}[s^{\xi_1}]^x = g(s)^x.$$

En réinjectant ceci dans l'expression précédente, on trouve

$$g_{Z_{n+1}}(s) = \sum_{x \in \mathbb{N}} \mathbb{P}(Z_n = x) g(s)^x = g_{Z_n}(g(s))$$

et on conclut par hypothèse de récurrence. □

Notons que $g_{Z_n}(0) = \mathbb{P}(Z_n = 0)$, et que cette probabilité croît vers la probabilité $q = \mathbb{P}(A^c)$ d'extinction (rappelons que si $Z_n = 0$ alors $Z_m = 0$ pour tout $m \geq n$). On en déduit que q est la limite de la suite récurrence $q_{n+1} = g(q_n)$ initiée en $q_0 = 0$. Or, la convexité de g implique que

- si $m \leq 1$ alors g a un unique point fixe sur $[0, 1]$, donné par 1. Noter que nous avons utilisé ici le fait que $\mu(1) < 1$, ce qui permet d'éviter le cas trivial où g est la fonction identité. Dans ce cas, la suite (q_n) , à valeurs dans $[0, 1]$, converge vers $q = 1$.
- Si $m > 1$ alors g admet exactement deux points fixes sur $[0, 1]$, que l'on note q' et 1, avec $q' < 1$. Le point q' est alors attractif ($0 < g'(q_0) < 1$) et le point 1 est répulsif, de plus, la suite (q_n) converge vers q' . On a donc $q' = q \in [0, 1]$.

Le théorème s'ensuit en notant que $\mathbb{P}(A) = 1 - q$. On voit que l'on a même donné une caractérisation de q , comme étant le plus petit point fixe de g sur $[0, 1]$.

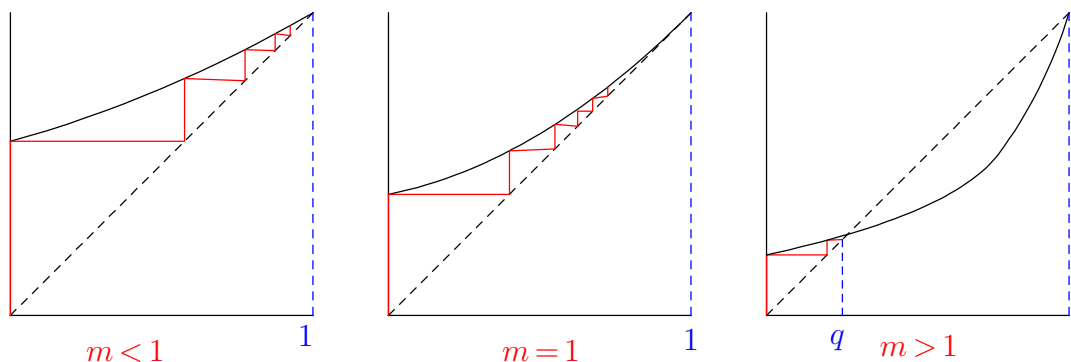


Figure 10.1. Illustration des trois phases : en noir, la courbe représentative de g , en rouge, l'évolution de la suite récurrence (q_n) .

On distingue trois *phases* dans le processus de Galton-Watson.

Phase sous-critique : $m < 1$

Dans ce cas, on a extinction presque sûre ($q = 1$), et les propriétés usuelles des suites récurrentes montrent que la probabilité que Z_n soit non nul converge vers 0 exponentiellement vite. En effet, on a que $1 - q_n \leq m^n$ par applications successives du théorème des accroissements finis, et donc

$$\mathbb{P}(Z_n > 0) = 1 - q_n \leq m^n.$$

L'extinction est donc très rapide.

Phase critique : $m = 1$

Dans ce cas, on a toujours extinction presque sûre. Néanmoins, la convergence de $\mathbb{P}(Z_n > 0)$ vers 0 est typiquement plus lente qu'exponentielle. Pour illustrer ceci, supposons que $\sum_{k \in \mathbb{N}} k^2 \mu(k) < \infty$. Notons que $g''(1-) = \sum_{k \in \mathbb{N}} k(k-1)\mu(k) = \sigma^2$ est la variance de la loi μ , puisque $m = 1$, et cette variance est finie. En développant g au voisinage de 1, on a

$$g(s) = 1 - (1 - s) + \frac{(1 - s)^2}{2} \sigma^2 + o((1 - s)^2)$$

et donc

$$\frac{1}{1-g(s)} - \frac{1}{1-s} = \frac{\sigma^2}{2} + o(1).$$

En sommant on obtient donc

$$\frac{1}{1-q_n} - 1 = n \frac{\sigma^2}{2} + o(n),$$

d'où l'on déduit que

$$\mathbb{P}(Z_n > 0) = 1 - q_n = \frac{2}{\sigma^2 n} (1 + o(1)).$$

La décroissance vers 0 est donc beaucoup plus lente qu'exponentielle. Ce résultat est dû à Kolmogorov.

Phase sur-critique : $m > 1$

Dans ce cas, nous avons vu que le processus survit avec probabilité strictement positive, s'il est issu de $Z_0 = 1$ individu. On peut donner des propriétés plus fines, en décrivant l'allure du processus conditionné par l'événement de survie ou d'extinction. Nous ne donnons pas les preuves ici, mais juste les idées intuitives.

Tout d'abord, on peut constater que le processus de branchement jouit d'une *propriété de branchement* stipulant que, si $Z = (Z_0, Z_1, \dots)$ et $Z' = (Z'_0, Z'_1, \dots)$ sont deux processus de branchement indépendants, de même loi de reproduction μ , et issus respectivement de $Z_0 = x$ et $Z'_0 = x'$ individus, alors $Z + Z' = (Z_0 + Z'_0, Z_1 + Z'_1, \dots)$ est un processus de branchement également (issu de $x + y$), de loi de reproduction μ . Ceci revient intuitivement à dire que deux « familles » distinctes se reproduisent indépendamment dans le futur, en suivant la dynamique du processus de branchement. De cela, on tire par exemple que si $Z_0 = x$ presque sûrement, alors la probabilité de survie est donnée par

$$\mathbb{P}(A) = 1 - q^x,$$

puisqu'on a extinction si et seulement si chacune des x lignées indépendantes s'éteint.

À quoi ressemble le processus (Z_0, Z_1, \dots) si'il est issu de $Z_0 = 1$, et si on le conditionne à s'éteindre ? Nous affirmons que

$$\mathbb{P}(Z_1 = x | A^c) = \frac{1}{q} \mathbb{P}(Z_1 = x, A^c) = \frac{1}{q} \mu(x) q^x = \mu(x) q^{x-1}.$$

Intuitivement, cela signifie que, si l'on sait que l'ancêtre commun à la population a eu x enfants, alors sa descendance s'éteint si et seulement si les x lignées de ces enfants, qui sont indépendantes, s'éteignent, ce qui arrive avec probabilité q^x . On pourrait itérer cet argument, et obtenir le résultat suivant.

Proposition 10.3. *Conditionnellement à A^c , le processus (Z_0, Z_1, \dots) est un processus de branchement sous-critique de loi de reproduction $\mu_q(x) = q^{x-1} \mu(x)$, $x \in \mathbb{N}$.*

Noter que la fonction génératrice de μ_q n'est autre que $g(qs)/q$. On l'interprète en disant que c'est la portion de g comprise entre $[0, q]$, remise à l'échelle linéairement pour en faire une fonction de $[0, 1]$ dans lui-même. En particulier, la moyenne de μ_q est $g'(q)$ qui est bien dans $[0, 1]$, et le processus est sous-critique !

Que se passe-t-il alors si l'on conditionne le processus par l'événement A de non-extinction ? Cette fois, on a

$$\begin{aligned}\mathbb{P}(Z_1 = x | A) &= \frac{1}{1-q} \mathbb{P}(Z_1 = x, A) \\ &= \frac{1}{1-q} \mu(x) \sum_{k=1}^x \binom{x}{k} (1-q)^k q^{x-k} \\ &= \mu(x) \frac{1-q^x}{1-q}.\end{aligned}$$

Ici, l'on a exprimé le fait que, si l'ancêtre commun a eu x enfants, la population totale survit si au moins un de ces enfants a une lignée qui survit. Or les x lignées sont indépendantes et ont la probabilité $1-q$ de survivre, le nombre de lignées qui survivent suit donc une loi binomiale. On constate alors que la dernière formule définit, lorsque x décrit \mathbb{N}^* , une mesure de probabilités sur \mathbb{N}^* , dont la fonction génératrice est donnée par

$$\frac{g(s) - g(qs)}{1-q}.$$

Il faut prendre néanmoins garde au fait que chacun des x enfants de l'ancêtre commun ne va pas se reproduire en suivant cette loi : seuls ceux dont la lignée ne s'éteint pas vont être dans ce cas. On peut donc être plus précis dans le calcul précédent, et noter Z_1^{ext} le nombre d'individus nés de l'ancêtre commun dont la descendance s'éteint, et Z_1^{surv} les autres enfants de l'ancêtre commun. On a alors, pour $x \geq 0$ et $y \geq 1$.

$$\mathbb{P}(Z_1^{\text{ext}} = x, Z_1^{\text{surv}} = y | A) = \mu(x+y) \binom{x+y}{y} (1-q)^{y-1} q^x.$$

Ceci se résume plus facilement comme une fonction génératrice à deux variables :

$$\begin{aligned}g^{\text{surv}}(s, t) &= \mathbb{E}[s^{Z_1^{\text{ext}}} t^{Z_1^{\text{surv}}} | A] = \sum_{x \geq 0, y \geq 1} s^x t^y \mathbb{P}(Z_1^{\text{ext}} = x, Z_1^{\text{surv}} = y | A) \\ &= \frac{1}{1-q} \sum_{k \geq 1} \mu(k) \sum_{y=1}^k \binom{k}{y} (t(1-q))^y (sq)^{k-y} \\ &= \frac{1}{1-q} \sum_{k \geq 1} \mu(k) ((sq + t(1-q))^k - (sq)^k) \\ &= \frac{g(sq + t(1-q)) - g(sq)}{1-q}.\end{aligned}$$

Pour $s = 1$, on obtient la fonction génératrice de Z_1^{surv} sachant A , et cette dernière est $(g(q + t(1-q)) - g(q)) / (1-q)$. Une nouvelle fois, on interprète cette fonction comme la partie de g comprise entre q et 1 , translatée et remise à l'échelle pour en faire la fonction génératrice d'une variable aléatoire. On note $\bar{\mu}_q$ la loi de probabilité associée. On constate que cette fonction est nulle en 0 , c'est-à-dire que la variable aléatoire associée ne peut pas prendre la valeur 0 , et sa dérivée à gauche en 1 est m . On pourrait compléter ce résultat par le résultat suivant, appelé la *décomposition de Harris*. On ne donne pas pour une fois d'énoncé formel.

Conditionnellement à l'événement A de survie du processus, on peut décrire la population de la façon suivante. On a deux types d'individus, les « mortels » et les « immortels ». Les individus immortels donnent naissance à des individus mortels et immortels selon la loi sur \mathbb{N}^2 définie par les coefficients de la fonction génératrice $g^{\text{surv}}(s, t)$, indépendamment les uns des autres. Les individus mortels engendrent des processus de branchement sous-critiques de loi de reproduction μ_q . Enfin, la restriction de la généalogie du processus aux individus immortels forme un processus de branchement sur-critique de loi de reproduction $\bar{\mu}_q$.

Références

- P. Barbé, M. Ledoux, *Probabilité*
- R. Durrett, *Probability: theory and examples*
- W. Feller, *An Introduction to Probability Theory and its Applications*, vol. 1 et 2.
- G. Grimmett, D. Stirzaker, *Probability and random processes*
- J.-F. Le Gall, *Intégration, probabilités et processus aléatoires*, notes de cours disponibles sur <http://www.math.u-psud.fr/~jflegall/IPPA2.pdf>
- M. Pinsky, *Introduction to Fourier analysis and wavelets*
- W. Rudin, *Real and complex analysis*