CentraleSupélec Rennes 2025-2026

PROBABILITÉS

Partie probabilités du cours CIP de CentraleSupélec Rennes.

Arthur Maritch-Roy

Table des matières

Préambule

Ces notes de cours sont très largement inspirées des notes de cours de Jean-Christophe Breton (probabilités L3) et Jean-François le Gall (cours de l'ENS) accessibles sur leurs pages web respectives, ainsi que du livre *Probabilités 2*, de Jean-Yves Ouvrard (éditions Cassini).

Elles ont pour but d'aider les étudiants à suivre la deuxième partie du cours de Convergence, Intégration et Probabilités du cursus FISA de CentraleSupélec (campus de Rennes). Étant donné que le volume horaire est court (trois séances de 1h30), on passe souvent rapidement sur les preuves. Pour l'examen, il ne sera pas nécessaire de connaître les preuves des théorèmes, mais seulement de bien connaître les énoncés et comment ils s'appliquent.

*
* *

Introduction

Contexte historique

Même si on retrouve des traces de probabilités (via des jeux de hasard) dans les civilisations antiques et au moyen-âge, le point de départ dans l'élaboration d'une théorie des probabilités est souvent placé en 1654, avec une correspondance entre Blaise Pascal et Pierre de Fermat sur le problème des partis.
En 1655, Christian Huygens en prend connaissance et rédige un premier traité de probabilités de 1657.

En 1718, Abraham De Moivre publie *The Doctrine of Chances*, qui contient des problèmes combinatoires, des probabilités conditionnelles et une première version de ce qu'on appellera le théorème limite central : le nombre de réussite à des jeux successifs de pile ou face forme, après un grand nombre de réalisations, une courbe en cloche. Lagrange introduit vers 1770 les lois de probabilité continues, qui formalisent cette cloche en une loi normale, dont la densité est une gaussienne, nom donné plus tard en l'honneur de Carl Friedrich Gauss.

Les probabilités ne sont complètement formalisées qu'à partir de 1933 par Andreï Kolmogorov grâce notamment aux travaux de Borel et Lebesgue sur la théorie de la mesure et de l'intégration. Longtemps ensuite les probabilités n'étaient considérées que comme une sous-branche des mathématiques. Par exemple, la première médaille Fields décernée à un probabiliste ne date que de 2006, avec Wendelin Werner et Andreï Okounkov.

Que sont les probabilités ?

L'exemple du lancer de dé. Lorsqu'on lance un dé sur une table, même si la donnée de sa position et de sa vitesse initiale donne un unique résultat par les lois de la mécanique, celui-ci est très sensible aux données initiales et cela le rend "imprévisible" ; c'est la notion de hasard. L'objectif des probabilités est de combler ce défaut, et d'essayer de modéliser ces expériences aléatoires. Pour cela, on attribue une "mesure de probabilité" à chacune des éventualités possibles. La somme des probabilités est postulée comme égale à 1, et pour un dé cubique et équilibré, la bonne mesure de probabilité à attribuer à chacune des faces semble être $1/6$.

On peut alors, par des calculs qui semblent intuitifs, donner la probabilités d'autres événements. La probabilité que le résultat du dé soit pair est par exemple $1/2$. On a ici additionné les mesures des faces paires, au nombre de trois. Si on lance deux dés de manière indépendante, on s'attend à ce que la probabilité d'obtenir un puis un soit le produit des probabilités pour chacun des résultats, soit $1/36$. Avec ces deux postulats, on peut calculer d'autres probabilités, comme la probabilité d'obtenir 7 via le lancer de deux dés.

Mais comment alors vérifier ces calculs ? Que veulent-ils dire ? L'intuition derrière les probabilités est que si on lance un dé plein de fois, la proportion du nombre de lancers où le résultat est pair sur le nombre de lancers total est environ $1/2$, et va se rapprocher de $1/2$ au fur et à mesure que l'on lance des nouveaux dés.

Un des objectifs de ce cours est de retrouver cette intuition, via un théorème qui s'appelle la *loi des grands nombres*.

Les probabilités à l'école. Au début, lorsque l'on commence à faire des probabilités (au collège), on travaille sur un univers fini (les résultats possibles pour notre expérience aléatoire sont en nombre fini) et on fait des calculs de base. Ensuite, au lycée, on aborde des calculs plus compliqués avec les probabilités conditionnelles par exemple et on introduit le formalisme des variables aléatoires. En classes prépa, on définit un peu mieux les objets que l'on manipule, en travaillant avec des tribus, mais uniquement sur un univers dénombrable. En effet, il faut attendre d'avoir étudié la théorie de la mesure et l'intégration pour travailler sur un univers quelconque ; c'est ce que nous allons faire ici.

1 Espace de probabilité

Dans cette partie, on définit l'outil de base pour faire des probabilités. On va définir l'ensemble des événements auxquels on va vouloir attribuer une mesure de probabilité. Pour coller à l'intuition, les événements à considérer vont devoir satisfaire des relations ensemblistes particulières. C'est l'objet de la définition de tribu. Ensuite, on définit une mesure (au sens de la théorie de la mesure) de probabilité sur l'ensemble des événements que l'on considère.

1.1 Définition

Définition 1. Soit $(\Omega, \mathcal{A})$ un espace mesurable (c'est-à-dire un ensemble Ω sur lequel on a une tribu $\mathcal{A} \subset \mathcal{P}(\Omega)$, une tribu étant par définition un ensemble de parties contenant Ω, stable par passage au complémentaire et par réunion *dénombrable*). Une *mesure de probabilité* est une mesure définie sur cet espace telle que $\mathbb{P}(\Omega) = 1$. On dit alors que $(\Omega, \mathcal{A}, \mathbb{P})$ est un *espace de probabilité*.

Remarque 2. On rappelle qu'une mesure est une fonction $\mu : \mathcal{A} \to \mathbb{R}_+$ telle que la mesure de l'ensemble vide soit 0 et telle que la mesure d'une union dénombrable disjointe d'éléments de la tribu soit la somme des mesures.

Exemple 3.
- Sur $(E, \mathcal{P}(E))$, où E est fini, une mesure de probabilité n'est rien d'autre qu'une famille $(p_e)_{e \in E}$ telle que $\sum_{e \in E} p_e = 1$. De même pour un ensemble dénombrable, où la somme est à comprendre au sens des familles sommables.
- Le lancer d'un dé équilibré se modélise par l'espace de probabilité $(\{1, \dots, 6\}, \mathcal{P}(\{1, \dots, 6\}), \mathbb{P})$ où $\mathbb{P}(\{a\}) = 1/6$ pour tout $a \in \{1, \dots, 6\}$. On a alors par exemple $\mathbb{P}(\{2, 4, 6\}) = \mathbb{P}(\{2\}) + \mathbb{P}(\{4\}) + \mathbb{P}(\{6\}) = 1/2$.
- Sur $([0, 1], \mathcal{B}([0, 1]))$, la mesure de Lebesgue λ_1 définit une mesure de probabilité.
- Sur $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d), \lambda_d)$ (λ_d désigne la mesure de Lebesgue sur $\mathbb{R}^d$), pour f une fonction mesurable positive dont l'intégrale $\int_{\mathbb{R}^d} f \, \mathrm{d}\lambda_d = 1$, la mesure

$$\mathbb{P}_f : A \in \mathcal{B}(\mathbb{R}^d) \mapsto \int_A f \, \mathrm{d}\lambda_d$$

  définit une mesure de probabilité.

On peut vérifier tous ces points en exercice.

Proposition 4. (propriétés élémentaires)
Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace de probabilité.

1. Pour $A \in \mathcal{A}, \mathbb{P}(A^c) = 1 - \mathbb{P}(A)$
2. Pour $A, B \in \mathcal{A}$, $\mathbb{P}(A \backslash B) = \mathbb{P}(A) - \mathbb{P}(A \cap B)$
3. Pour $A, B \in \mathcal{A}$, $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$
4. Si $A \subset B$, $\mathbb{P}(A) \leq \mathbb{P}(B)$.

Démonstration. Avec les notations de l'énoncé :

1. On a $\Omega = A \sqcup A^c$ donc $1 = \mathbb{P}(\Omega) = \mathbb{P}(A) + \mathbb{P}(A^c)$.
2. On a $A = (A \cap B) \sqcup (A \backslash B)$ et on procède de même.
3. On utilise $A \cup B = (A \cap B) \sqcup (B \backslash A) \sqcup (A \backslash B)$.
4. On utilise $B = A \sqcup (B \backslash A)$.

$\square$

Proposition 5. (formule du crible)
On peut généraliser la formule pour la probabilité d'une union à un nombre fini quelconque d'événements : si $(\Omega, \mathcal{A}, \mathbb{P})$ est un espace de probabilité, et si $A_1, \ldots, A_n \in \mathcal{A}$:

$$\mathbb{P}\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{k=1}^{n} (-1)^{k-1} \sum_{\substack{J \subset \{1,\ldots,n\} \\ |J|=k}} \mathbb{P}\left(\bigcap_{j\in J} A_j\right) = \sum_{\substack{J \subset \{1,\ldots,n\} \\ J \neq \emptyset}} (-1)^{|J|-1} \mathbb{P}\left(\bigcap_{j\in J} A_j\right).$$

Démonstration. Ce résultat se montre par récurrence sur n. On peut aussi le prouver directement à l'aide des fonctions indicatrices, ce qui est l'objet d'un exercice du TD. □

Proposition 6. (continuité de la mesure)
Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace de probabilité, et $(A_n)_n$ une suite d'événements. Alors

1. $$\mathbb{P}\left(\bigcup_{n=1}^{N} A_n\right) \xrightarrow[N\to\infty]{} \mathbb{P}\left(\bigcup_{n=1}^{\infty} A_n\right).$$

   En particulier, si (A_n) est croissante pour l'inclusion ($A_n \subset A_{n+1} \forall n$),

   $$\mathbb{P}(A_N) \xrightarrow[N\to\infty]{} \mathbb{P}\left(\bigcup_{n=1}^{\infty} A_n\right).$$

2. $$\mathbb{P}\left(\bigcap_{n=1}^{N} A_n\right) \xrightarrow[N\to\infty]{} \mathbb{P}\left(\bigcap_{n=1}^{\infty} A_n\right).$$

   En particulier, si (A_n) est décroissante pour l'inclusion ($A_{n+1} \subset A_n \forall n$),

   $$\mathbb{P}(A_N) \xrightarrow[N\to\infty]{} \mathbb{P}\left(\bigcap_{n=1}^{\infty} A_n\right).$$

3. $$\mathbb{P}\left(\bigcup_{n=1}^{\infty} A_n\right) \leq \sum_{n=1}^{+\infty} \mathbb{P}(A_n).$$

Démonstration. Ce sont des réécritures de résultats de théorie de la mesure. □

Remarque 7. Attention, le deuxième résultat est toujours vrai dans le cas d'une mesure de probabilité, mais n'est pas vrai en général. Par exemple, sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \lambda)$, si $A_n =]n, \infty[$, on a $\lambda(\bigcap_{n=1}^{N} A_n) = +\infty$ mais $\lambda(\bigcap_{n=1}^{\infty} A_n) = \lambda(\emptyset) = 0$.

Définition 8. Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace de probabilité et $A \in \mathcal{A}$.
Si $\mathbb{P}(A) = 0$, on dit que A est *négligeable*. Si $\mathbb{P}(A) = 1$, on dit qu'il est *presque sûr*.

Proposition 9. Une union dénombrable d'événements négligeables est négligeable. Une intersection dénombrable d'événements presque sûrs est presque sûre.

Démonstration. Cela découle de la proposition précédente. □

Remarque 10. Un événement négligeable n'est pas forcément vide, comme le montre l'exemple de $\{1/2\}$ dans l'espace $([0,1], \mathcal{B}([0,1]), \lambda)$.

Définition 11. (limites supérieure et inférieure d'ensembles)
Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace de probabilité. Soit $(A_n) \in \mathcal{A}^{\mathbb{N}}$ une suite d'événements. On définit :

$$\limsup_{n\to\infty} A_n = \bigcap_{n\geq 0} \left(\bigcup_{k\geq n} A_k \right)$$

la *limite supérieure* des A_n et

$$\liminf_{n\to\infty} A_n = \bigcup_{n\in\mathbb{N}} \left(\bigcap_{k\geq n} A_k \right)$$

la *limite inférieure* des A_n. En ayant à l'esprit que l'intersection correspond au fait que tous les événements se réalisent et que l'union correspond au fait qu'un événement se réalise, on voit que $\limsup A_n$ est l'événement "une infinité de A_n se réalise" et que $\liminf A_n$ est l'événement "tous les A_n se réalisent à partir d'un certain rang". Remarquons que l'on a toujours $\liminf A_n \subset \limsup A_n$.

Théorème 12 (premier lemme de Borel-Cantelli)**.** Soit (A_n) une suite d'événements telle que $\sum_{n\in\mathbb{N}} \mathbb{P}(A_n) < \infty$. Alors

$$\mathbb{P}\left(\limsup_{n\to\infty} A_n \right) = 0.$$

Démonstration. On note, pour $n \in \mathbb{N}$, $B_n = \bigcup_{k\geq n} A_k$. Les (B_n) forment une suite d'événements décroissante pour l'inclusion. On a donc

$$\mathbb{P}\left(\limsup_{n\to\infty} A_n \right) = \lim_{n\to\infty} \mathbb{P}(B_n).$$

Mais on a aussi

$$\mathbb{P}(B_n) \leq \sum_{k=n}^{\infty} \mathbb{P}(A_k) \xrightarrow[n\to\infty]{} 0,$$

puisque c'est un reste de série convergente. Par comparaison des limites, on a donc bien le résultat voulu. □

Exemple 13. On considère, pour $n \in \mathbb{N}^*$, un dé équilibré à n^2 faces. En lançant tous ces dés successivement, de manière indépendante, et puisque la somme des inverses des carrés converge, la probabilité d'obtenir une infinité de 1 est nulle.

Remarque 14. Quelle espace de probabilité utilisé pour définir rigoureusement l'expérience aléatoire précédente ? C'est ce que nous allons voir dans la suite.

Théorème 15. (Unicité de la mesure de probabilité)
Deux mesures de probabilité $\mathbb{P}_1$ et $\mathbb{P}_2$ définies sur l'espace probabilisable $(\Omega, \mathcal{A})$ et coïncidant sur un π-système (c'est-à-dire une famille d'événements stable par intersection finie) $\mathcal{C}$ coïncident sur $\sigma(\mathcal{C})$ la tribu engendrée par $\mathcal{C}$.

Démonstration. La preuve utilise le lemme des classes monotones. Si vous ne le connaissez pas, vous pouvez passer cela.
Soit $\mathcal{M} = \{A \in \mathcal{A}, \mathbb{P}_1(A) = \mathbb{P}_2(A)\}$ est une classe monotone (contient Ω, est stable par différence et union dénombrable croissante) contenant $\mathcal{C}$. D'après le lemme des classes monotones, elle contient $\sigma(\mathcal{C})$. □

Définition 16. Soit $((\Omega_n, \mathcal{A}_n))_n$ une suite d'espaces mesurables et $\Omega = \prod_{n\in\mathbb{N}^*} \Omega_n$. On appelle *cylindre* tout ensemble de la forme $A_1 \times \cdots \times A_k \times \prod_{n\geq k+1} \Omega_k$, où les A_i sont des événements de $\mathcal{A}_i$ et où k est un entier quelconque. On appelle *tribu cylindrique* la tribu engendrée par les cylindres (les cylindres forment un π-système).

Exemple 17. Pour modéliser une infinité de lancers de pile ou face, on se place sur l'espace $\{0,1\}^{\mathbb{N}}$ muni de la tribu cylindrique associée. La probabilité que l'on considère est alors l'unique probabilité qui vaut 2^{-n} sur le cylindre $\{1\} \times \ldots \times \{1\} \times \{0,1\} \times \ldots$ (où $\{1\}$ apparaît n fois).

Remarque 18. Il existe un théorème général, appelé théorème d'extension de Kolmogorov, qui permet de définir des probabilités sur la tribu cylindrique à partir des probabilités des cylindres. Dans le suite, on supposera bien définis tous les espaces de probabilité produit que l'on considérera.

1.2 Probabilités discrètes, fonction de répartition

On a déjà vu à quoi correspondait une mesure de probabilité sur un ensemble dénombrable. En particulier, sur $\mathbb{N}$, une probabilité $\mathbb{P}$ est la donnée de la suite $(\mathbb{P}(\{n\}) =: p_n)_n$ qui doit être telle que $\sum_n p_n = 1$. À partir de séries numériques classiques, on obtient des probabilités classiques, dont les utilisations sont courantes.

La **loi de Bernoulli** $\mathcal{B}(p)$ ($p \in [0,1]$) : $\mathbb{P}(\{0\}) = 1-p$, $\mathbb{P}(\{1\}) = p$. Elle correspond à une expérience aléatoire binaire, avec probabilité de succès p.

La **loi binomiale** $\mathcal{B}(n,p)$ ($p \in [0,1], n \in \mathbb{N}^*$) : pour $k \in [\![0,n]\!]$, $\mathbb{P}(\{k\}) = \binom{n}{k} p^k (1-p)^{n-k}$. Elle correspond au nombre de succès après n expériences de Bernoulli indépendantes.

La **loi géométrique** $\mathcal{G}(p)$ ($p \in [0,1]$) : pour $k \in \mathbb{N}^*$, $\mathbb{P}(\{k\}) = (1-p)^{k-1} p$. Elle correspond au rang du premier succès lors d'une suite d'expériences de Bernoulli indépendantes.

La **loi de Poisson** $\mathcal{P}(\lambda)$ ($\lambda > 0$) : pour $k \in \mathbb{N}$, $\mathbb{P}(\{k\}) = \frac{\mathrm{e}^{-\lambda}\lambda^k}{k!}$. Elle modélise les événements rares.

Sur l'espace probabilisable $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, il est moins facile d'avoir l'intuition de ce que peut être une probabilité. La donnée des $(\mathbb{P}(\{x\}))_{x\in\mathbb{R}}$ n'est pas suffisante, comme le montre l'exemple de la loi uniforme sur un segment évoquée en préambule (tous les $\mathbb{P}(\{x\})$ valent 0). Le bon objet à étudier est la *fonction de répartition.*

Définition 19. Soit μ une mesure de probabilité sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. On définit sa *fonction de répartition*, notée F_μ, par

$$F_\mu : x \mapsto \mu(]-\infty, x]).$$

Proposition 20. La fonction de répartition caractérise la mesure. Si deux mesures sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ ont même fonction de répartition alors elles sont égales.

Démonstration. Cela découle du critère d'unicité des probabilités (théorème 15), puisque $\{]-\infty, x], x \in \mathbb{R}\}$ est un π-système qui engendre la tribu borélienne (pour montrer ce dernier point, on peut utiliser la structure des ouverts de $\mathbb{R}$). □

Proposition 21. Soit μ une mesure de probabilité sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. Alors

1. F_μ est croissante
2. F_μ a pour limite 1 en $+\infty$ et 0 en $-\infty$.
3. F_μ est continue à droite en tout point de $\mathbb{R}$.

Démonstration. Le premier point découle du fait que $]-\infty, x] \subset]-\infty, y]$ si $x \leq y$ et de la croissance de la mesure. Pour le deuxième point, on écrit

$$0 = \mu(\emptyset) = \mu\left(\bigcap_{n\in\mathbb{N}}]-\infty, -n]\right) = \lim_{n\to\infty} F(-n),$$

et on fait la même chose de l'autre côté avec $]-\infty, n]$.
Pour le dernier point, c'est toujours la même idée avec $]-\infty, x+\frac{1}{n}]$ pour x fixé et n qui tend vers l'infini. $\square$

Remarque 22. En général, F_μ n'est pas continue à gauche. En effet, si $A_n =]-\infty, x-\frac{1}{n}]$, alors (A_n) est croissante pour l'inclusion mais d'union $]-\infty, x[$. On a alors $\lim_{y\to x^-} F_\mu(y) = F_\mu(x) - \mu(\{x\})$. Ainsi F_μ est continue si et seulement si $\mu(\{x\}) = 0$ pour tout $x \in \mathbb{R}$. On dit alors que μ est *sans atome*.

Exemple 23. La fonction de répartition d'une loi discrète $\mu = \sum_i p_i \delta_{x_i}$ est constante par morceaux, et saute au niveau des x_i pour lesquels p_i est non nul.

La fonction de répartition de la loi uniforme $\mathcal{U}(0,1)$ est la fonction égale à 0 sur $\mathbb{R}_-$, à l'identité sur $]0,1[$ et à 1 sur $[1,\infty[$.

Théorème 24. Soit $F : \mathbb{R} \to \mathbb{R}$ une fonction croissante dont la limite en $+\infty$ et 1 et en $-\infty$ est 0 et continue à droite en tout point de $\mathbb{R}$. Alors F est la fonction de répartition d'une mesure de probabilité μ.

Démonstration. On définit logiquement μ sur les intervalles de la forme suivante : $\mu(]a,b]) = F(b) - F(a)$. On montre que μ s'étend alors en une mesure de probabilité sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ (on ne le fera pas ici, on prouvera le résultat avec le formalisme des variables aléatoires, cf. section 2.). $\square$

Une autre famille importante de mesures de probabilité sur $\mathbb{R}$ sont les mesures à densité.

Définition 25. Soit μ une mesure sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. On dit que μ est *absolument continue* par rapport à la mesure de Lebesgue si pour tout $A \in \mathcal{B}(\mathbb{R})$:

$$\lambda(A) = 0 \Rightarrow \mu(A) = 0.$$

Exemple 26. Si f est une fonction mesurable positive, alors

$$\mu : A \mapsto \int_A f \,\mathrm{d}\lambda$$

définit une mesure absolument continue par rapport à la mesure de Lebesgue.

Le théorème qui suit fournit une réciproque à cet exemple. Toutes les mesures absolument continues par rapport à la mesure de Lebesgue s'écrivent de cette façon.

Théorème 27 (de Radon-Nikodym)**.** Soit μ une mesure σ-finie absolument continue par rapport à la mesure de Lebesgue. Il existe une unique (à égalité presque partout près) fonction f mesurable positive telle que pour tout $A \in \mathcal{B}(\mathbb{R})$, $\mu(A) = \int_A f \,\mathrm{d}\lambda$. La fonction f est appelée *dérivée de Radon-Nikodym* de μ par rapport à λ. On la note parfois $f = \frac{\mathrm{d}\mu}{\mathrm{d}\lambda}$.

On rappelle qu'une mesure σ-finie est une mesure μ telle qu'il existe une suite croissante (E_n) d'union pleine telle que $\mu(E_n)$ soit finie pour tout n.

Démonstration. La démonstration complète est difficile et a peu d'intérêt puisque l'on n'utilisera pas ce théorème. Voyons quand même l'unicité, qui utilise un argument qui peut resservir. Soient f et g deux fonctions qui vérifient le résultat de l'énoncé. On a alors, en considérant l'événement $\{f > g\} = \{x, f(x) > g(x)\} \in \mathcal{B}(\mathbb{R})$.

$$\mu(\{f > g\}) = \int_{f>g} f(x)\,\mathrm{d}x = \int_{f>g} g(x)\,\mathrm{d}x.$$

Donc $\int_{\mathbb{R}} \mathbf{1}_{f(x)>g(x)}(f(x) - g(x))\,\mathrm{d}x = 0$. L'intégrande étant positive, elle est forcément nulle donc $\lambda(\{f > g\}) = 0$. En faisant de même dans l'autre sens, on a $\lambda(\{f < g\}) = 0$. Ainsi, on a bien $f = g$ presque partout. $\square$

On parlera donc, quand elle existe bien, de *la* dérivée de Radon-Nikodym de μ par rapport à λ.

Définition 28. Quand μ est une mesure de probabilité absolument continue par rapport à la mesure de Lebesgue, la fonction $\frac{\mathrm{d}\mu}{\mathrm{d}\lambda}$ est d'intégrale 1 sur $\mathbb{R}$ et est appelée *densité* de μ sur $\mathbb{R}$.

Remarque 29. Plus généralement, on peut parler de densité sur $\mathbb{R}^d$, le théorème de Radon-Nikodym est toujours vrai.

Voyons à présent quelques exemples de densité connues.

La densité **gaussienne** : $g_{\mu,\sigma^2}(x) = \frac{1}{\sqrt{2\pi\sigma^2}}\mathrm{e}^{-\frac{(x-\mu)^2}{2\sigma^2}}$. La probabilité associée est appelée la *loi normale* de paramètres μ et σ^2, notée $\mathcal{N}(\mu, \sigma^2)$.

La densité **exponentielle** : $e_\lambda(x) = \lambda\mathrm{e}^{-\lambda x}\mathbf{1}_{x>0}$. La probabilité associée est appelée la *loi exponentielle* de paramètre λ, notée $\mathcal{E}(\lambda)$.

La densité **uniforme** : $u_{a,b}(x) = \frac{1}{b-a}\mathbf{1}_{[a,b]}(x)$. La probabilité associée est appelée la *loi uniforme* sur $[a, b]$, notée $\mathcal{U}(a, b)$.

Nous verrons plus tard pourquoi la gaussienne possède une place centrale dans la théorie des probabilités, et on verra en TD pourquoi la loi normale apparaît souvent.

Remarque 30. On n'a étudié que des exemples de mesures discrètes et à densité, mais une loi peut être ni l'un ni l'autre.

1.3 Indépendance

L'indépendance est une notion fondamentale en probabilités, qui peut souvent causer des nœuds au cerveau lorsqu'il s'agit de montrer proprement quelque chose. Il faut donc être au point sur les définitions.

Définition 31. Deux éléments A et B de la tribu $\mathcal{A}$ (on se place toujours par défaut sur un espace de probabilité $(\Omega, \mathcal{A}, \mathbb{P})$) sont *indépendants* si $\mathbb{P}(A \cap B) = \mathbb{P}(A)\,\mathbb{P}(B)$.

Remarque 32. C'est une définition qui peut se justifier ; si l'on considère un lancer de deux dés, on s'attend à ce que pour calculer la probabilité de faire deux as, on regarde la probabilité de faire un as une fois que le premier as a été réalisé (idée de raisonnement conditionnel).

Définition 33. Une famille $(A_i)_{i\in I}$ d'événements est une famille d'événements *mutuellement indépendants* si pour tout $J \subset I$, J fini, on a

$$\mathbb{P}\left(\bigcap_{j\in J} A_j\right) = \prod_{j\in J} \mathbb{P}(A_j).$$

Remarque 34. Des événements peuvent être indépendants deux à deux sans être (mutuellement) indépendants. Par exemple, pour un lancer de deux dés classiques, on peut considérer

— $A =$ le premier dé donne un chiffre impair

— $B =$ le deuxième dé donne un chiffre impair

— $C =$ la somme des dés est impaire

et montrer par exemple que $A \cap B \cap C = \emptyset$.

Définition 35. Deux tribus $\mathcal{A}$ et $\mathcal{B}$ définies sur un même ensemble Ω sont *indépendantes* si pour tout $A \in \mathcal{A}$ et $B \in \mathcal{B}$, A et B sont indépendants. Une famille de tribus $(\mathcal{A}_i)_{i\in I}$ (sur un même ensemble Ω) est une famille de tribus (mutuellement) indépendantes si pour tout $(A_i) \in \prod_i \mathcal{A}_i$, les (A_i) sont des événements indépendants.

Proposition 36. Si un π-système $\mathcal{A}$ est indépendant d'une tribu $\mathcal{G}$ alors $\sigma(\mathcal{A})$ est indépendant de $\mathcal{G}$.

Démonstration. On utilise le lemme des classes monotones avec

$$\mathcal{M} = \{A \in \mathcal{F}, A \text{ est indépendant de } \mathcal{G}\}$$

($\mathcal{F}$ dénote la tribu ambiante). $\square$

Remarque 37. On peut adapter le résultat et montrer que si deux π-systèmes sont indépendants alors les tribus qu'ils engendrent le sont également.

Théorème 38 (lemme des coalitions)**.** Soient $(\mathcal{F}_i)_{i\in I}$ des tribus indépendantes et $I = \bigcup_{a\in A} I_a$ une partition de I. Alors les tribus

$$(\mathcal{G}_a)_{a\in A} := \left(\sigma\left(\bigcup_{i\in I_a} \mathcal{F}_i\right)\right)$$

sont indépendantes.

Démonstration. On se ramène à un argument avec des π-systèmes en considérant des intersections finies. Pour une version de la preuve, se référer au poly de Jean-Christophe Breton, accessible sur sa page web. $\square$

Théorème 39 (deuxième lemme de Borel-Cantelli)**.** Soit $(A_n)_n$ une suite d'événements indépendants telle que $\sum_{n=1}^{\infty} \mathbb{P}(A_n) = +\infty$. Alors

$$\mathbb{P}\left(\limsup_{n\to\infty} A_n\right) = 1.$$

Presque sûrement, une infinité de A_n se réalisent.

Démonstration.

$$\begin{aligned}\mathbb{P}\left(\limsup_{n\to\infty} A_n\right) &= 1 - \mathbb{P}\left(\left(\limsup_{n\to\infty} A_n\right)^c\right)\\ &= 1 - \mathbb{P}\left(\liminf_{n\to\infty} A_n^c\right).\end{aligned}$$

On a donc

$$\begin{aligned}\mathbb{P}\left(\liminf_{n\to\infty} A_n^c\right) &= \lim_{n\to\infty} \mathbb{P}\left(\bigcap_{k=n}^{\infty} A_k^c\right)\\ &= \lim_{n\to\infty} \lim_{q\to\infty} \mathbb{P}\left(\bigcap_{k=n}^{q} (1 - \mathbb{P}(A_k))\right)\\ &\leq \lim_{n\to\infty} \lim_{q\to\infty} \mathrm{e}^{-\sum_{k=n}^{q} \mathbb{P}(A_k)}.\end{aligned}$$

On obtient alors le résultat puisque $\sum_{k=n}^{\infty} \mathbb{P}(A_k) = +\infty$ pour tout n. $\square$

Application 40. Dans une infinité de pile ou face, n'importe quelle séquence finie se produit une infinité de fois.

1.4 Probabilités conditionnelles

On se place toujours dans un espace $(\Omega, \mathcal{A}, \mathbb{P})$.

Proposition 41. Pour $B \in \mathcal{A}$ un événement de probabilité non nulle, on peut définir une nouvelle probabilité sur $(\Omega, \mathcal{A})$ par

$$A \in \mathcal{A} \mapsto \mathbb{P}_B(A) = \mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}.$$

Démonstration. Ce ne sont que des vérifications simples. □

Définition 42. On appelle cette mesure la probabilité conditionnelle *sachant* B. On appelle $\mathbb{P}(A|B)$ la probabilité de A sachant B.

Exemple 43. Si on lance un dé classique et que A correspond à faire un et B à avoir un résultat impair, on a

$$\mathbb{P}(A|B) = \frac{1/6}{1/2} = \frac{1}{3},$$

ce qui correspond à l'intuition du "sachant".

Définition 44. Un système complet d'événements est une partition dénombrable de Ω dans laquelle chaque élément est un événement de probabilité non nulle.

Théorème 45 (formules des probabilités totales)**.** Soit $A \in \mathcal{A}$ et $(B_n)_n$ un système complet d'événements. Alors

$$\mathbb{P}(A) = \sum_{n \in \mathbb{N}} \mathbb{P}(A \cap B_n) = \sum_{n \in \mathbb{N}} \mathbb{P}(B_n)\,\mathbb{P}(A|B_n).$$

Démonstration. Puisque (B_n) est un système complet d'événements, A est l'union disjointe des $(A \cap B_n)$. Ainsi on obtient le résultat en utilisant le fait que $\mathbb{P}$ est une probabilité ainsi que la définition de la probabilité conditionnelle. □

Exemple 46. Pour calculer la probabilité que la somme de deux dés vaille 9, on conditionne par le système complet d'événements donné par le résultat du premier dé.

Proposition 47. Soient A, B deux événements, avec $\mathbb{P}(B) \neq 0$. Alors A et B sont indépendants si et seulement si $\mathbb{P}(A|B) = \mathbb{P}(A)$.

Démonstration. Découle de la définition. □

Proposition 48 (formule de Bayes)**.** Soient A et B deux événements de probabilité non nulle. Alors

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

Démonstration. Cela découle des définitions :

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

□

Remarque 49. Lorsque l'on fait des probabilités conditionnelles à partir d'un énoncé en français (avec des mots), il faut faire attention à ne pas confondre $\mathbb{P}(A|B), \mathbb{P}(A \cap B), \mathbb{P}(B|A)$. Parfois, les énoncés peuvent perturber l'intuition et il faut bien introduire les bons événements pour utiliser les bonnes formules. On verra en TD comment bien faire cela.

2 Variables aléatoires

Le point de vue des variables aléatoires est le suivant : on ne s'intéresse plus à toute notre expérience aléatoire, qui peut être trop grosse, mais seulement à une partie. Par exemple, on ne s'intéresse plus aux valeurs de nos deux dés mais seulement à la somme de ces valeurs. On ne s'intéresse plus à une personne mais seulement à sa taille. On va donc construire des applications de Ω dans un ensemble quelconque. Elles vont devoir être compatibles avec notre structure d'espace de probabilité, c'est-à-dire être mesurables.

2.1 Définition et loi d'une variable aléatoire

On se place sur $(\Omega, \mathcal{A}, \mathbb{P})$.

Définition 50. Une variable aléatoire est une application mesurable

$$X : (\Omega, \mathcal{A}) \to (E, \mathcal{E})$$

où $(E, \mathcal{E})$ est un espace mesurable.

Rappels de théorie de la mesure : si $\sigma(\mathcal{C}) = \mathcal{E}$, il suffit, pour montrer que X est une variable aléatoire, de montrer que $X^{-1}(C) \in \mathcal{A}$ pour tout $C \in \mathcal{C}$. Par exemple, pour une variable $X : (\Omega, \mathcal{A}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$, on vérifiera seulement que $\{X \leq t\} \in \mathcal{A}$ pour tout t. On se rappellera aussi qu'une limite simple de variable aléatoires est une variable aléatoire sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$.

Définition 51. Pour $X : (\Omega, \mathcal{A}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$, on parlera de *variable aléatoire réelle.* Si $X(\Omega)$ est fini ou dénombrable, on parlera de variable aléatoire discrète.

Exemple 52. Si $\Omega = \{1, \ldots, 6\}$, $\mathcal{A} = \mathcal{P}(\Omega)$, alors $X : (\omega_1, \omega_2) \mapsto \omega_1 + \omega_2$ est une variable aléatoire. Si $\Omega = \mathbb{R}$, $\mathcal{A} = \mathcal{B}(\mathbb{R})$, alors $X : \omega \mapsto \omega^2 \sin(\omega)$ est une variable aléatoire.

Maintenant, on veut savoir, si on met une probabilité sur notre espace de départ, la probabilité que notre variable aléatoire X prenne telle ou telle valeur. C'est l'objet de la notion de loi.

Définition 53. Soit $X : (\Omega, \mathcal{A}) \to (E, \mathcal{E})$ une variable aléatoire. On munit l'espace de départ d'une probabilité $\mathbb{P}$. La *loi* de X est la mesure image de $\mathbb{P}$ par X, notée $\mathbb{P}_X$:

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

On dit alors que X *suit la loi* $\mathbb{P}_X$, et on note $X \sim \mathbb{P}_X$.

Dans le cas où $X = (X_1, \ldots, X_n) \in \mathbb{R}^n$, on parle de *loi jointe* et les lois des (X_i) sont appelées les lois *marginales.* Connaître la loi d'un couple $(X, Y) \in E_1 \times E_2$ donne accès aux lois marginales par

$$\mathbb{P}_X(A) = \mathbb{P}_{X,Y}(A \times E_2).$$

Définition 54. La fonction de répartition d'une variable réelle X est la fonction de répartition de sa loi $\mathbb{P}_X$, notée alors F_X :

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

Définition 55. Soit F une fonction de répartition. On appelle *fonction quantile* (appelée fonction *inverse généralisée* analyse) associée à F la fonction

$$u \in]0, 1[\mapsto \inf\{t \in \mathbb{R}, F(t) > u\}.$$

Remarque 56. Lorsque la fonction F est continue, la fonction quantile correspond à l'inverse F^{-1} de F (fonction réciproque définie par le théorème de la bijection continue).

Proposition 57. Soit $U : ([0,1], \mathcal{B}([0,1]), \lambda) \to ([0,1], \mathcal{B}([0,1]))$ la variable aléatoire qui à x associe x. Alors si q est la fonction quantile de la fonction de répartition F, $q(U)$ est une variable aléatoire qui a pour fonction de répartition F.

Démonstration. Remarquons que $q(U)$ est bien une variable aléatoire puisque U en est une et q est monotone donc mesurable. Pour $t \in \mathbb{R}$, on a $\{U < F(t)\} \subset \{q(U) \leq t\}$. Par croissance de F, $F(s) > U$ pour tout $s > t$, donc $\{q(U) \leq t\} \subset \{U < F(s)\}$. Ainsi

$$F(t) = \mathbb{P}(U < F(t)) \leq \mathbb{P}(q(U) \leq t) \leq \mathbb{P}(U < F(s)) = F(s).$$

En faisant tendre s vers t, on obtient, par continuité à droite de F :

$$\mathbb{P}(q(U) \leq t) = F(t) \qquad \forall t \in \mathbb{R},$$

ce qui est exactement le résultat souhaité. □

On se sert souvent de cette proposition pour simuler des lois avec un ordinateur.

Remarque 58. On a utilisé le formalisme des variables aléatoires pour montrer le Théorème 24.

Définition 59. Si $\mathbb{P}_X$ est à densité, on dit que X aussi, et $f = \frac{\mathrm{d}\mathbb{P}_X}{\mathrm{d}\lambda}$ est appelée *densité* de X.

Proposition 60. Si X est à densité de densité f, alors pour tout $x \in \mathbb{R}$, $F_X(x) = \int_{-\infty}^{x} f(t)\,\mathrm{d}t$. En particulier, F_X est continue et dérivable sauf en les points de discontinuité de f.

Démonstration. Tout provient de l'expression

$$F(b) - F(a) = \mathbb{P}_X(]a,b]) = \int_a^b f(t)\,\mathrm{d}t.$$

□

Proposition 61. Si (X,Y) est un couple aléatoire de densité $(x,y) \mapsto f(x,y)$, les marginales sont données par

$$\mathbb{P}_X(A) = \int_{A\times\mathbb{R}} f(x,y)\,\mathrm{d}x\mathrm{d}y \text{ et } \mathbb{P}_Y(B) = \int_{\mathbb{R}\times B} f(x,y)\,\mathrm{d}y\mathrm{d}x.$$

Démonstration. C'est une conséquence du théorème de Fubini pour les fonctions positives. □

2.2 Lois usuelles

On a déjà vu des lois discrètes classiques et des densités classiques. Lorsque l'on a une loi μ, on rappelle que l'on écrit $X \sim \mu$ pour dire que X a pour loi μ. Ainsi, $X \sim \mathcal{P}(\lambda)$ veut dire que $\forall k \in \mathbb{N}$, $\mathbb{P}(X = k) = \mathrm{e}^{-\lambda}\frac{\lambda^k}{k!}$, et $X \sim \mathcal{E}(\lambda)$ veut dire que pour tout couple $a < b$, $\mathbb{P}(X \in [a,b]) = \int_a^b \lambda \mathrm{e}^{-\lambda x}\,\mathrm{d}x$.

Se représenter les choses de la sorte est plus parlant que via $(\Omega, \mathcal{A}, \mathbb{P})$. Dorénavant, on se fiche de $(\Omega, \mathcal{A}, \mathbb{P})$ et on dira juste : soit X une variable aléatoire réelle qui suit telle ou telle loi. Dans certains cas, la variable peut avoir un sens dans la vie réelle. Par exemple, la durée de vie d'une ampoule peut être modélisée par une loi $\mathcal{E}(\lambda)$.

Par ailleurs, on supposera que lorsque l'on parle de plusieurs variables aléatoires, celles-ci existent toujours, sur un espace de probabilité que l'on ne précise pas la plupart du temps.

2.3 Moments d'une variable aléatoire

Rappel : (formule de transfert). Si $(A, \mathcal{A})$ et $(B, \mathcal{B})$ sont deux espaces mesurables et $\varphi : A \to B$ est mesurable alors pour μ une mesure sur A, la mesure $\mu_\varphi = \mu \circ \varphi^{-1}$ en est bien une (mesure image de μ par φ) et une fonction $h : (B, \mathcal{B}) \to (\mathbb{C}, \mathcal{B}(\mathbb{C}))$ (on prend ici les complexes car ce sera utile pour parler de fonction caractéristique, cf. plus tard) mesurable est μ_φ-intégrable si et seulement si $h \circ \varphi$ est μ-intégrable, avec

$$\int_A h \circ \varphi \, \mathrm{d}\mu = \int_B h \, \mathrm{d}\mu_\varphi.$$

Définition 62. Soit X une variable aléatoire *positive*. On définit son *espérance*, notée $\mathbb{E}(X)$, par son intégrale sur Ω :

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

Exemple 63. Si $A \in \mathcal{A}$, alors $\mathbb{E}(\mathbf{1}_A) = \mathbb{P}(A)$.

Définition 64. Si X est de signe quelconque, on dit qu'elle est *intégrable* si $\mathbb{E}(|X|) < +\infty$ et on note alors $\mathbb{E}(X) = \int_\Omega X \, \mathrm{d}\,\mathbb{P}$. Si $\mathbb{E}(X) = 0$, on dit que X est *centrée.* On oublie parfois le mot *intégrable* en probabilité, au profit de l'expression *admettre une espérance*, mais on n'oubliera pas qu'admettre une espérance est une notion d'intégrabilité au sens d'intégrale de Lebesgue.

Exemple 65. Toute variable aléatoire bornée admet une espérance.

Proposition 66. Soient X, Y des variables aléatoires réelles intégrables ou positives, a, b des réels.

1. Si $0 \leq X \leq Y$, alors $0 \leq \mathbb{E}(X) \leq \mathbb{E}(Y)$.
2. $\mathbb{E}(aX + bY) = a\,\mathbb{E}(X) + b\,\mathbb{E}(Y)$ (linéarité de l'espérance)
3. Si X est intégrable, $|\mathbb{E}(X)| \leq \mathbb{E}(|X|)$.

Démonstration. Ce sont des résultats de théorie de l'intégration. □

Théorème 67 (inégalité de Markov)**.** Soit X une variable aléatoire *positive*, alors pour tout $\lambda > 0$,

$$\mathbb{P}(X \geq \lambda) \leq \frac{\mathbb{E}(X)}{\lambda}.$$

Démonstration. Soit $\lambda > 0$. On remarque que $\lambda \mathbf{1}_{X \geq \lambda} \leq X$. Par croissance de l'espérance,

$$\lambda\,\mathbb{P}(X \geq \lambda) \leq \mathbb{E}(X),$$

ce qui permet de conclure. □

Proposition 68. Soit $X : (\Omega, \mathcal{A}, \mathbb{P}) \to \mathbb{R}$ de loi $\mathbb{P}_X$. Alors X est intégrable si et seulement si $x \mapsto |x|$ est $\mathbb{P}_X$-intégrable sur $\mathbb{R}$ (c'est-à-dire $\int_\mathbb{R} |x| \, \mathrm{d}\,\mathbb{P}_X(x)$ est fini), et on a

$$\mathbb{E}(X) = \int_\mathbb{R} x \, \mathrm{d}\,\mathbb{P}_X(x).$$

Démonstration. Cela découle de la formule de transfert annoncée plus haut. □

Exemple 69. Si X a pour loi $f \, \mathrm{d}\lambda$, avec f la densité de X, alors X est intégrable si et seulement si $x \mapsto xf(x)$ est intégrable sur $\mathbb{R}$, avec $\mathbb{E}(X) = \int_\mathbb{R} xf(x) \, \mathrm{d}x$. Si $\mathbb{P}_X = \sum_{i \in I} p_i \, \delta_{x_i}$ avec I dénombrable, alors X a une espérance si et seulement si $\sum |x_i| p_i < +\infty$ et l'espérance en question vaut $\sum_{i \in I} x_i p_i$.

Remarque 70. Vu que les espérances sont des intégrales, les théorèmes de Fatou, convergences monotone et dominée s'appliquent.

Théorème 71. Soient X et Y deux variables aléatoires réelles. Elles ont la même loi si et seulement si pour toute fonction mesurable bornée, ou continue bornée, $\mathbb{E}(h(X)) = \mathbb{E}(h(Y))$.

Démonstration. Pour les fonctions mesurables, il suffit de prendre les indicatrices, et pour les fonctions continues, on peut approcher les indicatrices $\mathbf{1}_{\cdot \leq x}$ en croissant par des fonctions continues. Le théorème de convergence monotone donne alors que les variables aléatoires ont même fonction de répartition, et donc ont même loi. $\square$

Théorème 72 (inégalité de Jensen)**.** Soit $\varphi : \mathbb{R} \to \mathbb{R}$ une fonction convexe, X une variable aléatoire réelle intégrable telle que $\mathbb{E}(|\varphi(X)|) < +\infty$. Alors

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

Démonstration. Puisque φ est convexe, en tout point x, pour tout t, il existe $\delta > 0$ tel que

$$\varphi(x) \geq \varphi(t) + \delta(x - t).$$

Pour $t = \mathbb{E}(X)$ et $x = X(\omega)$, on a en intégrant :

$$\mathbb{E}(\varphi(X)) \geq \varphi(\mathbb{E}(X)) + \delta \int_\Omega (X - \mathbb{E}(X)) \,\mathrm{d}\,\mathbb{P},$$

d'où le résultat puisque le dernier terme vaut zéro. $\square$

Exemple 73. On se sert surtout de ce résultat pour la fonction carré.

Définition 74. On dit qu'une variable aléatoire X admet un moment d'ordre $p \geq 1$ lorsque $|X|^p$ admet une espérance finie. Son moment d'ordre p est alors (pour p entier) $\mathbb{E}(X^p)$. L'ensemble des variables qui ont un moment d'ordre p est $\mathrm{L}^p(\Omega, \mathcal{A}, \mathbb{P})$ (où l'égalité entre deux variables aléatoires est l'égalité presque partout). L'ensemble des variables (presque sûrement) bornées est $\mathrm{L}^\infty(\Omega, \mathcal{A}, \mathbb{P})$. On retrouve alors les inégalités de Hölder, Minkowski qui font de ces espaces des espaces de Banach. Comme les mesures de probabilité sont finies, on a $\mathrm{L}^p \subset \mathrm{L}^q$ si $p \geq q$. En particulier, l'espace $\mathrm{L}^2(\Omega, \mathcal{A}, \mathbb{P})$ est un espace de Hilbert, et le produit scalaire est défini par :

$$\langle X, Y \rangle_{\mathrm{L}^2} := \mathbb{E}(XY).$$

Exemple 75. La gaussienne et la loi exponentielle admettent des moments à tous les ordres.

Définition 76. Pour $X \in \mathrm{L}^2(\Omega, \mathcal{A}, \mathbb{P})$, on définit sa *variance* par

$$\mathrm{Var}(X) = \mathbb{E}((X - \mathbb{E}(X))^2)$$

et son *écart-type* par $\sigma_X = \sqrt{\mathrm{Var}(X)}$.

Remarque 77. L'espérance indique la position moyenne de X, tandis que la variance mesure l'écartement moyen à cette position moyenne. La variance se calcule avec des intégrales de la même manière que l'espérance. Une variable de variance 1 est dite *réduite*.

Proposition 78. Soit $X \in \mathrm{L}^2$, et a, b des réels.

1. $\mathrm{Var}(X) \geq 0$
2. $\mathrm{Var}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2$
3. $\mathrm{Var}(aX) = a^2 \,\mathrm{Var}(X)$
4. $\mathrm{Var}(X + b) = \mathrm{Var}(X)$

5. $\mathrm{Var}(X) = 0$ si et seulement si $X = \mathbb{E}(X)$ presque sûrement.

Démonstration. Ce ne sont que des vérifications simples. $\square$

Définition 79. Soient $X, Y \in \mathrm{L}^2$. On définit leur *covariance* par

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

L'application Cov est une forme bilinéaire symétrique sur $\mathrm{L}^2(\Omega, \mathcal{A}, \mathbb{P})$ avec $\mathrm{Cov}(X, X) = \mathrm{Var}(X)$. On a aussi

$$\mathrm{Var}(X + Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) + 2\,\mathrm{Cov}(X, Y).$$

Deux variables dont la covariance est nulle sont dites *décorrélées*. On a alors $\mathrm{Var}(X + Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)$.

Théorème 80 (inégalité de Bienaymé-Tchebychev)**.** Soit $X \in \mathrm{L}^2$. Alors

$$\mathbb{P}(|X - \mathbb{E}(X)| \geq t) \leq \frac{\mathrm{Var}(X)}{t^2}.$$

Démonstration. On applique l'inégalité de Markov à la variable $(X - \mathbb{E}(X))^2$ pour $\lambda = t^2$. $\square$

Définition 81. Pour X une variable aléatoire réelle, on définit sa *fonction caractéristique* φ_X par

$$\varphi_X(t) := \mathbb{E}(\mathrm{e}^{\mathrm{i}tX}) = \int_{\mathbb{R}} \mathrm{e}^{\mathrm{i}tx}\,\mathrm{d}\,\mathbb{P}_X(x),$$

qui est bien définie car $|\mathrm{e}^{\mathrm{i}tX}| = 1$.

Remarque 82.
1. On définit la fonction caractéristique de $X \in \mathbb{R}^d$ par $\varphi_X(t) = \mathbb{E}(\mathrm{e}^{\mathrm{i}\langle t, X\rangle})$ où $t \in \mathbb{R}^d$ et $\langle \cdot, \cdot \rangle$ désigne le produit scalaire canonique. Les résultats énoncés resteront vrais pour des vecteurs aléatoires.
2. La fonction caractéristique est, comme la fonction de répartition, à définir sur les mesures de probabilité.
3. La fonction caractéristique est une sorte de transformée de Fourier de la loi.

Théorème 83. Si $a < b$ dont deux réels et si μ est une mesure de probabilité, et φ la fonction caractéristique associée, alors

$$\frac{1}{2}\mu(\{a, b\}) + \mu(]a, b[) = \lim_{T \to +\infty} \frac{1}{2\pi} \int_{-T}^{T} \frac{\mathrm{e}^{-\mathrm{i}ta} - \mathrm{e}^{-\mathrm{i}tb}}{\mathrm{i}t} \varphi(t)\,\mathrm{d}t.$$

Ainsi, la fonction caractéristique caractérise la loi.

Démonstration. Il n'est pas utile de retenir cette preuve, mais elle fait un bon exercice d'intégration.

On note $I_T(x) = \frac{1}{2\pi}\int_{-T}^{T} \frac{\mathrm{e}^{-\mathrm{i}ta} - \mathrm{e}^{-\mathrm{i}tb}}{\mathrm{i}t} \mathrm{e}^{\mathrm{i}tx}\,\mathrm{d}t$. La fonction, $(x, t) \mapsto \frac{\mathrm{e}^{-\mathrm{i}ta} - \mathrm{e}^{-\mathrm{i}tb}}{\mathrm{i}t} \mathrm{e}^{\mathrm{i}tx}$ est bornée sur $[-T, T] \times \mathbb{R}$. Alors elle est intégrable contre la mesure finie $\lambda_{[-T,T]} \otimes \mu$. Par le théorème de Fubini :

$$\frac{1}{2\pi} \int_{-T}^{T} \frac{\mathrm{e}^{-\mathrm{i}ta} - \mathrm{e}^{-\mathrm{i}tb}}{\mathrm{i}t} \varphi(t)\,\mathrm{d}t = \frac{1}{2\pi} \int_{\mathbb{R}} I_T(x)\,\mathrm{d}\mu(x).$$

On va chercher à faire une convergence dominée. Étudions pour cela en détail l'application I_T. Pour $x \notin \{a, b\}$, on a, par changement de variable :

$$I_T(x) = \frac{1}{2\pi} \int_{-T(x-a)}^{T(x-a)} \frac{\sin(u)}{u}\,\mathrm{d}u - \frac{1}{2\pi} \int_{-T(x-b)}^{T(x-b)} \frac{\sin(u)}{u}\,\mathrm{d}u,$$

ce qui tend vers $\mathbf{1}_{]a,b[}(x)$ lorsque T tend vers l'infini. Pour $x \in \{a,b\}$, on fait le calcul à la main et somme toute :

$$\lim_{T\to\infty} I_T(x) = \frac{1}{2}\mathbf{1}_{\{a,b\}}(x) + \mathbf{1}_{]a,b[}(x).$$

Ainsi, on a en particulier le caractère borné de $(T,x) \mapsto I_T(x)$. Par convergence dominée, on obtient alors

$$\lim_{T\to\infty} \frac{1}{2\pi}\int_{-T}^{T} \frac{\mathrm{e}^{-\mathrm{i}ta} - \mathrm{e}^{-\mathrm{i}tb}}{\mathrm{i}t}\varphi(t)\,\mathrm{d}t = \frac{1}{2\pi}\int_{\mathbb{R}} \lim_{T\to\infty} I_T(x)\,\mathrm{d}\mu(x),$$

on a donc le résultat annoncé dans l'énoncé. En refaisant tout ce raisonnement pour $\frac{1}{2\pi}\int_{-T}^{T} \mathrm{e}^{-\mathrm{i}tb}\varphi(t)\,\mathrm{d}t$, on trouve que $\lim_{T\to\infty} \frac{1}{2\pi}\int_{-T}^{T} \mathrm{e}^{-\mathrm{i}tb}\varphi(t)\,\mathrm{d}t = \mu(\{b\})$. Ainsi, si deux mesures réelles finies ont la même fonction caractéristique, elles coïncident sur les segments. On conclut alors quant à l'égalité des mesures par l'argument classique des classes monotones. □

Proposition 84. Soit $X \in \mathbb{R}$ une variable aléatoire.

1. $|\varphi_X(t)| \leq 1$ pour tout $t \in \mathbb{R}$.
2. $\varphi_X(-t) = \overline{\varphi_X(t)} \qquad \forall t \in \mathbb{R}$.
3. $\varphi_X(0) = 1$.
4. φ_X est uniformément continue.

Démonstration. On ne démontre que le point 4. Soit $t \in \mathbb{R}$ et $h > 0$,

$$\varphi_X(t+h) - \varphi_X(t) = \mathbb{E}\left(\mathrm{e}^{\mathrm{i}tX}\mathrm{e}^{\frac{\mathrm{i}}{2}hX}2\mathrm{i}\sin\left(\frac{hX}{2}\right)\right).$$

D'où

$$|\varphi_X(t+h) - \varphi_X(t)| \leq 2\,\mathbb{E}\left(\left|\sin\left(\frac{hX}{2}\right)\right|\right) \leq h|X| \wedge 2,$$

qui tend vers 0 indépendamment de t (le symbole $\wedge$ désigne le minimum). □

Remarque 85. Si $\varphi_X \in \mathrm{L}^1(\mathbb{R})$, le théorème d'inversion de Fourier classique donne que X admet une densité f et que $f(x) = \frac{1}{2\pi}\int_{\mathbb{R}} \mathrm{e}^{-\mathrm{i}tx}\varphi_X(t)\,\mathrm{d}t$ pour tout $x \in \mathbb{R}$.

Proposition 86. Si une variable aléatoire réelle X admet un moment d'ordre $k \in \mathbb{N}^*$, alors φ_X est k fois dérivable, avec pour $j \leq k$:

$$\varphi_X^{(j)}(t) = \mathrm{i}^j\,\mathbb{E}(X^j\mathrm{e}^{\mathrm{i}tX}).$$

En particulier, $\varphi_X^{(j)}(0) = \mathrm{i}^j\,\mathbb{E}(X^j)$.

Démonstration. C'est un théorème de convergence dominée avec la majoration

$$\left|\mathrm{e}^{\mathrm{i}t} - 1 - \mathrm{i}t - \ldots - \frac{(\mathrm{i}t)^{n-1}}{(n-1)!}\right| \leq \frac{|t|^n}{n!}.$$

□

Ce résultat présente aussi une réciproque partielle, que l'on ne présente pas ici.

Pour les variables aléatoires à valeurs dans $\mathbb{N}$, une autre transformée est commode, c'est la fonction (ou série) génératrice.

Définition 87. Soit X une variable aléatoire à valeurs dans $\mathbb{N}$. Sa *fonction génératrice* définie (au moins) sur $\mathrm{D}(0,1)$ est la série entière $\mathbb{E}(z^X) = \sum_{n\in\mathbb{N}} \mathbb{P}(X=n)z^n =: \mathrm{G}_X(z)$. Par unicité du développement en série entière sur un voisinage de zéro, la série génératrice caractérise la loi de X.

Théorème 88. Pour $x \in \mathbb{R}$, on pose $(x)_k = x(x-1)\cdots(x-k+1)$. Si $\mathbb{E}((X)_k)$ est fini alors G_X est dérivable k fois à gauche en 1 avec $\mathbb{E}((X)_k) = \mathrm{G}_X^{(k)}(1)$.

Démonstration. Il s'agit d'un théorème de dérivation sous la somme. □

Remarque 89. La réciproque est vraie, mais on ne le montre pas ici.

Récapitulatif des lois usuelles :

Nom de la loi	Loi	Espérance	Variance	Fonction d'intérêt
Loi de Bernoulli $X \sim \mathcal{B}(p)$	$\mathbb{P}(X=0) = 1-p, \mathbb{P}(X=1) = p$	$\mathbb{E}(X) = p$	$\mathrm{Var}(X) = 1-p$	$\mathrm{G}_X(t) = pt + (1-p)$
Loi binomiale $X \sim \mathcal{B}(n,p)$	$\mathbb{P}(X=k) = \binom{n}{k} p^k (1-p)^{n-k}$	$\mathbb{E}(X) = np$	$\mathrm{Var}(X) = np(1-p)$	$\mathrm{G}_X(t) = (pt+1-p)^n$
Loi de Poisson $X \sim \mathcal{P}(\lambda)$	$\mathbb{P}(X=k) = \mathrm{e}^{-\lambda} \frac{\lambda^k}{k!}$	$\mathbb{E}(X) = \lambda$	$\mathrm{Var}(X) = \lambda$	$\mathrm{G}_X(t) = \mathrm{e}^{\lambda(t-1)}$
Loi géométrique $X \sim \mathcal{G}(p)$	$\mathbb{P}(X=k) = (1-p)^{k-1} p^k$	$\mathbb{E}(X) = 1/p$	$\mathrm{Var}(X) = (1-p)/p^2$	$\mathrm{G}_X(t) = \frac{pt}{1-(1-p)t}$
Loi uniforme $X \sim \mathcal{U}(a,b)$	$f_X(x) = \frac{1}{b-a} \mathbf{1}_{[a,b]}(x)$	$\mathbb{E}(X) = \frac{a+b}{2}$	$\mathrm{Var}(X) = \frac{(b-a)^2}{12}$	$\varphi_X(t) = \frac{\mathrm{e}^{ibt} - \mathrm{e}^{iat}}{\mathrm{i}(b-a)t}$
Loi exponentielle $X \sim \mathcal{E}(\lambda)$	$f_X(x) = \lambda \mathrm{e}^{-\lambda x} \mathbf{1}_{x>0}$	$\mathbb{E}(X) = \frac{1}{\lambda}$	$\mathrm{Var}(X) = \frac{1}{\lambda^2}$	$\varphi_X(t) = \frac{\lambda}{\lambda - \mathrm{i}t}$
Loi normale $X \sim \mathcal{N}(m, \sigma^2)$	$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \mathrm{e}^{-\frac{(x-\mu)^2}{2\sigma^2}}$	$\mathbb{E}(X) = m$	$\mathrm{Var}(X) = \sigma^2$	$\varphi_X(t) = \mathrm{e}^{imt - \frac{1}{2}\sigma^2 t^2}$
Loi de Cauchy $X \sim \mathcal{C}(a,b)$	$f_X(x) = \frac{a}{\pi(a^2 + (x-b)^2)}$	non défini	non définie	$\varphi_X(t) = \mathrm{e}^{\mathrm{i}bt - a\lvert t \rvert}$

2.4 Indépendance de variables aléatoires

Voyons, pour clôturer ce chapitre, comment on formalise la notion d'indépendance dans le cadre de variables aléatoires.

Définition 90. Des variables aléatoires $(X_i)_{i \in I}$ sont indépendantes si les tribus $(\sigma(X_i))_{i \in I}$ le sont.

Puisque les tribus sont en général des gros objets que l'on ne connaît pas, on utilisera plus souvent la proposition suivante.

Proposition 91. Un vecteur aléatoire $(X_1, \ldots, X_n)$ est à composantes indépendantes si et seulement si

$$\mathbb{P}_{X_1, \ldots, X_n} = \mathbb{P}_{X_1} \otimes \cdots \otimes \mathbb{P}_{X_n}$$

Démonstration. Soit $A = A_1 \times \cdots \times A_n$ un événement de la tribu produit. On a

$$\mathbb{P}_{X_1, \ldots, X_n}(A) = \mathbb{P}(X_1 \in A_1, \ldots, X_n \in A_n) = \mathbb{P}(X_1 \in A_1) \cdots \mathbb{P}(X_n \in A_n) = \mathbb{P}_{X_1} \otimes \cdots \otimes \mathbb{P}_{X_n}(A).$$

Puisque les événements de la forme précédente engendrent la tribu produit, on obtient le résultat. □

Proposition 92. Deux variables X et Y discrètes sont indépendantes si et seulement si pour tout $x \in X(\Omega)$ et pour tout $y \in Y(\Omega)$,

$$\mathbb{P}(X = x, Y = y) = \mathbb{P}(X = x)\,\mathbb{P}(Y = y).$$

Deux variables à densité sur $\mathbb{R}$ sont indépendantes si et seulement si (X, Y) est à densité, de densité $x, y \mapsto f_X(x) f_Y(y)$.

Démonstration. Cela découle de la proposition précédente. $\square$

Proposition 93 (lemme des coalitions, version variables aléatoires)**.** Soient $(X_1, \dots, X_n)$ des variables aléatoires indépendantes, et $f : \mathbb{R}^k \to \mathbb{R}, g : \mathbb{R}^{n-k} \to \mathbb{R}$. Alors $Y := f(X_1, \dots, X_k)$ et $Z := g(X_{k+1}, \dots, X_n)$ sont indépendantes.

Démonstration. La variable Y est $\sigma\left(\bigcup_{i=1}^{k} \sigma(X_i)\right)$-mesurable et Z est $\sigma\left(\bigcup_{i=k+1}^{n} \sigma(X_i)\right)$-mesurable. Ces deux tribus sont indépendantes par le lemme des coalitions classique (Théorème 38), donc Y et Z sont indépendantes par définition de ce que sont des variables indépendantes. $\square$

Exemple 94. Si (X, Y, Z) sont trois variables aléatoires réelles indépendantes, $X + Z^2$ est indépendante de $\sin(Y)$.

Proposition 95. Soient $(X_i)_{i\in I}$ des variables aléatoires. Elles sont indépendantes si et seulement si pour toute famille $J \subset I$ finie et pour toutes fonctions mesurables (h_i) bornées (ou intégrables),

$$\mathbb{E}\left(\prod_{j\in J} h_j(X_j)\right) = \prod_{j\in J} \mathbb{E}(h_j(X_j)).$$

Démonstration. Pour le sens réciproque, il suffit de considérer les indicatrices. Pour le sens direct, on utilise la théorème de Fubini :

$$\begin{aligned}\mathbb{E}\left(\prod_{j\in J} h_j(X_j)\right) &= \int \prod_{j\in J} h_j(x_j)\,\mathrm{d}\,\mathbb{P}_{X_1,\dots,X_n}(x_1, \dots, x_n)\\ &= \prod_{j\in J} \int h_j(x_j)\,\mathrm{d}\,\mathbb{P}_{X_j}(x_j)\\ &= \prod_{j\in J} \mathbb{E}(h_j(X_j)).\end{aligned}$$

$\square$

Exemple 96. Si deux variables $X, Y \in \mathrm{L}^2$ sont indépendantes, $\mathbb{E}(XY) = \mathbb{E}(X)\,\mathbb{E}(Y)$ et donc $\mathrm{Var}(X + Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)$ et $\mathrm{Cov}(X, Y) = 0$. La réciproque est fausse, prendre $X \sim \mathcal{U}(-1, 1)$ et $Y = X^2$.

Proposition 97. Deux variables réelles X, Y sont indépendantes si et seulement si pour tout $t, s \in \mathbb{R}$,

$$\varphi_{(X,Y)}(t, s) = \varphi_X(t)\varphi_Y(s).$$

De même pour n variables aléatoires $X_1, \dots, X_n$.

Démonstration. On utilise la caractérisation de la loi par la fonction caractéristique et le théorème de Fubini. $\square$

Rappel (convolution de mesures) : Soient μ et ν deux mesures quelconques définies sur un espace vectoriel mesurable $(E, \mathcal{A})$. On définit la *convoluée* de μ et ν par

$$\mu * \nu(A) = \int_E \mu(A - x)\,\mathrm{d}\nu(x),$$

avec $A - x = \{a - x, a \in A\}$. La fonction d'ensemble $\mu * \nu$ est une mesure, et c'est une probabilité si μ et ν le sont. De plus,

$$\mu * \nu(A) = \iint \mathbf{1}_A(x + y)\,\mathrm{d}\nu(x)\mathrm{d}\mu(x).$$

Proposition 98. Soient X et Y deux variables aléatoires indépendantes à valeurs dans $\mathbb{R}^n$. Alors la loi de $X+Y$ est $\mathbb{P}_{X+Y} = \mathbb{P}_X * \mathbb{P}_Y$.

Démonstration. Pour $A \in \mathcal{B}(\mathbb{R}^n)$,

$$\begin{aligned}\mathbb{P}_{X+Y}(A) = \mathbb{E}(\mathbf{1}_A(x+y)) &= \iint \mathbf{1}_A(x+y)\,\mathrm{d}\,\mathbb{P}_{(X,Y)}(x,y) \\ &= \iint \mathbf{1}_A(x+y)\,\mathrm{d}\,\mathbb{P}_X(x)\mathrm{d}\,\mathbb{P}_Y(y) \\ &= \mathbb{P}_X * \mathbb{P}_Y(A).\end{aligned}$$

□

Remarque 99. En particulier, si X a pour densité f et Y a pour densité g et si X et Y sont indépendantes, $X+Y$ a pour densité $f*g$, où $*$ désigne ici la convolution dans L^1, définie via le théorème de Fubini :

$$f * g : x \mapsto \int_{\mathbb{R}} f(t)g(x-t)\,\mathrm{d}t.$$

Proposition 100. Soient X, Y deux variables aléatoires réelles. Si elles sont indépendantes, alors $\varphi_{X+Y} = \varphi_X\varphi_Y$. Idem pour une collection de n variables.

Démonstration. Soit $t \in \mathbb{R}$.

$$\varphi_{X+Y}(t) = \mathbb{E}(\mathrm{e}^{\mathrm{i}t(X+Y)}) = \mathbb{E}(\mathrm{e}^{\mathrm{i}tX}\mathrm{e}^{\mathrm{i}tY}) = \mathbb{E}(\mathrm{e}^{\mathrm{i}tX})\,\mathbb{E}(\mathrm{e}^{\mathrm{i}tY}) = \varphi_X(t)\varphi_Y(t).$$

□

Proposition 101. Si X et Y sont deux variables aléatoires indépendantes à valeurs dans $\mathbb{N}$, alors $\mathrm{G}_{X+Y}(t) = \mathrm{G}_X(t)\mathrm{G}_Y(t)$ pour tout $t \in [0,1]$. De même pour n variables.

Démonstration. Soit $t \in [0,1]$:

$$\mathrm{G}_{X+Y}(t) = \mathbb{E}(t^{X+Y}) = \mathbb{E}(t^X)\,\mathbb{E}(t^Y) = \mathrm{G}_X(t)\mathrm{G}_Y(t).$$

□

Application 102. On peut se servir de ces deux résultats pour montrer qu'une somme de lois normales indépendantes est une loi normale, où les moyennes et les variances se somment, ainsi que pour montrer qu'une somme de lois de Poisson indépendantes est une loi de Poisson, où les paramètres se somment.

Remarque 103. Une loi binomiale de paramètres n, p est la somme de n lois de Bernoulli de paramètre p indépendantes.

Théorème 104 (loi faible des grands nombres)**.** Soient $(X_1, \ldots, X_n)$ des variables L^2 indépendantes et de même loi (on dit souvent i.i.d pour indépendantes et identiquement distribuées). Alors

$$\mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^n X_i - \mathbb{E}(X_1)\right| \geq \varepsilon\right) \leq \frac{\mathrm{Var}(X_1)}{n\varepsilon^2}.$$

Démonstration. C'est une application de l'inégalité de Bienaymé-Tchebychev à la variable aléatoire $\frac{1}{n}\sum_{i=1}^n X_i$. □

Ce théorème explique que la moyenne dite *empirique*, notée $\overline{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$ va approcher la *vraie* moyenne, c'est-à-dire l'espérance, avec grande probabilité quand n tend vers l'infini. En un sens, que l'on va préciser dans la section suivante, les moyennes empiriques convergent vers l'espérance. L'objectif de la prochaine section est de donner des notions de convergence pour les variables aléatoires, et d'obtenir des théorèmes de convergence, comme celui que l'on vient de voir.

3 Convergence de variables aléatoires

3.1 Convergence en probabilité, presque sûre

On commence par deux modes de convergence propres aux probabilités. On se place par défaut sur un espace de probabilité $(\Omega, \mathcal{A}, \mathbb{P})$.

Définition 105. On dit qu'une suite $(X_n)_n$ de variable aléatoires réelles converge *en probabilité* vers une variable aléatoire X (définie sur le même espace de probabilité) si pour tout $\varepsilon > 0$:

$$\mathbb{P}(|X_n - X| \geq \varepsilon) \xrightarrow[n\to\infty]{} 0.$$

On note $X_n \xrightarrow[n\to\infty]{\mathbb{P}} X$.

Remarque 106. La limite en probabilité est unique (décomposer l'événement $|X - Y| \geq \varepsilon$ si X, Y sont deux limites en probabilité). Le résultat de la loi faible des grands nombres est une convergence en probabilité. On peut montrer que la convergence en probabilité est métrisable ($X_n \xrightarrow{\mathbb{P}} X$ si et seulement si $d(X_n, X)$ tend vers 0 pour une certaine distance d sur l'ensemble des variables aléatoires), et fait même de l'espace des variables aléatoires réelles sur $(\Omega, \mathcal{A}, \mathbb{P})$ un espace complet.

Énonçons maintenant la définition d'une autre forme de convergence, qui est plus simple à manipuler mais plus forte.

Définition 107. On dit que $(X_n)_n$ converge *presque sûrement* vers X (définie sur le même espace de probabilité) si

$$\mathbb{P}(\lim_{n\to\infty} X_n = X) = 1,$$

c'est-à-dire l'événement $\{\omega \in \Omega, X_n(\omega) \xrightarrow[n\to\infty]{} X(\omega)\}$ est de probabilité 1. On note $X_n \xrightarrow[n\to\infty]{\text{ps}} X$.

Proposition 108. La convergence presque sûre entraîne la convergence en probabilité, mais la réciproque est fausse.

Démonstration. Pour $\varepsilon > 0$, la suite $(\mathbf{1}_{\{X_n - X > \varepsilon\}})_n$ est une suite de variable aléatoires qui converge presque sûrement vers 0 si X_n converge presque sûrement vers X. Ainsi, par convergence dominée,

$$\mathbb{P}(|X_n - X| > \varepsilon) = \mathbb{E}(\mathbf{1}_{\{X_n - X > \varepsilon\}}) \xrightarrow[n\to\infty]{} 0.$$

□

Ainsi, la limite presque sûre est unique. Pour obtenir une convergence presque sûre à partir d'une convergence en probabilité, on utilise le résultat suivant, qui fournit donc une réciproque partielle au résultat que l'on vient de montrer.

Théorème 109. Soit $(X_n)_n$ une suite de variables aléatoires réelles.

1. Si pour tout $\varepsilon > 0$, la série $\sum_n \mathbb{P}(|X_n - X| > \varepsilon)$ converge, alors X_n converge ps vers X.
2. Si les $(X_n)_n$ sont de plus indépendantes, alors X_n converge ps vers 0 si et seulement si la série $\sum_n \mathbb{P}(|X_n| > \varepsilon)$ converge pour tout $\varepsilon > 0$.

Démonstration. Les deux résultats sont deux reformulations des deux lemmes de Borel-Cantelli.

□

Exemple 110. Comme annoncé précédemment, il existe des suites qui convergent en probabilité mais pas presque sûrement. Une suite de variables indépendantes de lois $(\mathcal{B}(1/n))_n$ tend en probabilité vers 0, mais ne tend pas presque sûrement vers 0 (ni vers aucune autre variable aléatoire par unicité de la limite en probabilité).

Remarque 111. Puisque toute suite admet une sous-suite dont la série converge, le résultat précédent montre que de toute suite de variables aléatoires qui converge en probabilité, on peut extraire une sous-suite qui converge presque sûrement.

Proposition 112. Soient (X_n) des variables aléatoires réelles, et $g : \mathbb{R} \to \mathbb{R}$ une fonction continue. Si X_n converge ps vers X, alors $g(X_n)$ converge ps vers $g(X)$. Idem pour la convergence en probabilité.

Démonstration. Soit $\omega \in \Omega$ tel que $X_n(\omega) \xrightarrow[n\to\infty]{} X(\omega)$. Alors par continuité de g, $g(X_n(\omega)) \xrightarrow[n\to\infty]{} g(X(\omega))$, ce qui donne le résultat.

Maintenant, si $X_n \xrightarrow{\mathbb{P}} X$, soit $\varepsilon > 0$. Pour $\delta > 0$, on considère

$$B_\delta = \{x \in \mathbb{R}, \exists y \in \mathbb{R}, |x-y| < \delta, |g(x) - g(y)| > \varepsilon\}.$$

Par continuité de g,

$$\bigcap_{\substack{\delta>0\\ \delta\in\mathbb{Q}}} B_\delta = \emptyset.$$

De plus,

$$\mathbb{P}(|g(X_n) - g(X)| > \varepsilon) \leq \mathbb{P}(|X_n - X| \geq \delta) + \mathbb{P}(X \in B_\delta).$$

Ainsi,

$$\limsup_{n\to\infty} \mathbb{P}(|g(X_n) - g(X)| \geq \varepsilon) \leq \mathbb{P}(X \in B_\delta) \xrightarrow[\delta\to 0]{} 0,$$

d'où le résultat. □

Remarque 113. Les convergences ps et en probabilité se généralisent pour des variables à valeurs dans un espace métrique quelconque.

3.2 Convergence en norme

Définition 114. Pour $(X_n)_n$ une suite de variables aléatoires réelles et $1 \leq p < \infty$, on dit que X_n converge vers X en norme p si

$$\|X_n - X\|_p = \mathbb{E}(|X_n - X|^p)^{1/p} \xrightarrow[n\to\infty]{} 0.$$

On note alors $X_n \xrightarrow{\mathrm{L}^p} X$.

Remarque 115. Il s'agit simplement de la convergence pour la distance de l'espace métrique complet $(\mathrm{L}^p(\Omega, \mathcal{A}, \mathbb{P}), \|\cdot\|_p)$.

Proposition 116. Si $p \geq q$, la convergence L^p entraîne la convergence L^q.

Démonstration. C'est la même preuve que pour les fonctions. □

Proposition 117. La convergence L^p entraîne la convergence en probabilité.

Démonstration. Cela découle de l'inégalité de Markov :

$$\mathbb{P}(|X_n - X| \geq \varepsilon) \leq \frac{\|X_n - X\|_p^p}{\varepsilon^p}.$$

□

Remarque 118. La convergence en norme n'est pas préservée par une application continue. Soit par exemple (X_n) une suite de variables indépendantes telle que X_n vaut 1 avec probabilité $1 - 1/n^2$ et n avec probabilité $1/n^2$ et $g : x \mapsto x^2$. On peut montrer que X_n tend vers 1 dans L^2 mais que ce n'est pas le cas pour la suite des carrés. Ainsi, (X_n^2) donne un exemple de suite qui converge en probabilité mais pas dans L^2. Pour passer d'une convergence ps ou en probabilité à une convergence en norme, on peut utiliser le théorème de convergence dominée.

3.3 Convergence en loi

Par souci de simplicité, on énoncera les définitions pour des variables aléatoires réelles, mais quand c'est possible cela s'étend aux variables aléatoires dans $\mathbb{R}^d$.
On désigne par $\mathrm{C_b}(\mathbb{R})$ l'ensemble des fonctions continues bornées de $\mathbb{R}$.

Définition 119. Une suite (X_n) de variables aléatoires réelles *converge en loi* vers une variable aléatoire X (noté $X_n \xrightarrow{\text{loi}} X$) si pour tout $\varphi \in \mathrm{C_b}(\mathbb{R})$,

$$\mathbb{E}(\varphi(X_n)) \xrightarrow[n\to\infty]{} \mathbb{E}(\varphi(X)).$$

Remarque 120. La convergence en loi est en fait une convergence de mesures. Les variables aléatoires n'ont pas besoin d'être définies sur le même espace de probabilité. La limite n'est pas unique, mais c'est la loi limite $\mathbb{P}_X$ qui est unique (cf. Théorème 71).

Exemple 121. Si X_n possède une densité f_n par rapport à la mesure de Lebesgue. On suppose que (f_n) converge simplement vers une fonction f avec une hypothèse de domination uniforme par une fonction intégrable sur les (f_n). Alors par convergence dominée, les (X_n) convergent en loi vers une variable ayant pour densité f. Il existe un résultat, appelé *lemme de Scheffé*, qui permet de se débarrasser de l'hypothèse de domination.

Avant d'énoncer le lemme de portemanteau, on rappelle que si (x_n) est une suite réelle, sa limite inférieure, notée $\liminf_{n\to} x_n \in [-\infty,+\infty]$, est sa plus petite valeur d'adhérence. Sa limite supérieure, notée $\limsup_{n\to} x_n \in [-\infty,+\infty]$, est sa plus grande valeur d'adhérence. Une suite converge si et seulement si sa $\limsup$ et sa $\liminf$ sont égales.

Proposition 122 (lemme portemanteau)**.** Soient (X_n) des variables aléatoires de lois respectives (μ_n) et X de loi μ. On a équivalence entre :

1. X_n converge en loi vers X.
2. Pour tout ouvert G de $\mathbb{R}$, $\liminf_{n\to\infty} \mu_n(G) \geq \mu(G)$
3. Pour tout fermé F de $\mathbb{R}$, $\limsup_{n\to\infty} \mu_n(F) \leq \mu(F)$
4. Pour tout borélien B tel que $\mu(\partial B) = 0$, $\lim_{n\to\infty} \mu_n(B) = \mu(B)$.

Démonstration. $(1 \Rightarrow 2)$. Soit G un ouvert de $\mathbb{R}$, et (φ_p) une suite de fonctions continues bornées qui approche $\mathbf{1}_G$ en croissant (on la construit en étudiant la structure des ouverts de $\mathbb{R}$). Alors

$$\liminf_{n\to\infty} \mu_n(G) \geq \sup_p \left(\liminf_{n\to\infty} \int \varphi_p \,\mathrm{d}\mu_n \right) = \sup_p \int \varphi_p \,\mathrm{d}\mu = \mu(G).$$

$(2 \Rightarrow 3)$. Par passage au complémentaire.
$(2{,}3 \Rightarrow 4)$. Soit B un borélien de $\mathbb{R}$ tel que $\mu(\partial B) = 0$:

$$\limsup_{n\to\infty} \mu_n(B) \leq \limsup_{n\to\infty} \mu_n(\overline{B}) \leq \mu(\overline{B}).$$

De même $\mu(\mathring{B}) \leq \liminf_{n\to\infty} \mu_n(B)$. Puisque $\mu(\partial B) = 0$, on a $\mu(B) = \mu(\mathring{B}) = \mu(\overline{B})$.
$(4 \Rightarrow 1)$. Soit φ continue bornée, on suppose φ positive quitte à faire $\varphi = \varphi_+ - \varphi_-$. Soit K tel que $0 \leq \varphi \leq K$. Par Fubini, on a

$$\begin{aligned}\int \varphi(x) \,\mathrm{d}\mu(x) &= \int_{\mathbb{R}} \int_0^K \mathbf{1}_{t \leq \varphi(x)} \,\mathrm{d}t\mathrm{d}\mu(x) \\ &= \int_0^K \mu(\{\varphi \geq t\}) \,\mathrm{d}t.\end{aligned}$$

Et de même pour tout n :

$$\int \varphi(x)\, \mathrm{d}\mu_n(x) = \int_0^K \mu_n(\{\varphi \geq t\})\, \mathrm{d}t.$$

On remarque alors que $\{\varphi \geq t\}$ est un borélien, de frontière $\{\varphi = t\}$, qui n'est de mesure strictement positive que sur un ensemble de mesure de Lebesgue nulle. On a donc

$$\mu_n(\{\varphi \geq t\}) \xrightarrow[n\to\infty]{} \mu(\{\varphi \geq t\})$$

pour λ-presque tout t. On a donc le résultat par convergence dominée. $\square$

Corollaire 123. Soit (X_n) une suite de variables aléatoires réelles et X une variable aléatoire réelle. X_n converge en loi vers X si et seulement si $F_{X_n}(x)$ converge vers $F_X(x)$ en tout point x où F_X est continue.

Démonstration. Le sens direct vient de 4. du lemme portemanteau. Réciproquement, pour $x \in \mathbb{R}$:

$$\liminf_{n\to\infty} F_{X_n}(x^-) \geq F_X(x^-) \text{ et } \limsup_{n\to\infty} F_{X_n}(x) \leq F_X(x),$$

donc le point 2. du lemme portemanteau est satisfait pour un intervalle ouvert. Mais un ouvert de $\mathbb{R}$ est une réunion dénombrable d'intervalles ouverts disjoints de $\mathbb{R}$, ce qui permet de conclure. $\square$

Proposition 124. La convergence en probabilité entraîne la convergence en loi.

Démonstration. Soit (X_n) une suite de variables aléatoires réelles qui converge vers une variable aléatoire X. Soit $a \in \mathbb{R}$, pour $\varepsilon > 0$, on a :

$$\mathbb{P}(X \leq a - \varepsilon) - \mathbb{P}(|X_n - X| > \varepsilon) \leq \mathbb{P}(X_n \leq a) \leq \mathbb{P}(X \leq a + \varepsilon) + \mathbb{P}(|X_n - X| > \varepsilon).$$

Soit a un point où F_X est continue et $\varepsilon' > 0$. On peut calibrer $\varepsilon > 0$ tel que

$$|\,\mathbb{P}(X \leq a + \varepsilon) - \mathbb{P}(X \leq a)|, |\,\mathbb{P}(X \leq a - \varepsilon) - \mathbb{P}(X \leq a)| < \varepsilon'.$$

Puisque X_n converge en probabilité vers X, il existe N tel que

$$\forall n \geq N, \mathbb{P}(|X_n - X| > \varepsilon) \leq \varepsilon'.$$

Ainsi, pour $n \geq N$,

$$|\,\mathbb{P}(X_n \leq a) - \mathbb{P}(X \leq a)| < 2\varepsilon'.$$

On a donc convergence des fonctions de répartition en tout point de continuité de la fonction de répartition limite. Le résultat précédent nous permet donc de conclure. $\square$

Proposition 125. Si (X_n) est une suite de variables aléatoires réelles (définies sur le même espace de probabilité) qui converge en loi vers une constante $a \in \mathbb{R}$, alors elle converge en probabilité vers cette constante.

Démonstration. Par le point 2. du lemme portemanteau (Proposition 122) :

$$\liminf_{n\to\infty} \mathbb{P}_{X_n}(]a-\varepsilon, a+\varepsilon[) \geq \mathbb{P}(a \in]a-\varepsilon, a+\varepsilon[) = 1 \qquad \forall \varepsilon > 0,$$

ce qui veut dire que X_n converge en probabilité vers X. $\square$

Proposition 126 (continuous mapping theorem)**.** Si (X_n) est une suite de variables aléatoires réelles (ou dans $\mathbb{R}^d$) qui converge en loi vers X, alors pour toute fonction continue $g : \mathbb{R} \to \mathbb{R}$, $(g(X_n))$ converge en loi vers $g(X)$.

Démonstration. Soit h une fonction continue bornée, alors $h \circ g$ est une fonction continue bornée donc $\mathbb{E}(h(g(X_n)))$ converge vers $\mathbb{E}(h(g(X)))$, ce qui signifie que $g(X_n)$ converge en loi vers $g(X)$, puisque ceci vaut pour tout h. □

Proposition 127. Soient des variables aléatoires (X_n) de lois respectives (μ_n) et X de loi μ. Soit $H \subset \mathrm{C_b}(\mathbb{R})$ tel que $\overline{H} \supset \mathrm{C_c}(\mathbb{R})$ (fonctions continues à support compact). Alors on a équivalence entre :

1. $X_n \xrightarrow{\text{loi}} X$
2. $\forall \varphi \in \mathrm{C_c}(\mathbb{R}), \int \varphi \,\mathrm{d}\mu_n \xrightarrow[n\to\infty]{} \int \varphi \,\mathrm{d}\mu$
3. $\forall \varphi \in H, \int \varphi \,\mathrm{d}\mu_n \xrightarrow[n\to\infty]{} \int \varphi \,\mathrm{d}\mu$.

Démonstration. Les implications $(1 \Rightarrow 2)$ et $(1 \Rightarrow 3)$ ont déjà été vues (elles découlent de la définition). Supposons que 2. est satisfaite. Soient $0 \leq f_k \leq 1$ des fonctions continues à support compact qui convergent en croissant vers 1. Pour φ continue bornée (positive, quitte à décomposer en partie négative/positive), on a alors

$$\forall k \in \mathbb{N}, \int \varphi f_k \,\mathrm{d}\mu_n \xrightarrow[n\to\infty]{} \int \varphi f_k \,\mathrm{d}\mu.$$

On a donc, pour tout k :

$$\limsup_{n\to\infty} \left| \int \varphi \,\mathrm{d}\mu_n - \int \varphi \,\mathrm{d}\mu \right| \leq 2\|\varphi\|_\infty \left(1 - \int f_k \,\mathrm{d}\mu\right).$$

On prend alors $k \to \infty$ pour avoir $(2 \Rightarrow 1)$.
Maintenant, traitons $(3 \Rightarrow 2)$: soit $\varphi_\varepsilon \in H$ telle que $\|\varphi - \varphi_\varepsilon\| \leq \varepsilon$ (où $\varphi \in \mathrm{C_c}(\mathbb{R})$). Alors on a, en intercalant $\int \varphi_\varepsilon \,\mathrm{d}\mu_n$ et $\int \varphi_\varepsilon \,\mathrm{d}\mu$:

$$\limsup_{n\to\infty} \left| \int \varphi \,\mathrm{d}\mu_n - \int \varphi \,\mathrm{d}\mu \right| \leq 2\varepsilon,$$

ce qui permet de conclure. □

Remarque 128. On peut utiliser ce résultat pour H l'ensemble des fonctions C^∞ à support compact (noté $\mathrm{C}_\mathrm{c}^\infty(\mathbb{R})$) ou bien lipschitziennes à support compact, lipschitziennes bornées.

Maintenant, un théorème très pratique lorsqu'il s'agit de montrer des convergences en loi particulières (pour des lois usuelles par exemple).

Théorème 129 (de Lévy)**.** Soient $(X_n)_n$ et X des variables aléatoires réelles. Alors X_n converge en loi vers X si et seulement si la suite de fonctions (φ_{X_n}) converge simplement vers φ_X.

Démonstration. Le sens direct découle de la définition de la convergence en loi, puisque les exponentielles complexes sont continues bornées. Supposons donc que $\varphi_n(\xi) := \varphi_{X_n}(\xi) \xrightarrow[n\to\infty]{} \varphi_X(\xi) =: \varphi(\xi)$ pour tout $\xi \in \mathbb{R}$. Soit $f \in \mathrm{C_c}(\mathbb{R})$ et

$$g_\sigma(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \mathrm{e}^{-x^2/2\sigma^2}.$$

C'est un résultat classique que de montrer que $g_\sigma * f$ converge simplement et même uniformément vers f lorsque σ tend vers 0. De plus, on peut montrer que pour ν une mesure de probabilité sur $\mathbb{R}$:

$$\int g_\sigma * f \,\mathrm{d}\mu = \int_\mathbb{R} f(x) \left[\frac{1}{\sqrt{2\pi\sigma^2}} \int_\mathbb{R} \mathrm{e}^{\mathrm{i}\xi x} g_{1/\sigma}(\xi) \widehat{\nu}(-\xi) \,\mathrm{d}\xi \right] \mathrm{d}x,$$

où $\widehat{\nu}(\xi) = \int_{\mathbb{R}} \mathrm{e}^{\mathrm{i}t\xi}\,\mathrm{d}\nu(t)$. Par convergence dominée, on sait que (pour $\mu_n = \mathbb{P}_{X_n}$ et $\mu = \mathbb{P}_X$) :

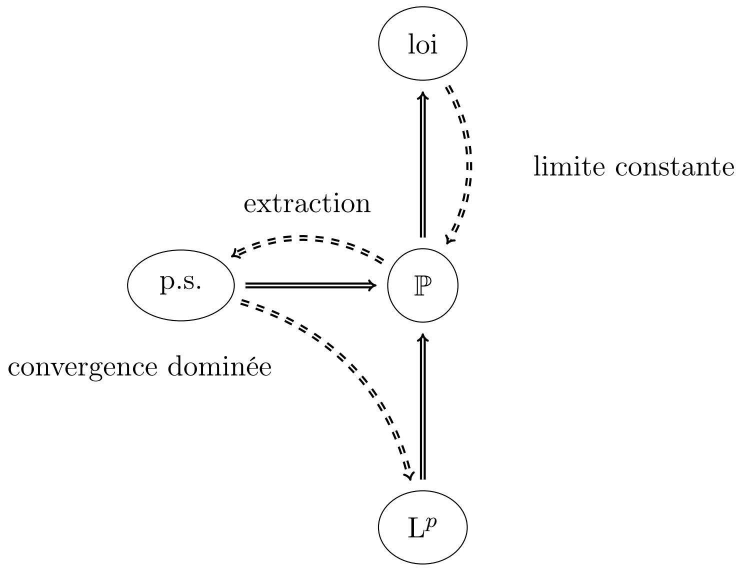
$$\int_{\mathbb{R}} \mathrm{e}^{\mathrm{i}\xi x} g_{1/\sigma}(\xi)\widehat{\mu}_n(-\xi)\,\mathrm{d}\xi \xrightarrow[n\to\infty]{} \int_{\mathbb{R}} \mathrm{e}^{\mathrm{i}\xi x} g_{1/\sigma}(\xi)\widehat{\mu}(-\xi)\,\mathrm{d}\xi.$$

En utilisant une deuxième fois le théorème de convergence dominée, on montre alors que

$$\int_{\mathbb{R}} g_\sigma * f\,\mathrm{d}\mu_n \xrightarrow[x\to\infty]{} \int_{\mathbb{R}} g_\sigma * f\,\mathrm{d}\mu.$$

Pour $H = \{g_\sigma * f, \sigma > 0, f \in \mathrm{C_c}(\mathbb{R})\}$, on peut appliquer la Proposition 127 qui permet de conclure. $\square$

Graphe des différents modes de convergence :



3.4 Théorèmes limite et statistiques

La loi faible des grands nombres nous donne une convergence dite *faible* (c'est-à-dire en probabilité) de moyennes empiriques $\overline{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$ lorsque les $(X_i)_i$ sont i.i.d et L^2. On va chercher à donner une version plus forte de ce résultat, c'est-à-dire enlever l'hypothèse L^2 et passer à une convergence presque sûre.

Théorème 130 (loi forte des grands nombres, LGN)**.** Soient $(X_n)_n$ des variables aléatoires indépendantes et de même loi avec $\mathbb{E}(|X_1|) < \infty$. Alors

$$\frac{X_n}{n} \xrightarrow[n\to\infty]{\text{ps}} \mathbb{E}(X_1).$$

Démonstration. Quitte à remplacer X_1 par $-X_1$, on peut se ramener à montrer que pour tout $\varepsilon > 0$,

$$\limsup_{n\to\infty} \frac{1}{n}\sum_{i=1}^n (X_i - \mathbb{E}(X_i) - \varepsilon) \leq 0 \text{ ps.}$$

On pose alors $Y_i = X_i - \mathbb{E}(X_i) - \varepsilon$ (on a $\mathbb{E}(Y_i) < 0$) et

$$G_n = \max\left(Y_0, Y_0 + Y_1, \dots, \sum_{i=0}^{n-1} Y_i\right) \text{ et } H_n = \max\left(Y_1 + \dots, \sum_{i=1}^{n} Y_i\right),$$

qui sont deux suites croissantes de variables aléatoires réelles. On a aussi, pour tout $n \in \mathbb{N}$:

$$G_{n+1} = Y_0 + \max(0, H_n) \text{ soit } G_{n+1} - H_n = Y_0 - \min(0, H_n).$$

Soit A l'événement

$$A = \left\{\sup_n G_n = +\infty\right\} = \left\{\sup_n H_n = +\infty\right\}.$$

On voit alors que l'on a

$$(G_{n+1} - H_n)\mathbf{1}_A \xrightarrow[n\to\infty]{} Y_0\mathbf{1}_A$$

et à partir d'un certain rang :

$$|(G_{n+1} - H_n)\mathbf{1}_A| \leq |Y_0| \in \mathrm{L}^1.$$

On peut donc appliquer le théorème de convergence dominée, qui donne

$$\mathbb{E}((G_{n+1} - H_n)\mathbf{1}_A) \xrightarrow[n\to\infty]{} \mathbb{E}(Y_0\mathbf{1}_1) = \mathbb{P}(A)\,\mathbb{E}(Y_0) \leq 0,$$

puisque Y_0 est indépendante de A. Par ailleurs, $(G_n, \mathbf{1}_A)$ a même loi que $(H_n, \mathbf{1}_A)$, donc

$$\mathbb{E}((G_{n+1} - H_n)\mathbf{1}_A) = \mathbb{E}((G_{n+1} - G_n)\mathbf{1}_A) \geq 0.$$

Nécessairement, $\mathbb{P}(A) = 0$ soit $\sup_n G_n < +\infty$ ps. On a donc

$$\frac{1}{n}\sum_{i=1}^{n} Y_i \leq \frac{G_n}{n} \xrightarrow[n\to\infty]{} 0$$

presque sûrement, ce qui permet de conclure. □

Remarque 131. Ce théorème permet de répondre à une interrogation de départ : peut-on retrouver les probabilités d'un événement en regardant la fréquence de réalisation de l'événement sur un grand nombre d'expériences ? La réponse est oui. Par exemple, pour vérifier qu'un dé est équilibré, on peut le lancer un grand nombre de fois N, et compter le nombre de fois où on tombe sur 1, on calcule alors

$$\widehat{p}_1 = \frac{1}{N}\sum_{i=1}^{N} \mathbf{1}_{X_i=1}.$$

Par la loi des grands nombres, on est sûrs que $\widehat{p}_1$ approche $\mathbb{E}(\mathbf{1}_{X_1=1}) = \mathbb{P}(X_1 = 1) = 1/6$ quand N devient grand. Ce théorème est à la base de ce qu'on appelle les *statistiques*, qui consistent à modéliser les expériences aléatoires avec des probabilités, et de retrouver ces probabilités avec des méthodes d'estimation.

Exemple 132. On suppose que la taille d'un individu pris au hasard suit une loi P avec un moment d'ordre 1. Pour estimer l'espérance inconnue de cette loi, notée m, à partir d'un échantillon de la population $(X_1, \ldots, X_n)$ i.i.d. de loi P, la chose que l'on peut faire est calculer la moyenne sur l'échantillon (moyenne empirique) :

$$\overline{X}_n = \frac{1}{n}\sum_{k=1}^{n} X_k,$$

qui va approcher m quand n tend vers l'infini. Une nouvelle question se pose alors ; combien de personnes dois-je mesurer pour avoir accès à m à une précision σ donnée ? La théorème qui fournit la réponse à cette question est le théorème limite central, que voici.

Théorème 133 (théorème limite central (TLC), théorème central limite (TCL))**.** Soient $(X_n)_n$ des variables aléatoires i.i.d avec $\mathbb{E}(X_1) = m$, $\operatorname{Var}(X_1) = \sigma^2$. Alors

$$\sqrt{n}(\overline{X}_n - m) \xrightarrow[n\to\infty]{\text{loi}} \mathcal{N}(0, \sigma^2),$$

ou de manière équivalente :

$$\sqrt{n}\left(\frac{\overline{X}_n - m}{\sigma}\right) \xrightarrow[n\to\infty]{\text{loi}} \mathcal{N}(0, 1).$$

Démonstration. Quitte à centrer et réduire (retrancher m et diviser par σ), on se ramène à $m = 0, \sigma = 1$. Soit $t \in \mathbb{R}$ et $n \in \mathbb{N}^*$:

$$\begin{aligned}
\varphi_{\sqrt{n}\overline{X}_n}(t) &= \mathbb{E}(\mathrm{e}^{\mathrm{i}\sqrt{n}t\overline{X}_n}) \\
&= \prod_{j=1}^{n} \varphi_{X_j}\left(\frac{t}{\sqrt{n}}\right) \quad \text{par indépendance} \\
&= \varphi_{X_1}\left(\frac{t}{\sqrt{n}}\right)^n \quad \text{par identique distribution.}
\end{aligned}$$

Par le théorème de dérivation sous l'intégrale, φ est de classe C^2 et on a donc par la formule de Taylor-Young, pour n qui tend vers l'infini :

$$\begin{aligned}\varphi_{\sqrt{n}\overline{X}_n}(t) &= \left(1 - \frac{t^2}{2n} + \mathrm{o}(t^2/n)\right)^n \\ &= \mathrm{e}^{n\log\left(1-\frac{t^2}{2n}+\mathrm{o}(t^2/n)\right)},\end{aligned}$$

où le log est un logarithme complexe, bien défini pour n assez grand. On a alors, en faisant un développement limité du logarithme :

$$\varphi_{\sqrt{n}\overline{X}_n}(t) \xrightarrow[n\to\infty]{} \mathrm{e}^{-t^2/2},$$

qui n'est autre que la fonction caractéristique de la loi normale centrée réduite $\mathcal{N}(0,1)$. On conclut donc avec le théorème de Lévy. $\square$

Ce théorème nous dit donc que des sommes de variables indépendantes bien renormalisées ont toujours la même loi à la limite, ce qui s'observe en pratique. Cela permet de construire en statistiques ce que l'on appelle des intervalles de confiance (asymptotiques). Sous les hypothèses du théorème limite central, on a, puisque la convergence en loi entraîne la convergence des fonctions de répartition :

$$\mathbb{P}\left(\left|\sqrt{n}\left(\frac{\overline{X}_n - m}{\sigma}\right)\right| > t_\alpha\right) \xrightarrow[n\to\infty]{} \mathbb{P}(|\mathcal{N}(0,1)| > t_\alpha).$$

Si on calibre t_α de telle sorte que $\mathbb{P}(|\mathcal{N}(0,1)| > t_\alpha) = \alpha$, on a alors :

$$\mathbb{P}\left(m \in \left[\overline{X}_n - \frac{\sigma t_\alpha}{\sqrt{n}}, \overline{X}_n + \frac{\sigma t_\alpha}{\sqrt{n}}\right]\right) \xrightarrow[n\to\infty]{} 1-\alpha.$$

Si on connaît σ, on sait comment localiser m avec grande probabilité. Plus notre taille d'échantillon est grande, plus on est précis dans cette localisation. La vitesse de convergence vers la moyenne dans la loi des grands nombres est en $\frac{1}{\sqrt{n}}$. Si on veut être deux fois plus précis, il faudra un échantillon quatre fois plus grand.

*   *
*