Probabilités et Statistique

Benoîte de Saporta & Benjamin Charlier

24 février 2025

Table des matières

Chapitre 2

Variables Aléatoires

Dans cette section, on introduit la notion de variable aléatoire. Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace probabilisé et $(E, \mathcal{E})$ un espace probabilisable.

2.1 Définition

Définition 2.1.1. *Une* ***variable aléatoire*** *X sur $(\Omega, \mathcal{F}, \mathbb{P})$ à valeurs dans $(E, \mathcal{E})$ est une application mesurable de $(\Omega, \mathcal{F})$ dans $(E, \mathcal{E})$ telle que*

$$\forall A \in \mathcal{E}, \quad X^{-1}(A) \in \mathcal{F}.$$

On note alors l'événement "X appartient à A" par

$$X^{-1}(A) = \{\omega \in \Omega; X(\omega) \in A\} = (X \in A).$$

Dans toute la suite de ce cours, on ne considèrera que les deux cas suivants

1. $(E, \mathcal{E}) = (\mathbb{R}, B(\mathbb{R}))$, on parle alors de **variable aléatoire réelle**,
2. $(E, \mathcal{E}) = (\mathbb{R}^d, B(\mathbb{R}^d))$, on parle alors de **vecteur aléatoire**.

D'autres cas sont cependant possibles, comme des variables aléatoires à valeurs dans des espaces de fonctions (on parle alors de processus stochastique) ou même des variables aléatoires à valeurs dans des ensembles de mesures (on parle alors de mesures aléatoires). Vous verrez des exemples en M1 pour celles et ceux qui choisiront un cursus proba-stat.

Exemple 2.1.1. On lance deux fois une pièce. On modélise cette expérience par $\Omega = \{P, F\}^2$ muni de la tribu de ses parties et de la loi uniforme. On considère les applications X_1 et X_2 de Ω dans $\mathbb{R}$ définies par

- $X_1(\omega_1, \omega_2) = \mathbb{1}_P(\omega_1)$ le nombre de P au premier lancer,
- $X_2(\omega_1, \omega_2) = \mathbb{1}_P(\omega_2)$ le nombre de P au deuxième lancer.

Alors X_1 et X_2 sont des variables aléatoires réelles à valeurs dans $\{0, 1\}$.

Exemple 2.1.2. Soit $\Omega =]0, 1]$ muni de la tribu des boréliens et de la mesure de Lebesgue λ. On définit l'application X de Ω dans $\mathbb{R}$ par $X(\omega) = -\log(\omega)$. Cette application est continue donc mesurable. Ainsi X est une variable aléatoire.

Proposition 2.1.1. *Soit X sur $(\Omega, \mathcal{F}, \mathbb{P})$ à valeurs dans $(E, \mathcal{E})$ et h une fonction mesurable de $(E, \mathcal{E})$ dans $(G, \mathcal{G})$. Alors $h(X)$ est une variable aléatoire sur $(\Omega, \mathcal{F}, \mathbb{P})$ à valeurs dans $(G, \mathcal{G})$.*

Démonstration. Une composée de fonctions mesurables est mesurable. □

Exemple 2.1.3. On reprend l'exemple 2.1.1 du lancer de pièce. Soit les applications h_1 et h_2 de $(E,\mathcal{E}) = (\{0,1\},\mathcal{P}(\{0,1\}))$ dans $(G,\mathcal{G}) = (\{0,1\}^2,\mathcal{P}(\{0,1\}^2))$, et h_3 de $(G,\mathcal{G})$ dans $(G,\mathcal{G})$ définies par

$$\begin{aligned} h_1(x_1) &= (x_1, 0), \\ h_2(x_2) &= (0, x_2), \\ h_3(x,y) &= x + y. \end{aligned}$$

Ces trois applications sont mesurables. Soit $Y = (X_1, X_2)$ le vecteur nombre de P au premier lancer, nombre de P au deuxième lancer. Comme $Y = h_3(h_1(X_1), h_2(X_2))$, Y est une variable aléatoire à valeurs dans $(G,\mathcal{G})$.
Soit h_4 de $(G,\mathcal{G}) = (\{0,1\}^2,\mathcal{P}(\{0,1\}^2)$ dans $(H,\mathcal{H}) = (\{0,1,2\},\mathcal{P}(\{0,1,2\})$ définie par

$$h_4(x,y) = x + y.$$

Alors $Z = X_1 + X_2 = h_4(X_1, X_2)$ est une variable aléatoire réelle.

Plus généralement, le vecteur $X = (X_1, \cdots, X_d)$ est un vecteur aléatoire si et seulement si chacune de ses coordonnées est une variable aléatoire (les applications coordonnées sont mesurables).

Définition 2.1.2. *Soit X une variable aléatoire sur $(\Omega,\mathcal{F},\mathbb{P})$ à valeurs dans $(E,\mathcal{E})$. La loi de X, notée $\mathbb{P}_X$ est la mesure image de $\mathbb{P}$ par X*

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

Proposition 2.1.2. *Soit X une variable aléatoire sur $(\Omega,\mathcal{F},\mathbb{P})$ à valeurs dans $(E,\mathcal{E})$. La loi de X est une loi de probabilité sur $(E,\mathcal{E})$.*

Démonstration. Pour tout $A \in \mathcal{E}$, par définition $\mathbb{P}_X(A) = \mathbb{P}(X \in A) \geq 0$ donc on a bien une application de $\mathcal{E}$ dans $\mathbb{R}_+$. De plus, $\mathbb{P}_X(E) = \mathbb{P}(X \in E) = \mathbb{P}(\Omega) = 1$ et si $(A_n)_{n\in\mathbb{N}}$ est une suite d'événements deux à deux disjoints de $\mathcal{E}$, alors $(X \in A_n)_{n\in\mathbb{N}}$ est une suite d'événements deux à deux disjoints de $\mathcal{F}$, donc

$$\mathbb{P}_X\Big(\bigcup_{n\in\mathbb{N}} A_n\Big) = \mathbb{P}\Big(X \in \bigcup_{n\in\mathbb{N}} A_n\Big) = \mathbb{P}\Big(\bigcup_{n\in\mathbb{N}} (X \in A_n)\Big) = \sum_{n\in\mathbb{N}} \mathbb{P}(X \in A_n) = \sum_{n\in\mathbb{N}} \mathbb{P}_X(A_n),$$

on a bien la propriété de σ-additivité, donc $\mathbb{P}_X$ est bien une probabilité sur $(E,\mathcal{E})$. $\square$

En pratique, on confond souvent une variable aléatoire et sa loi. Si $X = (X_1, \cdots, X_d)$ est un vecteur aléatoire, on appelle également la loi de X la **loi jointe** du vecteur et les lois des coordonnées X_i (ou de sous-vecteurs) les **lois marginales**.

Exemple 2.1.4. Dans l'exemple des pièces, X_1 et X_2 sont à valeurs dans $(E,\mathcal{E}) = (\{0,1\},\mathcal{P}(\{0,1\})$. On peut donc complètement décrire leur loi

$A \in \mathcal{P}(\{0,1\})$	$A = \emptyset$	$A = \{0\}$	$A = \{1\}$	$A = E$
$\mathbb{P}(X_1 \in A)$	$\mathbb{P}(\emptyset) = 0$	$\mathbb{P}(\{FP, FF\}) = \frac{1}{2}$	$\mathbb{P}(\{PP, PF\}) = \frac{1}{2}$	$\mathbb{P}(\Omega) = 1$
$\mathbb{P}(X_2 \in A)$	$\mathbb{P}(\emptyset) = 0$	$\mathbb{P}(\{PF, FF\}) = \frac{1}{2}$	$\mathbb{P}(\{PP, FP\}) = \frac{1}{2}$	$\mathbb{P}(\Omega) = 1$

$\mathbb{P}_{X_1}$ et $\mathbb{P}_{X_2}$ sont donc toutes les deux la loi uniforme sur $(\{0,1\},\mathcal{P}(\{0,1\})$. Remarquons que les lois de X_1 et X_2 sont égales mais $X_1 \neq X_2$.
La variable aléatoire Y est à valeurs dans $(G,\mathcal{G}) = (\{0,1\}^2,\mathcal{P}(\{0,1\}^2))$. On a

$A \subset E$	$A = \{(0,0)\}$	$A = \{(0,1)\}$	$A = \{(1,0)\}$	$A = \{(1,1)\}$
$\mathbb{P}(Y = A)$	$\mathbb{P}(\{FF\}) = \frac{1}{4}$	$\mathbb{P}(\{FP\}) = \frac{1}{4}$	$\mathbb{P}(\{PF\}) = \frac{1}{4}$	$\mathbb{P}(\{PP\}) = \frac{1}{4}$

Il s'agit encore de la loi uniforme sur E.
Pour Z, elle est à valeurs dans $(H, \mathcal{H}) = (\{0,1,2\}, \mathcal{P}(\{0,1,2\}))$

$A \subset E$	$A = \{0\}$	$A = \{1\}$	$A = \{2\}$
$\mathbb{P}(Z = A)$	$\mathbb{P}(\{FF\}) = \frac{1}{4}$	$\mathbb{P}(\{PF, FP\}) = \frac{1}{2}$	$\mathbb{P}(\{PP\}) = \frac{1}{4}$

On a donc complètement décrit la loi de Z.

Exemple 2.1.5. Soit $\Omega =]0,1]$ muni de la tribu des boréliens et de la mesure de Lebesgue λ et X la variable aléatoire $X(\omega) = -\log(\omega)$. Elle est à valeurs dans $\mathbb{R}_+$. Soit A un borélien de $\mathbb{R}_+$, on a

$$\begin{aligned}
\mathbb{P}(X \in A) &= \mathbb{P}(\{\omega \in]0,1]; -\log(\omega) \in A\}) \\
&= \lambda\{\omega \in]0,1]; -\log(\omega) \in A\} \\
&= \int_{\{\omega \in]0,1]; -\log(\omega) \in A\}} d\omega \\
&= \int_A e^{-u} du,
\end{aligned}$$

en faisant le changement de variable $u = -\log(\omega)$.

Définition 2.1.3. *Soit X une variable aléatoire sur $(\Omega, \mathcal{F}, \mathbb{P})$ à valeurs dans $(E, \mathcal{E})$.*

- *On dit que la loi de X est discrète si c'est une combinaison linéaire finie ou dénombrable de masses de Dirac. On dit alors que X est une* ***variable aléatoire discrète****.*
- *On dit que la loi de X est absolument continue si elle admet une densité par rapport à la mesure de Lebesgue sur E. On dit alors que X est une* ***variable aléatoire absolument continue****, ou une* ***variable aléatoire à densité****.*

Exemple 2.1.6. Les variables aléatoires X_1, X_2, Y et Z de l'exemple de pile ou face sont toutes des variables aléatoires discrètes. On peut en effet écrire leur loi comme

$$\begin{aligned}
\mathbb{P}_{X_1} = \mathbb{P}_{X_2} &= \frac{1}{2}\delta_0 + \frac{1}{2}\delta_1, \\
\mathbb{P}_Y &= \frac{1}{4}\delta_{(0,0)} + \frac{1}{4}\delta_{(0,1)} + \frac{1}{4}\delta_{(1,0)} + \frac{1}{4}\delta_{(1,1)}, \\
\mathbb{P}_Z &= \frac{1}{4}\delta_0 + \frac{1}{2}\delta_1 + \frac{1}{4}\delta_2.
\end{aligned}$$

Exemple 2.1.7. Soit $\Omega =]0,1]$ muni de la tribu des boréliens et de la mesure de Lebesgue λ et X la variable aléatoire $X(\omega) = -\log(\omega)$ de l'exemple 2.1.5. On a vu que

$$\mathbb{P}(X \in A) = \int_A e^{-u} du,$$

pour tout A borélien de $\mathbb{R}_+$. On en déduit que la loi de X a pour densité $e^{-u}\mathbb{1}_{u \geq 0}$ par rapport à la mesure de Lebesgue. X est donc une variable aléatoire absolument continue.

Contre-Exemple 2.1.1. Toujours avec les notations de l'Exemple 2.1.5, on pose $Y = \min\{X, 1\} = X \wedge 1$. On note de plus $\{1 \in A\} = \{\omega \in \Omega; 1 \in A\} = \begin{cases} \Omega, & \text{si } 1 \in A \\ \emptyset, & \text{si } 1 \notin A \end{cases}$. Calculons la loi de Y. On a

$$\begin{aligned}\mathbb{P}(Y \in A) &= \mathbb{P}(X \wedge 1 \in A) \\ &= \mathbb{P}\big(\{X \wedge 1 = 1\} \cap \{1 \in A\} \cup \{X \wedge 1 = X\} \cap \{X \in A\}\big) \\ &= \mathbb{P}(\{X \geq 1\} \cap \{1 \in A\}) + \mathbb{P}(\{X \leq 1\} \cap \{X \in A\})\end{aligned}$$

car $\mathbb{P}(X = 1) = 0$. On a donc

$$\begin{aligned}\mathbb{P}(Y \in A) &= \mathbb{1}_A(1) \int_{[1,+\infty[} e^{-u} du + \int_{A \cap [0,1[} e^{-u} du \\ &= \mathbb{1}_A(1) e^{-1} + \int_{A \cap [0,1[} e^{-u} du.\end{aligned}$$

On a en particulier

$$\mathbb{P}(Y = 1) = \mathbb{P}(X \geq 1) = e^{-1} > 0.$$

La variable aléatoire Y a donc une partie densité sur $]0, 1[$ et une partie discrète en 1. Ce n'est ni une variable aléatoire absolument continue ni une variable aléatoire discrète.

2.2 Exemples de variables aléatoires, lois usuelles

On donne maintenant une liste des lois les plus courantes ainsi que quelques unes de leurs propriétés de base.

Loi de Bernoulli $\mathcal{B}(p)$ Il s'agit de la loi correspondant à un tirage avec une issue favorable et une issue défavorable, le paramètre p correspond à la probabilité de réussite (issue favorable), $X(\Omega) = \{0, 1\}$,

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

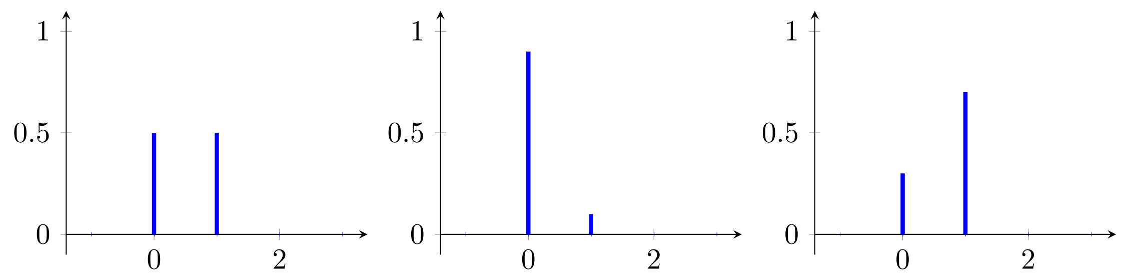

FIGURE 2.1 – Schéma de Bernoulli de paramètre $\mathcal{B}(0.5)$, $\mathcal{B}(0.1)$ et $\mathcal{B}(0.7)$

Loi Uniforme discrète $\mathcal{U}(E)$ sur un ensemble fini E de cardinal n, $\sum_{i \in E} \delta_i / n$. C'est le résultat d'un tirage avec équi-probabilité, par exemple si $n = 6$ le lancer d'un dé non truqué. Si $A \subset E$,

$$\mathbb{P}(X \in A) = \frac{|A|}{|E|}.$$

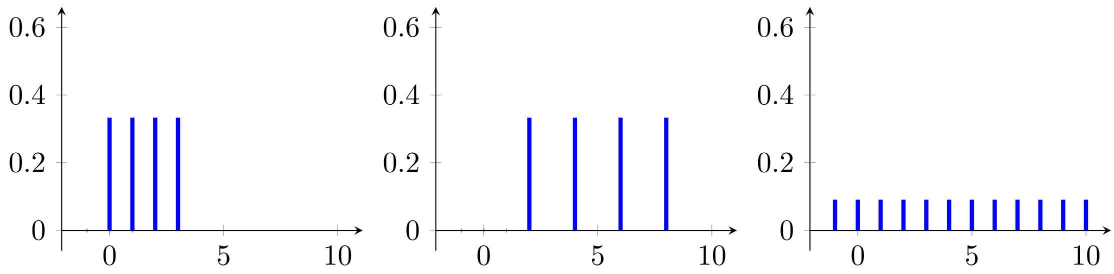

FIGURE 2.2 – Loi uniforme $\mathcal{U}(\{0,1,2,3\})$, $\mathcal{U}(\{2,4,6,8\})$ et $\mathcal{U}(\{-1,0,1,\cdots,10\})$

Loi Binomiale $\mathcal{B}(n,p)$ X représente le nombre de tirages favorables lors de n tirages avec remise dans une urne contenant une proportion p de boules favorables et $1-p$ de boules défavorables, $X(\Omega)=\{0,1,\cdots,n\}$,

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

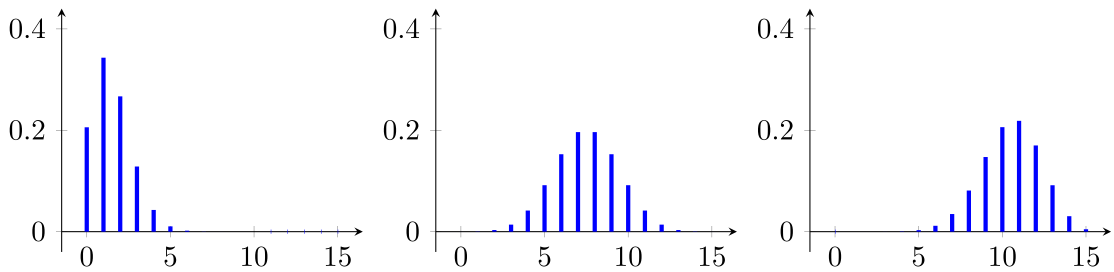

FIGURE 2.3 – Loi Binomiale de paramètre $\mathcal{B}(15,0.1)$, $\mathcal{B}(15,0.5)$ et $\mathcal{B}(15,0.7)$

La loi Binomiale $\mathcal{B}(1,p)$ est la loi de Bernoulli de paramètre p.

Loi Multinomiale $\mathcal{M}(n,m,p_1,\cdots,p_m)$ $X=(X_1,\cdots,X_m)$ suit la loi multinomiale si X_i est le nombre de boules de la couleur i tirées lors de n tirages avec remise dans une urne avec des boules de m couleurs et une proportion p_i de boules de la couleur i

$$\mathbb{P}(X=(n_1,\cdots,n_m))=\frac{n!}{n_1!\cdots n_m!}p_1^{n_1}\cdots p_m^{n_m},$$

$n_1+\cdots+n_m=n$. Chaque X_i suit la loi binomiale $\mathcal{B}(n,p_i)$.

Loi Hypergéométrique $\mathcal{H}(n,N_1,N)$ X suit la loi hypergéométrique si c'est le nombre de boules favorables tirée lors de n tirages sans remise dans une urne contenant N boules dont N_1 favorables (et $N_2=N-N_1$ défavorables), $X(\Omega)=\{0\vee(n-N_2),\cdots,n\wedge N_1\}$,

$$\mathbb{P}(X=k)=\frac{\binom{N_1}{k}\binom{N_2}{n-k}}{\binom{N}{n}}.$$

Si N et N_1 tendent vers l'infini avec $p=N_1/N$ on retrouve la loi binomiale (le **sans remise**) ne joue plus sur une urne infinie).

Loi Géométrique $\mathcal{G}(p)$ C'est la loi du premier succès : on réalise une succession de tirages (indépendants) à une issue défavorable et une issue favorable. C'est le rang du premier succès, $X(\Omega)=\mathbb{N}^*$,

$$\mathbb{P}(X=n)=(1-p)^{n-1}p.$$

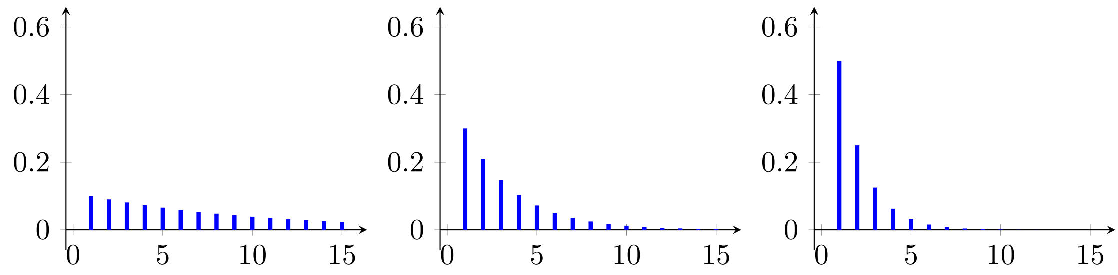

FIGURE 2.4 – Loi géométrique de paramètre $\mathcal{G}(0.1)$, $\mathcal{G}(0.3)$ et $\mathcal{G}(0.5)$

Loi de Poisson $\mathcal{P}(\lambda)$ On l'utilise pour des événements discrets rares, $X(\Omega) = \mathbb{N}$,

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

La loi binomiale de paramètres (n, p_n) avec $p_n = \lambda/n$ converge vers la loi de Poisson de paramètre λ (ce qui justifie la notion d'événement rare).

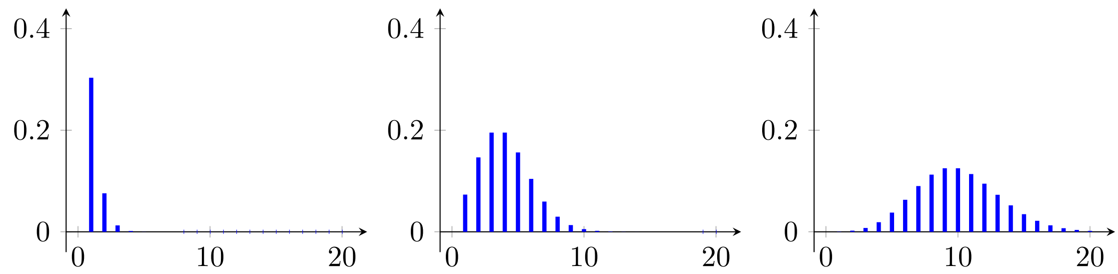

FIGURE 2.5 – Loi de Poisson de paramètre $\mathcal{P}(0.5)$, $\mathcal{P}(4)$ et $\mathcal{P}(10)$

Loi uniforme continue Soit E un sous-ensemble de $\mathbb{R}$ ou $\mathbb{R}^d$ de mesure de Lebesgue finie. La loi uniforme continue sur E est la loi de densité $\mathbb{1}_E/\lambda(E)$ par rapport à la mesure de Lebesgue sur $\mathbb{R}$ ou $\mathbb{R}^d$. Si A est un borélien de E

$$\mathbb{P}(X \in A) = \int_A \frac{\mathbb{1}_E(x)}{\lambda(E)} dx = \frac{\lambda(A)}{\lambda(E)}.$$

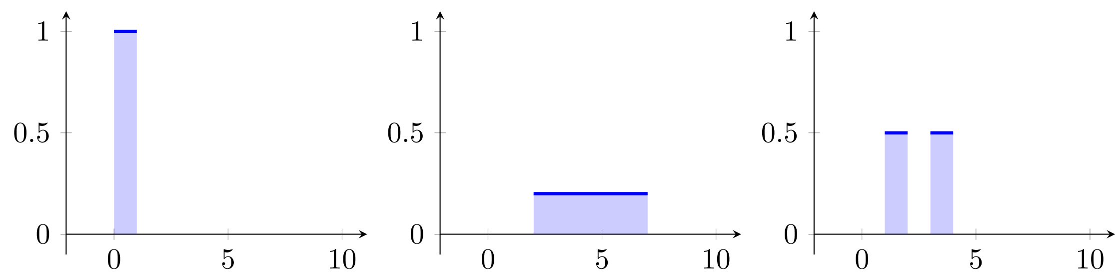

FIGURE 2.6 – Loi uniforme $\mathcal{U}([0,1])$, $\mathcal{U}([2,7])$ et $\mathcal{U}([1,2] \cup [3,4])$

Loi exponentielle $\mathcal{E}(\lambda)$ C'est une loi positive qui modélise des durées de vie d'objets (ampoule, composant électronique, ...) ou des temps d'attente. C'est une loi de densité

$$\lambda \exp(-\lambda x) \mathbb{1}_{x \geq 0}.$$

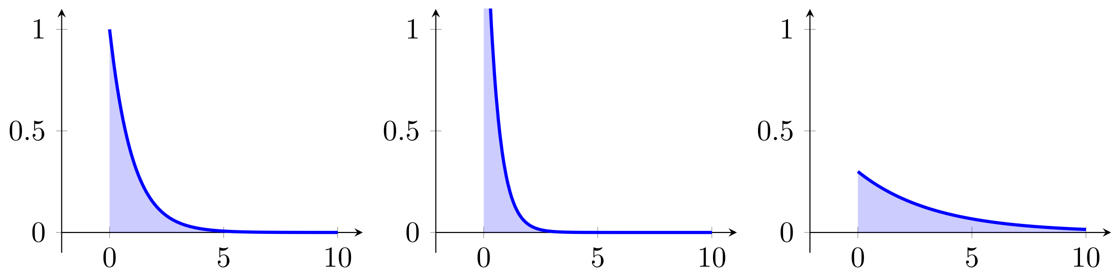

FIGURE 2.7 – Loi exponentielle de paramètre $\mathcal{E}(1)$, $\mathcal{E}(2)$ et $\mathcal{E}(0.3)$

Loi normale ou Gaussienne $\mathcal{N}(m, \sigma)$ C'est la loi qui a pour densité

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

Attention, selon les auteurs on paramétrise cette loi par moyenne/variance ou moyenne/écart type. Elle est universelle en un sens qu'on verra à la fin du semestre (Théorème centrale limite).

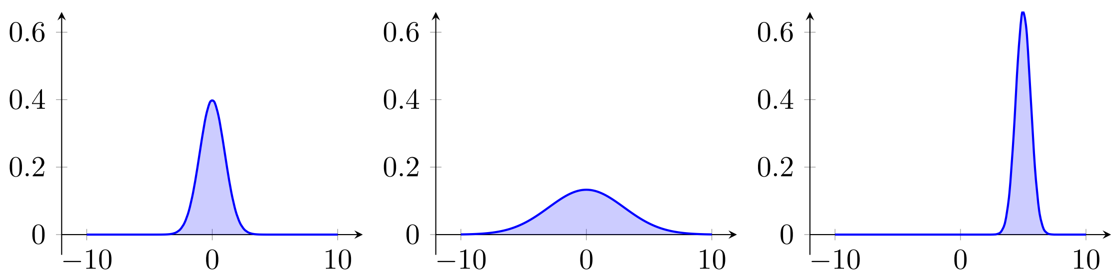

FIGURE 2.8 – Loi normale de paramètre $\mathcal{N}(0, 1)$, $\mathcal{N}(0, 3)$ et $\mathcal{B}(5, 0.6)$

Loi normale multi-dimensionnelle $\mathcal{N}(m, \Sigma)$ où m est un vecteur de dimension d et Σ une matrice carrée de taille d définie positive. Elle a pour densité sur $\mathbb{R}^d$

$$\frac{1}{\mathsf{det}(\Sigma)^{1/2}(2\pi)^{d/2}} \exp(-(x-m)^t\Sigma^{-1}(x-m)/2).$$

On manipulera cette loi surtout au semestre prochain (on parle de vecteurs gaussiens).

2.3 Indépendance de variables aléatoires

La notion d'indépendance peut-être définie également pour les variables aléatoires. Pour cela, on a besoin de définir la tribu engendrée par une variable aléatoire.

Proposition 2.3.1. *Soit X une variable aléatoire sur $(\Omega, \mathcal{F}, \mathbb{P})$ à valeurs dans $(E, \mathcal{E})$. Alors $\mathcal{F}_X = X^{-1}(\mathcal{E}) = \{X^{-1}(A); A \in \mathcal{E}\}$ est une sous-tribu de $\mathcal{F}$ sur Ω. On l'appelle la tribu engendrée par la variable aléatoire X. On la note $\sigma(X)$.*

Démonstration. Montrons que c'est une tribu.

1. $X^{-1}(\emptyset_E) = (X \in \emptyset_E) = \emptyset_\Omega$ donc $\emptyset_\Omega \in \mathcal{F}_X$.
2. Soit $B \in \mathcal{F}_X$. Par définition, il existe $A \in \mathcal{E}$ tel que $B = X^{-1}(A) = (X \in A)$. Alors $B^c = (X \in A)^c = (X \notin A) = (X \in A^c) = X^{-1}(A^c)$, donc $B^c \in \mathcal{F}_X$ car $A^c \in \mathcal{E}$.

3. Soit enfin une suite $(B_n)_{n\in\mathbb{N}}$ d'événements de $\mathcal{F}_X$. Il existe une suite $(A_n)_{n\in\mathbb{N}}$ d'événements de $\mathcal{E}$ telle que $B_n = (X \in A_n)$ pour tout $n \in \mathbb{N}$. Alors $\bigcup_{n\in\mathbb{N}} B_n = \bigcup_{n\in\mathbb{N}}(X \in A_n) = (X \in \bigcup_{n\in\mathbb{N}} A_n)$. Comme $\bigcup_{n\in\mathbb{N}} A_n$ est dans $\mathcal{E}$ qui est une tribu, alors $\bigcup_{n\in\mathbb{N}} B_n$ est dans $\mathcal{F}_X$.

On a donc montré que $\mathcal{F}_X$ est une tribu sur Ω. □

Exemple 2.3.1. Reprenons l'exemple des deux lancers de pièce. On a

$$\begin{aligned}\sigma(X_1) &= \{(X_1 \in \emptyset), (X_1 = 0), (X_1 = 1), (X_1 \in \{0,1\})\}\\ &= \{\emptyset, \{FP, FF\}, \{PP, PF\}, \Omega\},\\ \sigma(X_2) &= \{(X_2 \in \emptyset), (X_2 = 0), (X_2 = 1), (X_2 \in \{0,1\})\}\\ &= \{\emptyset, \{PF, FF\}, \{PP, FP\}, \Omega\}.\end{aligned}$$

Si une famille $\mathcal{A}$ engendre la tribu $\mathcal{E}$, alors $\sigma(X) = \sigma\{X^{-1}(A), A \in \mathcal{A}\}$. Donc

$$\begin{aligned}\sigma(Y) &= \sigma\{(Y = (0,0)), (Y = (0,1)), (Y = (1,0)), (Y = (1,1))\}\\ &= \sigma\{\{FF\}, \{FP\}, \{PF\}, \{FF\}\} = \mathcal{P}(\Omega)\\ \sigma(Z) &= \sigma\{(Z = 0), (Z = 1), (Z = 2)\} = \sigma\{\{FF\}, \{PF, FP\}, \{PP\}\}\\ &= \{\emptyset, \{PP\}, \{FF\}, \{PP, FF\}, \{PF, FP\}, \{PP, PF, FP\}, \{PF, FP, FF\}, \Omega\}.\end{aligned}$$

Remarquons que les quatre tribus $\mathcal{F}_{X_1}$, $\mathcal{F}_{X_2}$, $\mathcal{F}_Y$ et $\mathcal{F}_Z$ sont différentes. On a vu que X_1 et X_2 ont la même loi, mais elles n'engendrent pas la même tribu.

Définition 2.3.1. *Soit X et Y deux variables aléatoires sur $(\Omega, \mathcal{F}, \mathbb{P})$ à valeurs dans $(E_1, \mathcal{E}_1)$ et $(E_2, \mathcal{E}_2)$. On dit que X et Y sont indépendantes si les tribus $\sigma(X)$ et $\sigma(Y)$ le sont. Une famille quelconque $(X_i)_{i\in I}$ de variables aléatoires est indépendante si les tribus $\sigma(X_i)$ le sont.*

Même si X et Y ne prennent pas leurs valeurs dans les mêmes ensembles, les tribus $\sigma(X)$ et $\sigma(Y)$ sont des tribus sur Ω, on peut donc les comparer en terme d'indépendance.

Exemple 2.3.2. Reprenons l'exemple des deux lancers de pièce. On a vu

$$\begin{aligned}\sigma(X_1) &= \{(X_1 \in \emptyset), (X_1 = 0), (X_1 = 1), (X_1 \in \{0,1\})\} = \{\emptyset, \{FP, FF\}, \{PP, PF\}, \Omega\},\\ \sigma(X_2) &= \{(X_2 \in \emptyset), (X_2 = 0), (X_2 = 1), (X_2 \in \{0,1\})\} = \{\emptyset, \{PF, FF\}, \{PP, FP\}, \Omega\},\\ \sigma(Y) &= \mathcal{P}(\Omega),\\ \sigma(Z) &= \{\emptyset, \{PP\}, \{FF\}, \{PP, FF\}, \{PF, FP\}, \{PP, PF, FP\}, \{PF, FP, FF\}, \Omega\}.\end{aligned}$$

X_1 et X_2 sont indépendantes, X_1 et Y, X_1 et Z, Y et Z ne sont pas indépendantes car $\mathbb{P}(\{FP, FF\} \cap \{PP\}) = 0 \neq \mathbb{P}(\{FP, FF\})\mathbb{P}(\{PP\})$.

Proposition 2.3.2. *Soit X une variable aléatoire sur $(\Omega, \mathcal{F}, \mathbb{P})$ à valeurs dans $(E, \mathcal{E})$, h une fonction mesurable de $(E, \mathcal{E})$ dans $(G, \mathcal{G})$ et $Y = h(X)$. Alors $\sigma(Y) \subset \sigma(X)$.*

Démonstration. Soit $A \in \sigma(Y)$. Alors il existe un événement $B \in \mathcal{G}$ tel que $A = Y^{-1}(B) = (h \circ X)^{-1}(B) = X^{-1}(h^{-1}(B))$. Comme h est mesurable, $h^{-1}(B) \in \mathcal{E}$ donc $A \in \sigma(X)$. □

Exemple 2.3.3. Reprenons l'exemple des deux lancers de pièce. On a vu

$$\begin{aligned}\sigma(Y) &= \mathcal{P}(\Omega),\\ \sigma(Z) &= \{\emptyset, \{PP\}, \{FF\}, \{PP, FF\}, \{PF, FP\}, \{PP, PF, FP\}, \{PF, FP, FF\}, \Omega\}.\end{aligned}$$

$Z = Y_1 + Y_2$ et on a bien $\sigma(Z) \subset \sigma(Y)$.

Proposition 2.3.3. *Soit X et Y deux variables aléatoires indépendantes sur $(\Omega, \mathcal{F}, \mathbb{P})$ à valeurs dans $(E_1, \mathcal{E}_1)$ et $(E_2, \mathcal{E}_2)$, et g et h deux fonctions mesurables de $(E_1, \mathcal{E}_1)$ dans $(G_1, \mathcal{G}_1)$ et de $(E_2, \mathcal{E}_2)$ dans $(G_2, \mathcal{G}_2)$ respectivement. Alors $g(X)$ et $h(Y)$ sont indépendantes.*

Démonstration. On a vu que $\sigma(g(X)) \subset \sigma(X)$ et $\sigma(h(Y)) \subset \sigma(Y)$. C'est donc une conséquence directe de la définition d'indépendance des tribus. □

Exemple 2.3.4. Soit $X = (X_1, X_2)$ et $Y = (Y_1, Y_2)$ deux vecteurs aléatoires indépendants. Alors X_1 est indépendant de Y, Y_1 et Y_2.

Proposition 2.3.4. *Soit X et Y deux variables aléatoires sur $(\Omega, \mathcal{F}, \mathbb{P})$ à valeurs dans $(E_1, \mathcal{E}_1)$ et $(E_2, \mathcal{E}_2)$, et $\mathcal{C}_1$, $\mathcal{C}_2$ des familles de parties de E_1 et E_2 respectivement qui engendrent $\mathcal{E}_1$ et $\mathcal{E}_2$ respectivement et sont stables par intersection finie. Alors X et Y sont indépendantes si et seulement si les familles d'événements $X^{-1}(\mathcal{C}_1)$ et $Y^{-1}(\mathcal{C}_2)$ sont indépendantes.*

Démonstration. On a $\sigma(X^{-1}(\mathcal{C}_1)) = X^{-1}(\sigma(\mathcal{C}_1)) = X^{-1}(\mathcal{E}_1)$ et de même pour Y. De plus la propriété de stabilité par intersection finie se transmet à $X^{-1}(\mathcal{C}_1)$ et $Y^{-1}(\mathcal{C}_2)$. C'est donc une conséquence directe du résultat sur les tribus (Proposition **??**) qui dit que si $X^{-1}(\mathcal{C}_1)$ et $Y^{-1}(\mathcal{C}_2)$ sont deux familles d'événements de $\mathcal{F}$ stables par intersection finie, alors $X^{-1}(\mathcal{C}_1)$ et $Y^{-1}(\mathcal{C}_2)$ sont indépendantes si et seulement si $\sigma(X^{-1}(\mathcal{C}_1))$ et $\sigma(Y^{-1}(\mathcal{C}_2))$ le sont. □

Exemple 2.3.5. Pour des variables aléatoires X et Y à valeurs dans des ensembles dénombrables $E = (e_n, n \in \mathbb{N})$ et $G = (g_n, n \in \mathbb{N})$ munis de la tribu des parties, on sait que E engendre $\mathcal{P}(E)$ et G engendre $\mathcal{P}(G)$. Ils sont stables par intersection finie donc X et Y sont indépendantes si et seulement si pour tous n et m dans $\mathbb{N}$ les événements $(X = e_n)$ et $(Y = g_m)$ sont indépendants.
Pour des variables aléatoires réelles X et Y, comme les boréliens sont engendrés par exemple par les intervalles de la forme $[-\infty, t]$ qui sont stables par intersection finie, X et Y sont indépendantes si et seulement si pour tous s, t dans $\mathbb{R}$ les événements $(X \leq s)$ et $(Y \leq t)$ sont indépendants.

Proposition 2.3.5. *Soit X et Y deux variables aléatoires sur $(\Omega, \mathcal{F}, \mathbb{P})$ à valeurs dans $(E_1, \mathcal{E}_1)$ et $(E_2, \mathcal{E}_2)$. Alors les variables aléatoires X et Y sont indépendantes si et seulement si la loi du couple (X, Y) est égale à la loi produit des lois de X et Y*

$$\mathbb{P}_{(X,Y)} = \mathbb{P}_X \otimes \mathbb{P}_Y.$$

Démonstration. Par définition, X et Y sont indépendantes si et seulement si pour tous $A \in \mathcal{E}_1$ et $B \in \mathcal{E}_2$ on a $\mathbb{P}(\{X \in A\} \cap \{Y \in B\}) = \mathbb{P}(X \in A)\mathbb{P}(Y \in B)$. Or

$$\mathbb{P}(\{X \in A\} \cap \{Y \in B\}) = \mathbb{P}((X, Y) \in A \times B) = \mathbb{P}_{(X,Y)}(A \times B),$$
$$\mathbb{P}(X \in A)\mathbb{P}(Y \in B) = \mathbb{P}_X(A)\mathbb{P}_Y(B).$$

Donc X et Y sont indépendantes si et seulement si pour tous $A \in \mathcal{E}_1$ et $B \in \mathcal{E}_2$ on a $\mathbb{P}_{(X,Y)}(A \times B) = \mathbb{P}_X(A)\mathbb{P}_Y(B)$.

Rappel 1. [Théorème de caractérisation des mesures sur un π-système] Soit μ_1 et μ_2 deux mesures positives sur l'espace probabilisable $(\Omega, \mathcal{F})$ telles que $\mu_1(A) = \mu_2(A)$ pour tout $A \in \mathcal{C}$ où $\mathcal{C}$ est une famille de parties de Ω stable par intersection finie (i.e. un π-système) qui engendre $\mathcal{F}$. Si $\mu_1(\Omega) = \mu_2(\Omega) < \infty$, alors $\mu_1 = \mu_2$. En vertu du théorème d'unicité des mesures ceci est équivalent à $\mathbb{P}_{(X,Y)} = \mathbb{P}_X \otimes \mathbb{P}_Y$ car les pavés engendrent tous les boréliens. □

2.4 Fonction de répartition pour les variables aléatoires réelles

Dans cette partie, on ne s'intéresse qu'aux variables aléatoires réelles. Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace probabilisé et X une variable aléatoire sur $(\Omega, \mathcal{F}, \mathbb{P})$ à valeurs dans $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$.

Définition 2.4.1. *On appelle fonction de répartition de X ou de la loi $\mathbb{P}_X$ de X la fonction F_X de $\mathbb{R}$ dans $[0,1]$ définie par*

$$F_X(t) = \mathbb{P}_X(]-\infty, t]) = \mathbb{P}(X \leq t).$$

Exemple 2.4.1. La fonction de répartition de la masse de Dirac en x est $F(t) = \mathbb{1}_{[x,+\infty[}(t)$.

Proposition 2.4.1.
- *Si X est une variable aléatoire discrète de loi $\sum_{i\in\mathbb{N}} p_i \delta_{x_i}$ sa fonction de répartition est la fonction en escalier $F(t) = \sum_{i\in\mathbb{N}} p_i \mathbb{1}_{[x_i,+\infty[}(t)$.*
- *Si X est une variable aléatoire absolument continue de densité f, alors sa fonction de répartition est $F(t) = \int_{-\infty}^{t} f(u)du$. En particulier, la densité f est la dérivée de la fonction de répartition F.*

Démonstration. Si X est une variable aléatoire discrète de loi $\sum_{i\in\mathbb{N}} p_i \delta_{x_i}$, par définition, on a

$$\begin{aligned} F(t) &= \mathbb{P}(X \leq t) \\ &= \int_{-\infty}^{t} \sum_{i\in\mathbb{N}} p_i d\delta_{x_i}(u) \\ &= \sum_{i\in\mathbb{N}} p_i \int_{-\infty}^{t} d\delta_{x_i}(u) \\ &= \sum_{i\in\mathbb{N}} p_i \mathbb{1}_{[x_i,+\infty[}(t), \end{aligned}$$

où l'on a pu intervertir série et intégrale car tout est positif. Si X est une variable aléatoire absolument continue de densité f, par définition, on a

$$F(t) = \mathbb{P}(X \leq t) = \int_{-\infty}^{t} f(u)du,$$

ce qui entraine $F' = f$ puisque l'égalité est valable pour tout réel t. $\square$

Exemple 2.4.2. Reprenons l'exemple 2.1.1 du lancer de 2 pièces. On a

$$\begin{aligned} F_{X_1}(t) &= \tfrac{1}{2}\mathbb{1}_{[0,+\infty[}(t) + \tfrac{1}{2}\mathbb{1}_{[1,+\infty[}(t) \\ &= \tfrac{1}{2}\mathbb{1}_{[0,1[}(t) + \mathbb{1}_{[1,+\infty[}(t). \end{aligned}$$

Exemple 2.4.3. Pour l'exemple 2.1.5, on a

$$\begin{aligned} F_X(t) &= \int_{-\infty}^{t} e^{-u}\mathbb{1}_{u\geq 0}du \\ &= \int_{0}^{t} e^{-u}du \\ &= 1 - e^{-t}, \end{aligned}$$

qu'on avait déjà calculé directement.

Exemple 2.4.4. Revenons maintenant à l'exemple 2.1.1 où $Y = X \wedge 1$. On a vu que $\mathbb{P}(Y \in A) = \mathbb{1}_A(1)e^{-1} + \int_{A\cap[0,1[} e^{-u}du$, donc on a

$$\begin{aligned} F_Y(t) &= \mathbb{1}_{]-\infty,t]}(1)e^{-1} + \int_{]-\infty,t]\cap[0,1[} e^{-u}du \\ &= \mathbb{1}_{t\geq 1}(t)e^{-1} + \mathbb{1}_{[0,1[}(t)(1-e^{-t}) + \mathbb{1}_{t\geq 1}(t)(1-e^{-1})) \\ &= (1-e^{-t})\mathbb{1}_{[0,1[}(t) + \mathbb{1}_{[1,+\infty[}(t), \end{aligned}$$

qui n'est pas dérivable en 1, voir Figure 2.9.

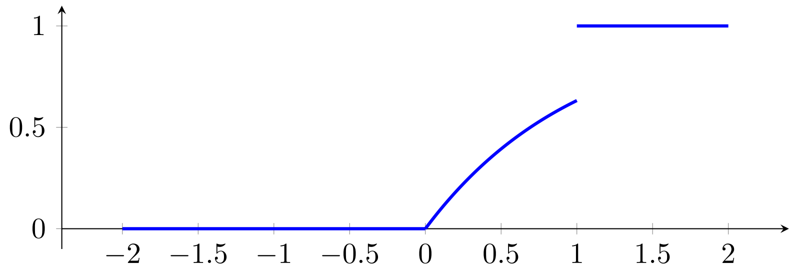

FIGURE 2.9 – Fonction de répartition de Y

Proposition 2.4.2. *Une fonction de répartition F vérifie les propriétés suivantes*

1. $0 \leq F \leq 1$,
2. *F est croissante, continue à droite avec une limite à gauche en tout point (càdlàg),*
3. $\lim_{t\to-\infty} F(t) = 0$ *et* $\lim_{t\to+\infty} F(t) = 1$.

Démonstration. Le premier point vient de $F(t) = \mathbb{P}(X \leq t) \in [0,1]$.
La croissance découle de la croissance des probabilités : $A \subset B \Rightarrow \mathbb{P}(A) \leq \mathbb{P}(B)$. En effet si $s \leq t$ alors $]-\infty, s] \subset]-\infty, t]$ donc $F(s) \leq F(t)$.
Pour la continuité à droite, remarquons qu'on peut écrire l'intervalle $]-\infty, t]$ comme une intersection décroissante

$$]-\infty, t] = \bigcap_{n\geq 1}]-\infty, t+\tfrac{1}{n}],$$

d'où

$$F(t) = \mathbb{P}(X \leq t) = \mathbb{P}(\bigcap_{n\geq 1} X \leq t+\tfrac{1}{n}) = \lim_{n\to\infty} \mathbb{P}(X \leq t+\tfrac{1}{n}) = \lim_{n\to\infty} F(t+\tfrac{1}{n}).$$

Ensuite, par croissance de F, on a

$$\lim_{h\to 0^+} F(t+h) = \lim_{n\to\infty} F(t+\tfrac{1}{n}) = F(t).$$

La limite à gauche est une conséquence du fait que F est croissante et bornée par 1. Enfin pour les limites en $\pm\infty$, on a

$$0 = \mathbb{P}(\emptyset) = \lim_{n\to\infty} \mathbb{P}(X \leq -n) = \lim_{n\to\infty} F(-n),$$

en utilisant la probabilité d'une limite décroissante et

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

en utilisant la probabilité d'une limite croissante. On conclut en utilisant encore la croissance pour obtenir la limite de $F(t)$. □

On peut également montrer la réciproque : une fonction vérifiant ces trois points est la fonction de répartition d'une variable aléatoire réelle. Pour cela, pour $a < b$ on pose $\mu(]a,b]) = F(b) - F(a)$ et on étend cette définition aux unions finies d'intervalles. C'est une fonction d'ensemble σ-additive et elle est positive car F est croissante. Alors μ se prolonge de façon unique en une mesure sur $\big(\mathbb{R}, \mathcal{B}(\mathbb{R})\big)$ (cf cours de théorie de la mesure). De plus $\mu(\mathbb{R}) = \lim_{t\to+\infty} F(t) - \lim_{t\to-\infty} F(t) = 1$ donc μ est une probabilité.

Proposition 2.4.3. *La fonction de répartition caractérise la loi : si X et Y sont deux variables aléatoires sur $(\Omega, \mathcal{F}, \mathbb{P})$ à valeurs dans $\big(\mathbb{R}, \mathcal{B}(\mathbb{R})\big)$ de fonction de répartition respective F_X et F_Y alors $F_X = F_Y$ si et seulement si $\mathbb{P}_X = \mathbb{P}_Y$.*

Démonstration. Si $\mathbb{P}_X = \mathbb{P}_Y$, alors par définition $F_X = F_Y$. Réciproquement, si $F_X = F_Y$ alors $\mathbb{P}_X$ et $\mathbb{P}_Y$ coïncident sur les intervalles de la forme $]-\infty, t]$. Or ces intervalles engendrent la tribu des boréliens,

Rappel 2. [Théorème de caractérisation des mesures sur un π-système] Soit μ_1 et μ_2 deux mesures positives sur l'espace probabilisable $(\Omega, \mathcal{F})$ telles que $\mu_1(A) = \mu_2(A)$ pour tout $A \in \mathcal{C}$ où $\mathcal{C}$ est une famille de parties de Ω stable par intersection finie (i.e. un π-système) qui engendre $\mathcal{F}$. Si $\mu_1(\Omega) = \mu_2(\Omega) < \infty$, alors $\mu_1 = \mu_2$. on a donc $\mathbb{P}_X = \mathbb{P}_Y$ pour tous les boréliens. □

Proposition 2.4.4. *Une fonction de répartition admet au plus un nombre dénombrable de points de discontinuité.*

Démonstration. Soit D_n l'ensemble des points de discontinuité avec un saut d'amplitude au moins égale à $\frac{1}{n}$

$$D_n = \{t \in \mathbb{R}; F(t) - \lim_{s\to t^-} F(s) \geq \tfrac{1}{n}\}.$$

Comme $0 \leq F \leq 1$, on a $card(D_n) \leq n$. L'ensemble des points de discontinuité $D = \bigcup_{n\in\mathbb{N}} D_n$ est donc dénombrable. □

Définition 2.4.2. *On appelle fonction de survie de X ou de la loi $\mathbb{P}_X$ de X la fonction S_X de $\mathbb{R}$ dans $[0,1]$ définie par*

$$S_X(t) = 1 - F_X(t) = \mathbb{P}_X(]t, +\infty[) = \mathbb{P}(X > t).$$

La fonction de survie caractérise également la loi et a des propriétés similaires à celles de la fonction de répartition. Pour certaines applications (en démographie, médecine, ...) ou pour certaines lois, elle peut-être plus pratique à manipuler que la fonction de répartition.

Exemple 2.4.5. Soit X et Y des variables aléatoires réelles indépendantes de même fonction de survie S. Alors la variable aléatoire $Z = \min\{X, Y\}$ a pour fonction de survie S^2. En effet, pour tout réel t on a $\mathbb{P}(Z > t) = \mathbb{P}(X > t, Y > t) = \mathbb{P}(X > t)\mathbb{P}(Y > t) = S(t)S(t)$. Le calcul est un peu plus compliqué avec la fonction de répartition car $(Z \leq t) \neq (X \leq t) \cap (Y \leq t)$.

Exemple 2.4.6. On dit que X suit une loi exponentielle de paramètre α si $X \geq 0$ et pout tout $t \geq 0$ on a $\mathbb{P}(X > t) = e^{-\alpha t}$. Cette loi est **sans mémoire** au sens suivant : pour tous $0 \leq s, t$

$$\mathbb{P}(X > t + s \mid X > s) = \mathbb{P}(X > t).$$

En effet, on peut calculer cette probabilité

$$\begin{aligned}\mathbb{P}(X > t+s \mid X > s) &= \frac{\mathbb{P}(X > t+s, X > s)}{\mathbb{P}(X > s)} \\ &= \frac{\mathbb{P}(X > t+s)}{\mathbb{P}(X > s)} \\ &= \frac{e^{-\alpha(t+s)}}{e^{-\alpha s}} \\ &= e^{-\alpha t} \\ &= \mathbb{P}(X > t).\end{aligned}$$

De plus, ce résultat admet la réciproque suivante : si X est une variable aléatoire sans mémoire, alors elle suit une loi exponentielle. La propriété d'absence de mémoire peut se récrire de façon équivalente avec la fonction de survie comme

$$S(t+s) = S(t)S(s),$$

et on sait que les seules fonctions multiplicatives (continues à gauche, ou décroissantes) sont les exponentielles. On pose $\alpha = -\log S(1) > 0$ et on obtient $S(t) = e^{-\alpha t}$.

2.5 Espérance

On note $|x|$ la valeur absolue de x si $x \in \mathbb{R}$ ou la norme de x si $x \in \mathbb{R}^d$ (pour une norme fixée, comme toutes les normes sont équivalentes en dimension finie, ce choix n'a pas d'importance). Soit X une variable aléatoire réelle ou un vecteur aléatoire.

Définition 2.5.1. *Si $X \in L^1(\Omega, \mathcal{F}, \mathbb{P})$, ie*

$$\int_\Omega |X(\omega)| d\mathbb{P}(\omega) < +\infty,$$

on dit que X est intégrable et alors $\int_\Omega X(\omega)d\mathbb{P}(\omega)$ s'appelle l'espérance de X, on la note $\mathbb{E}[X]$. Si X est une variable aléatoire réelle, $\mathbb{E}[X]$ est un réel, si X est un vecteur aléatoire, $\mathbb{E}[X]$ est un vecteur.

Exemple 2.5.1. Reprenons l'exemple des deux lancers de pièce. Comme X_1, X_2, Y et Z sont à valeurs positives, on a

$$\begin{aligned}\mathbb{E}[X_1] &= \int_\Omega X_1(\omega)d\mathbb{P}(\omega) \\ &= X_1(PP)\mathbb{P}(PP) + X_1(PF)\mathbb{P}(PF) + X_1(FP)\mathbb{P}(FP) + X_1(FF)\mathbb{P}(FF) \\ &= 1 \times \frac{1}{4} + 1 \times \frac{1}{4} + 0 \times \frac{1}{4} + 0 \times \frac{1}{4} \\ &= \frac{1}{2}.\end{aligned}$$

On montre de la même façon que $\mathbb{E}[X_2] = 1/2$, $\mathbb{E}[Z] = \mathbb{E}[X_1] + \mathbb{E}[X_2] = 1$ par linéarité de l'intégrale. On a aussi, en choisissant la norme 1 par exemple

$$\mathbb{E}[|Y|] = \int_\Omega |X_1(\omega)| + |X_2(\omega)| d\mathbb{P}(\omega) \leq \mathbb{E}[X_1] + \mathbb{E}[X_2] < +\infty,$$

d'où

$$\begin{aligned}
\mathbb{E}[Y] &= \int_\Omega (X_1(\omega), X_2(\omega)) d\mathbb{P}(\omega) \\
&= (X_1(PP), X_2(PP))\mathbb{P}(PP) + (X_1(PF), X_2(PF))\mathbb{P}(PF) \\
&\quad + (X_1(FP), X_2(FP))\mathbb{P}(FP) + (X_1(FF), X_2(FF))\mathbb{P}(FF) \\
&= (1,1) \times \frac{1}{4} + (1,0) \times \frac{1}{4} + (0,1) \times \frac{1}{4} + (0,0) \times \frac{1}{4} \\
&= (\frac{1}{2}, \frac{1}{2}) \\
&= (\mathbb{E}[X_1], \mathbb{E}[X_2]),
\end{aligned}$$

le dernier résultat étant une conséquence directe des propriétés de l'intégrale.

Exemple 2.5.2. Reprenons l'exemple logarithmique. La variable aléatoire X est positive et

$$\begin{aligned}
\mathbb{E}[X] &= \int_\Omega X(\omega) d\mathbb{P}(\omega) \\
&= \int_0^1 - \log(\omega) d\lambda(\omega) \\
&= [-\omega \log(\omega)]_0^1 + \int_0^1 \frac{1}{\omega} \omega d\omega \\
&= 1,
\end{aligned}$$

par intégration par partie. Donc X est intégrable et son espérance vaut 1.

Proposition 2.5.1. *L'espérance est linéaire : si X et Y sont des variables aléatoires intégrables, alors pour tous a et b dans $\mathbb{R}$ la variable aléatoire $Z = aX + bY$ est intégrable et $\mathbb{E}[Z] = a\mathbb{E}[X] + b\mathbb{E}[Y]$. Un vecteur $X = (X_1, \cdots, X_d)$ est intégrable si et seulement si chacune de ses coordonnées est intégrable, et dans ce cas $\mathbb{E}[X] = (\mathbb{E}[X_1], \cdots, \mathbb{E}[X_d])$.*

Démonstration. Ce sont des conséquences directes de la linéarité de l'intégrale et de la définition de l'intégrale d'un vecteur. □

Dans la plupart des exemples, on ne précise pas l'espace de probabilité $(\Omega, \mathcal{F}, \mathbb{P})$ et on ne peut donc pas intégrer directement sur Ω. Le résultat suivant permet de faire les calculs dans ce cas, et justifie le fait que l'on n'a pas besoin de préciser l'ensemble Ω.

Théorème 2.5.1 Théorème de transfert. *Soit $X \in L^1(\Omega, \mathcal{F}, \mathbb{P})$ à valeurs dans $(E, \mathcal{E})$ et h une fonction mesurable de E dans E. Alors $h(X) \in L^1(\Omega, \mathcal{F}, \mathbb{P})$ si et seulement si $h \in L^1(E, \mathcal{E}, \mathbb{P}_X)$ et dans ce cas on a*

$$\mathbb{E}[h(X)] = \int_\Omega h \circ X(\omega) d\mathbb{P}(\omega) = \int_\mathbb{R} h(x) \mathbb{P}_X(dx)$$

Démonstration. Comme $\mathbb{P}_X$ est la mesure image de $\mathbb{P}$ par X, c'est une conséquence directe du théorème de transfert vu en théorie de la mesure. □

Grâce à ce résultat, tous les calculs se font dans l'espace d'arrivée, c'est pour ça qu'on n'a pas besoin de s'occuper de décrire l'ensemble Ω. Les calculs d'espérance se font dans $\mathbb{R}$ ou $\mathbb{R}^d$ comme des intégrales contre la loi de X.

Exemple 2.5.3. Pour une variable discrète, $\mathbb{P}_X = \sum p_i \delta_{x_i}$ on a (sous réserve d'existence)

$$\mathbb{E}[X] = \int x \mathbb{P}_X(dx) = \sum x_i p_i.$$

Pour une variable aléatoire à densité f, $\mathbb{P}_X = f.\lambda$ donc on a (sous réserve d'existence)

$$\mathbb{E}[X] = \int x \mathbb{P}_X(dx) = \int x f(x) dx.$$

Exemple 2.5.4. Espérance des lois usuelles
Loi de Bernoulli $\mathcal{B}(p)$ **:** $\mathbb{E}[X] = p$

Loi Uniforme discrète $\mathcal{U}(\{1, \cdots, n\})$ **:** $\mathbb{E}[X] = (n+1)/2$

Loi Binomiale $\mathcal{B}(n,p)$ **:** $\mathbb{E}[X] = np$

Loi Multinomiale $\mathcal{M}(n, m, p_1, \cdots, p_m)$ **:** $\mathbb{E}[X] = (np_1, \cdots, np_m)$

Loi Hypergéométrique $\mathcal{H}(n, N_1, N)$ **:** $\mathbb{E}[X] = nN_1/N$

Loi Géométrique $\mathcal{G}(p)$ **:** $\mathbb{E}[X] = 1/p$

Loi de Poisson $\mathcal{P}(\lambda)$ **:** $\mathbb{E}[X] = \lambda$

Loi uniforme continue $\mathcal{U}([a,b])$ **:** $\mathbb{E}[X] = (a+b)/2$

Loi exponentielle $\mathcal{E}(\lambda)$ **:** $\mathbb{E}[X] = 1/\lambda$

Loi normale $\mathcal{N}(m, \sigma^2)$ **:** $\mathbb{E}[X] = m$

Loi normale multi-dimensionnelle $\mathcal{N}(m, \Sigma)$ **:** $\mathbb{E}[X] = m$

Exemple 2.5.5. Si X est un vecteur aléatoire (ou une variable aléatoire réelle) et si $A \in \mathcal{B}(\mathbb{R}^d)$ ou $A \in \mathcal{B}(\mathbb{R})$, alors $\mathbb{1}_A$ est une fonction mesurable et

$$\begin{aligned} \mathbb{E}[\mathbb{1}_A(X)] &= \int_\Omega \mathbb{1}_A(X(\omega)) d\mathbb{P}(\omega) \\ &= \int \mathbb{1}_A(x) d\mathbb{P}_X(x) \\ &= \mathbb{P}_X(A) \\ &= \mathbb{P}(X \in A). \end{aligned}$$

En particulier, la fonction de répartition d'une variable aléatoire réelle peut s'écrire comme $F_X(t) = \mathbb{P}(X \in]-\infty, t]) = \mathbb{E}[\mathbb{1}_{]-\infty,t]}(X)]$.

Définition 2.5.2. *Soit X une variable aléatoire intégrable (réelle ou vectorielle). On dit que X est centrée si $\mathbb{E}[X] = 0$.*

Exemple 2.5.6. Soit X une variable aléatoire intégrable (réelle ou vectorielle), alors $X - \mathbb{E}[X]$ est centrée.

Définition 2.5.3. *Soit X une variable aléatoire réelle. Si X^p est intégrable ($X^p \in L^1(\Omega, \mathcal{F}, \mathbb{P})$ ou $X \in L^p(\Omega, \mathcal{F}, \mathbb{P})$), on appelle $\mathbb{E}[X^p]$ le moment d'ordre p de X et $\mathbb{E}[(X - \mathbb{E}[X])^p]$ son* *moment centré d'ordre p. En particulier, le moment centré d'ordre 2 s'appelle la variance*

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

et sa racine carrée l'écart-type

$$\sigma_X = (\mathbb{E}[(X - \mathbb{E}[X])^2])^{1/2}.$$

Une variable aléatoire réelle centrée de variance 1 est dite ***centrée réduite****.*

Proposition 2.5.2. *Soit X une variable aléatoire réelle de carré intégrable. Alors*

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

Démonstration. En développant le carré, on obtient

$$\begin{aligned}
\mathsf{Var}(X) &= \mathbb{E}[(X - \mathbb{E}[X])^2] \\
&= \mathbb{E}[X^2 + \mathbb{E}[X]^2 - 2X\mathbb{E}[X]] \\
&= \mathbb{E}[X^2] + \mathbb{E}[X]^2 - 2\mathbb{E}[X]\mathbb{E}[X] \\
&= \mathbb{E}[X^2] - (\mathbb{E}[X])^2,
\end{aligned}$$

qui est bien la formule annoncée. □

Exemple 2.5.7. Variance des lois usuelles

Loi de Bernoulli $\mathcal{B}(p)$ ***:*** $\mathsf{Var}(X) = p(1-p)$

Loi Uniforme discrète $\mathcal{U}(\{1, \cdots, n\})$ ***:*** $\mathsf{Var}(X) = (n^2 - 1)/12$

Loi Binomiale $\mathcal{B}(n, p)$ ***:*** $\mathsf{Var}(X) = np(1-p)$

Loi Hypergéométrique $\mathcal{H}(n, N_1, N)$ ***:*** $\mathsf{Var}(X) = nN_1(N-n)(N-N_1)/(N^2(N-1))$

Loi Géométrique $\mathcal{G}(p)$ ***:*** $\mathsf{Var}(X) = (1-p)/p^2$

Loi de Poisson $\mathcal{P}(\lambda)$ ***:*** $\mathsf{Var}(X) = \lambda$

Loi uniforme continue $\mathcal{U}([a, b])$ ***:*** $\mathsf{Var}(X) = (b-a)^2/12$

Loi exponentielle $\mathcal{E}(\lambda)$ ***:*** $\mathsf{Var}(X) = 1/\lambda^2$

Loi normale $\mathcal{N}(m, \sigma^2)$ ***:*** $\mathsf{Var}(X) = \sigma^2$. En particulier, $\mathcal{N}(0, 1)$ s'appelle la loi normale centrée réduite.

Proposition 2.5.3. *La variance est une fonction quadratique : pour tout réel a on a $\mathsf{Var}(aX) = a^2 \mathsf{Var}(X)$. De plus $\mathsf{Var}(X + a) = \mathsf{Var}(X)$.*

Démonstration. On a

$$\begin{aligned}
\mathsf{Var}(aX) &= \mathbb{E}[(aX - \mathbb{E}[aX])^2] \\
&= a^2\mathbb{E}[(X - \mathbb{E}[X])^2],
\end{aligned}$$

De plus, on a

$$\begin{aligned}
\mathsf{Var}(X + a) &= \mathbb{E}[(X + a - \mathbb{E}[X + a])^2] \\
&= \mathbb{E}[(X - \mathbb{E}[X])^2],
\end{aligned}$$

qui est bien égal à $\mathsf{Var}(X)$. □

Exemple 2.5.8. Soit X une variable aléatoire de loi normale centrée réduite. Alors $\sigma X + m$ a pour espérance m et variance σ^2. Si X est une variable aléatoire de loi normale $\mathcal{N}(m, \sigma^2)$, alors $(X - m)/\sigma$ est centrée réduite.

Définition 2.5.4. *Soit X et Y deux variables aléatoires réelles de carré intégrable. On appelle **covariance** de X et Y la quantité*

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

Remarquons que la covariance dépend de loi du couple (X, Y) et pas seulement des lois marginales de X et Y. Cette quantité est bien définie par l'inégalité de Cauchy-Schwartz.

Proposition 2.5.4. *Soit X et Y deux variables aléatoires réelles de carré intégrable. Alors*

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

Démonstration. On a

$$\begin{aligned}
\mathsf{Cov}(X, Y) &= \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] \\
&= \mathbb{E}[XY - X\mathbb{E}[Y] - \mathbb{E}[X]Y + \mathbb{E}[X]\mathbb{E}[Y]] \\
&= \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y].
\end{aligned}$$

□

Proposition 2.5.5. *Soit X et Y deux variables aléatoires réelles de carré intégrable. Alors*

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

Démonstration. On a

$$\begin{aligned}
\mathsf{Var}(X + Y) &= \mathbb{E}[(X + Y - \mathbb{E}[X + Y])^2] \\
&= \mathbb{E}[((X - \mathbb{E}[X]) + (Y - \mathbb{E}[Y]))^2] \\
&= \mathbb{E}[(X - \mathbb{E}[X])^2] + \mathbb{E}[(Y - \mathbb{E}[Y])^2] + 2\mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] \\
&= \mathsf{Var}(X) + \mathsf{Var}(Y) + 2\,\mathsf{Cov}(X, Y),
\end{aligned}$$

qui est bien la formule annoncée. □

Exemple 2.5.9. Soit le vecteur $X = (X_1, X_2)$ de loi donnée par le tableau suivant

		X_1		
		-1	0	1
X_2	-1	0	$1/4$	0
	0	$1/4$	0	$1/4$
	1	0	$1/4$	0

On a $\mathbb{E}[X_1] = \mathbb{E}[X_2] = 0$ et $\mathbb{E}[X_1 X_2] = 0$, donc $\mathsf{Cov}(X_1, X_2) = 0$.

Définition 2.5.5. *Soit X et Y deux variables aléatoires réelles de carré intégrable. On dit que X et Y sont **non corrélées** si $\mathsf{Cov}(X, Y) = 0$.*

En particulier, si X et Y sont non corrélées alors $\mathsf{Var}(X + Y) = \mathsf{Var}(X) + \mathsf{Var}(Y)$.

Exemple 2.5.10. X_1 et X_2 ci-dessus sont non-corrélées.

Définition 2.5.6. *Soit* $X = (X_1, \cdots, X_d)$ *un vecteur (colonne) aléatoire de* $\mathbb{R}^d$ *tel que* $|X|^2$ *est intégrable. La matrice de variance-covariance de* X *est*

$$\mathsf{Var}(X) = \mathbb{E}[(X - \mathbb{E}[X])(X - \mathbb{E}[X])^t].$$

C'est une matrice symétrique semi-définie positive de taille $d \times d$ *dont la coordonnée* (i, j) *est* $\mathbb{E}[(X_i - \mathbb{E}[X_i])(X_j - \mathbb{E}[X]_j)] = \mathsf{Cov}(X_i, X_j)$ *(*$= \mathsf{Var}(X_i)$ *si* $i = j$*).*

Exemple 2.5.11. La matrice de variance-covariance de la loi normale multidimensionnelle $\mathcal{N}(m, \Sigma)$ est Σ.

On a vu que la fonction de répartition caractérise la loi. L'espérance sur une famille bien choisie de fonction caractérise également la loi.

Proposition 2.5.6. *Soit* X *et* Y *deux variables aléatoires. Alors* $\mathbb{P}_X = \mathbb{P}_Y$ *si et seulement si*

$$\mathbb{E}[\varphi(X)] = \mathbb{E}[\varphi(Y)],$$

pour toute fonction φ *de la forme* $\mathbb{1}_{]-\infty,t]}$ *si ce sont des variables réelles,* $\mathbb{1}_{]-\infty,t_1]\times\cdots\times]-\infty,t_d]}$ *si ce sont des vecteurs aléatoires.*

Démonstration. Comme $F_X(t) = \mathbb{E}[\mathbb{1}_{]-\infty,t]}(X)]$, c'est équivalent à dire que X et Y ont la même fonction de répartition, et on sait qu'elle caractérise la loi. □

On peut généraliser directement cet énoncé à toute famille de fonctions mesurables qui contient les $\mathbb{1}_{]-\infty,t]}$: les fonctions indicatrices d'intervalles, les fonctions indicatrices de boréliens, les fonctions mesurables bornées. C'est une conséquence directe du théorème d'unicité des mesure.

Rappel 3. [Théorème de caractérisation des mesures sur un π-système] Soit μ_1 et μ_2 deux mesures positives sur l'espace probabilisable $(\Omega, \mathcal{F})$ telles que $\mu_1(A) = \mu_2(A)$ pour tout $A \in \mathcal{C}$ où $\mathcal{C}$ est une famille de parties de Ω stable par intersection finie (i.e. un π-système) qui engendre $\mathcal{F}$. Si $\mu_1(\Omega) = \mu_2(\Omega) < \infty$, alors $\mu_1 = \mu_2$.
On peut également généraliser cet énoncé à toute famille de fonctions qui permet d'approcher les $\mathbb{1}_{]-\infty,t]}$ (via le théorème de convergence dominée ou le théorème de convergence monotone), par exemple, les fonctions continues bornées, les fonctions C^∞ bornées ou les fonctions C^∞ à support compact.

Exemple 2.5.12. Soit X une variable aléatoire de loi normale centrée réduite. Alors $Y = \sigma X + m$ suit la loi $\mathcal{N}(m, \sigma^2)$. Réciproquement, si Y est une variable aléatoire de loi normale $\mathcal{N}(m, \sigma^2)$, alors $X = (Y - m)/\sigma$ suit la loi $\mathcal{N}(0, 1)$. En effet, pour toute fonction φ mesurable bornée (par exemple), on obtient par le changement de variable $y = \sigma x + m$

$$\begin{aligned}
\mathbb{E}[\varphi(\sigma X + m)] &= \frac{1}{(2\pi)^{1/2}} \int_{-\infty}^{+\infty} \varphi(\sigma x + m) \exp(-x^2/2) dx \\
&= \frac{1}{(2\pi)^{1/2}} \int_{-\infty}^{+\infty} \varphi(y) \exp(-((y - m)/\sigma)^2/2) \frac{dy}{\sigma} \\
&= \frac{1}{\sigma(2\pi)^{1/2}} \int_{-\infty}^{+\infty} \varphi(y) \exp(-(y - m)^2/(2\sigma^2)) dy
\end{aligned}$$

$$= \int_{-\infty}^{+\infty} \varphi(y) f_Y(y) dy$$
$$= \mathbb{E}[\varphi(Y)],$$

où on reconnaît que Y suit la loi $\mathcal{N}(m, \sigma^2)$ en identifiant sa densité f_Y. La réciproque est complètement analogue. On peut ainsi très facilement identifier des lois (par leur densité) construites à partir de lois connues.

Une autre conséquence de cette propriété est la caractérisation de l'indépendance par l'espérance.

Corollaire 2.5.1. *Soit X et Y deux variables aléatoires. Alors X et Y sont indépendantes si et seulement si*

$$\mathbb{E}[\varphi(X)\psi(Y)] = \mathbb{E}[\varphi(X)]\mathbb{E}[\psi(Y)],$$

pour toutes fonctions φ et ψ mesurables bornées.

Démonstration. On a vu que deux variables aléatoires sont indépendantes si et seulement si le loi du couple (caractérisée par la première espérance) est égale au produit des probabilités marginales (caractérisées par les dernières espérances). □

Exemple 2.5.13. En particulier, si X et Y réelles sont indépendantes, alors $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$ donc X et Y sont non corrélées. La réciproque est fausse en général. Reprenons le couple (X_1, X_2). On a vu que X_1 et X_2 sont non-corrélées. Montrons qu'elles ne sont pas indépendantes. On choisit $\varphi(x_1) = \mathbb{1}_1(X_1)$ et $\psi(x_2) = \mathbb{1}_1(x_2)$. On a

$$\mathbb{E}[\varphi(X_1)\psi(X_2)] = \mathbb{P}((X_1, X_2) = (1, 1)) = 0,$$

et

$$\mathbb{E}[\varphi(X_1)] = \mathbb{P}(X_1 = 1) = 1/4, \qquad \mathbb{E}[\psi(X_2)] = \mathbb{P}(X_2 = 1) = 1/4.$$

On a $\mathbb{E}[\varphi(X_1)\psi(X_2)] \neq \mathbb{E}[\varphi(X_1)]\mathbb{E}[\psi(X_2)]$, donc X_1 et X_2 ne sont pas indépendantes.

2.6 Fonction caractéristique

Une famille intéressante de fonctions tests est la famille des exponentielles complexes à paramètre.

Définition 2.6.1. *Soit X une variable (ou vecteur) aléatoire. On appelle* ***fonction caractéristique*** *de X (ou de la loi de X) la fonction Φ_X de $\mathbb{R}$ (ou $\mathbb{R}^d$) dans $\mathbb{C}$ qui à u associe*

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

où $\langle u, X\rangle = uX$ dans le cas scalaire et $\langle u, X\rangle = \sum_{i=1}^d u_i X_i$ dans le cas vectoriel (c'est le produit scalaire). C'est une fonction à valeurs complexes, de module majoré par 1 et telle que $\Phi_X(0) = 1$.

La fonction caractéristique est en fait la **transformée de Fourier** de la loi.

Exemple 2.6.1. Fonction caractéristique pour les lois usuelles
Loi de Bernoulli $\mathcal{B}(p)$ **:** $\Phi_X(u) = 1 - p + pe^{iu}$

Loi Uniforme discrète : $\Phi_X(u) = \frac{1}{n} \sum\limits_{k=1}^{n} e^{kiu}$

Loi Binomiale $\mathcal{B}(n, p)$ **:** $\Phi_X(u) = (1 - p + pe^{iu})^n$

Loi Multinomiale $\mathcal{M}(n, m, p_1, \cdots, p_m)$ **:** $\Phi_X(u_1, \cdots, u_m) = \Big(\sum_{i=1}^{m} p_i e^{iu_i}\Big)^n$

Loi Géométrique $\mathcal{G}(p)$ **:** $\Phi_X(u) = \frac{pe^{iu}}{1-(1-p)e^{iu}}$

Loi de Poisson $\mathcal{P}(\lambda)$ **:** $\Phi_X(u) = e^{\lambda(e^{iu}-1)}$

Loi uniforme continue $\mathcal{U}([a, b])$ **:** $\Phi_X(u) = \frac{e^{iub}-e^{iua}}{iu(b-a)}$

Loi exponentielle $\mathcal{E}(\lambda)$ **:** $\Phi_X(u) = (1 - i\frac{u}{\lambda})^{-1}$

Loi normale $\mathcal{N}(m, \sigma)$ **:** $\Phi_X(u) = e^{imu-\sigma^2 u^2/2}$

Loi normale multi-dimensionnelle $\mathcal{N}(m, \Sigma)$ **:** $\Phi_X(u) = e^{i\langle u,m\rangle - \langle u,\Sigma u\rangle/2}$

Pour pouvoir montrer que la fonction caractéristique caractérise la loi d'une variable aléatoire, on a besoin de travailler un peu plus spécifiquement sur la fonction caractéristique des lois gaussiennes.

Proposition 2.6.1. *Soit Y une variable aléatoire de loi $\mathcal{N}(0, I_d)$ et σ un réel strictement positif. Alors la densité de σY peut s'écrire comme*

$$f(x) = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} e^{-i\langle x,u\rangle - \frac{\sigma^2}{2}\|u\|^2} du$$

Démonstration. La fonction caractéristique de Y est

$$\Phi_Y(u) = \mathbb{E}[e^{i\langle u,Y\rangle}] = e^{-\frac{1}{2}\|u\|^2}.$$

On a donc pour toute fonction mesurable bornée h

$$\begin{aligned}
\mathbb{E}[h(\sigma Y)] &= \int_{\mathbb{R}^d} h(\sigma y) \frac{e^{-\|y\|^2/2}}{(2\pi)^{\frac{d}{2}}} dy \\
&= \frac{1}{(2\pi)^{d/2}} \int_{\mathbb{R}^d} h(\sigma y) \Phi_Y(y) dy \\
&= \frac{1}{\sigma^d (2\pi)^{d/2}} \int_{\mathbb{R}^d} h(z) \Phi_Y(\tfrac{z}{\sigma}) dz.
\end{aligned}$$

en faisant le changement de variable $z = \sigma y$. On en déduit que la densité de σY peut s'écrire comme

$$\begin{aligned}
f(z) &= \frac{1}{\sigma^d (2\pi)^{d/2}} \Phi_Y(z/\sigma) \\
&= \frac{1}{\sigma^d (2\pi)^{d/2}} \int_{\mathbb{R}^d} e^{i\langle \frac{z}{\sigma}, v\rangle} \frac{1}{(2\pi)^{d/2}} e^{-\|v\|^2/2} dv \\
&= \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} e^{i\langle z,u\rangle} e^{-\|u\|^2 \sigma^2/2} du
\end{aligned}$$

en faisant le changement de variable $u = v/\sigma$. On fait enfin le changement de variable $y = -u$ pour obtenir

$$f(z) = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} e^{-i\langle z,y\rangle} e^{-\|y\|^2 \sigma^2/2} dy$$

qui est bien la forme voulue. $\square$

Proposition 2.6.2. *Soit X_1 et X_2 deux variables aléatoires. Alors $\mathbb{P}_{X_1} = \mathbb{P}_{X_2}$ si et seulement si $\Phi_{X_1} = \Phi_{X_2}$.*

Démonstration. Si $\mathbb{P}_{X_1} = \mathbb{P}_{X_2}$ alors par définition $\Phi_{X_1} = \Phi_{X_2}$.
Pour démontrer la réciproque de ce résultat, on va introduire une perturbation gaussienne et faire tendre sa variance vers 0. Soit Y une variable aléatoire de loi $\mathcal{N}(0, I_d)$ indépendante de X, et caractérisons la loi de la variable aléatoire $X + \sigma Y$, pour $\sigma > 0$. Pour toute fonction test continue bornée h, on a, en utilisant l'indépendance et la formule précédente pour la densité de σY

$$\begin{aligned}
\mathbb{E}[h(X+\sigma Y)] &= \int_{\mathbb{R}^d \times \mathbb{R}^d} h(x+z) d\mathbb{P}_X(x) d\mathbb{P}_{\sigma Y}(z) \\
&= \int_{\mathbb{R}^d \times \mathbb{R}^d} h(x+z) \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} e^{-i\langle z,u\rangle - \frac{\sigma^2}{2}\|u\|^2} dudzd\mathbb{P}_X(x)
\end{aligned}$$

On fait le changement de variable $v = x + z$ sur la variable z pour obtenir

$$\begin{aligned}
\mathbb{E}[h(X+\sigma Y)] &= \int_{\mathbb{R}^d \times \mathbb{R}^d} h(v) \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} e^{-i\langle v-x,u\rangle - \frac{\sigma^2}{2}\|u\|^2} dudvd\mathbb{P}_X(x) \\
&= \int_{\mathbb{R}^d \times \mathbb{R}^d} h(v) \frac{1}{(2\pi)^d} e^{-i\langle z,u\rangle - \frac{\sigma^2}{2}\|u\|^2} \Big(\int_{\mathbb{R}^d} e^{i\langle x,u\rangle} d\mathbb{P}_X(x) \Big) dudv \\
&= \int_{\mathbb{R}^d \times \mathbb{R}^d} h(v) \frac{1}{(2\pi)^d} e^{-i\langle z,u\rangle - \frac{\sigma^2}{2}\|u\|^2} \Phi_X(u) dudv \\
&= \int_{\mathbb{R}^d} h(v) \Big(\int_{\mathbb{R}^d} \frac{1}{(2\pi)^d} e^{-i\langle z,u\rangle - \frac{\sigma^2}{2}\|u\|^2} \Phi_X(u) du \Big) dv.
\end{aligned}$$

On fait maintenant tendre σ vers 0. Comme h est continue et bornée, le théorème de convergence dominée donne

$$\begin{aligned}
\mathbb{E}[h(X)] &= \mathbb{E}[\lim_{\sigma \to 0} h(X+\sigma Y)] \\
&= \lim_{\sigma \to 0} \mathbb{E}[h(X+\sigma Y)] \\
&= \lim_{\sigma \to 0} \int_{\mathbb{R}^d} h(v) \Big(\int_{\mathbb{R}^d} \frac{1}{(2\pi)^d} e^{-i\langle z,u\rangle - \frac{\sigma^2}{2}\|u\|^2} \Phi_X(u) du \Big) dv
\end{aligned}$$

qui ne dépend de X que par Φ_X. Donc si $\Phi_{X_1} = \Phi_{X_2}$ on a bien $\mathbb{P}_{X_1} = \mathbb{P}_{X_2}$. □

Corollaire 2.6.1. *Soit X et Y deux variables aléatoires. Alors elles sont indépendantes si et seulement si $\Phi_{(X,Y)} = \Phi_X \Phi_Y$*

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

Démonstration. C'est une conséquence directe du fait que X et Y sont indépendantes si et seulement si $\mathbb{P}_{(X,Y)} = \mathbb{P}_X \otimes \mathbb{P}_Y$. □

Proposition 2.6.3. *Soit X une variable aléatoire réelle de fonction caractéristique Φ_X.*

1. *Si $\mathbb{E}[|X|^n] < \infty$, alors Φ_X est n fois dérivable et pour $k \leq n$*

$$\Phi_X^{(k)}(t) = i^k \int x^k e^{itx} d\mathbb{P}_X(x) = i^k \mathbb{E}[X^k e^{itX}].$$

En particulier, $\Phi_X^{(k)}(0) = i^k \mathbb{E}[X^k]$.

> 2. *Réciproquement, si n est pair et Φ_X est n fois dérivable au voisinage de 0, alors X admet des moments de tout ordre inférieur ou égal à n.*

Démonstration. Sens 1. Par récurrence sur n

• Pour $n = 0$, $\Phi_X^{(0)}(t) = \Phi_X(t) = \mathbb{E}[(iX)^0 e^{itX}]$

• Hypothèse de récurrence au rang $n-1$ pour un n fixé

• Hérédité : on suppose que $|X|^n$ est intégrable et on va utiliser le théorème de dérivation sous l'intégrale. Par hypothèse de récurrence, on a

$$\Phi_X^{(n-1)}(t) = i^{n-1}\int x^{n-1}\exp(itx)d\mathbb{P}_X(x) = i^{n-1}\mathbb{E}[X^{n-1}\exp(itX)].$$

Or $\sup_t |\frac{d}{dt} i^{n-1}X^{n-1}\exp(itX)| = \sup_t |\frac{d}{dt} i^n X^n \exp(itX)| \leq |X|^n$ qui est bien intégrable. Donc on peut dériver sous l'intégrale pour obtenir directement

$$\Phi_X^{(n)}(t) = i^n\int x^n\exp(itx)d\mathbb{P}_X(x) = i^n\mathbb{E}[X^n\exp(itX)].$$

Sens 2. On montre par récurrence sur k que $\mathbb{E}[X^{2k}] < \infty$ si $2k \leq n$. Pour $k = 0$, la propriété est vraie. Supposons qu'elle est vraie pour $k-1$ fixé. Par hypothèse, la limite

$$\lim_{h\to 0}\frac{1}{h^2}(\Phi_X^{(2k-2)}(h) + \Phi_X^{(2k-2)}(-h) - 2\Phi_X^{(2k-2)}(0))$$

existe et vaut $\Phi_X^{(2k)}(0)$ (faire un DL à l'ordre 2 de $\Phi_X^{(2k-2)}(h)$ et $\Phi_X^{(2k-2)}(-h)$ et les additionner). Comme, d'après le premier cas, pour tout réel h on a

$$\Phi_X^{2k-2}(h) = (-1)^{k-1}\int x^{2k-2}\exp(ihx)d\mathbb{P}_X(x),$$

on a

$$(-1)^k\Phi_X^{2k}(0) = \lim_{h\to 0}\int x^{2k-2}\frac{\cos(hx)-1}{h^2}d\mathbb{P}_X(x).$$

On utilise ensuite le lemme de Fatou

Rappel 4. [Lemme de Fatou] Soit (f_n) une suite de fonctions mesurables positives et μ une mesure. Alors

$$\int \liminf_{n\to\infty} f_n d\mu \leq \liminf_{n\to\infty}\int f_n d\mu.$$

et $\lim(1 - cos(hx))/h^2 = x^2/2$ pour obtenir

$$\begin{aligned}\int x^{2k}d\mathbb{P}_X(x) &= \int \lim_{h\to 0} x^{2k-2}\frac{\cos(hx)-1}{h^2}d\mathbb{P}_X(x)\\ &\leq \lim_{h\to 0}\int x^{2k-2}\frac{\cos(hx)-1}{h^2}d\mathbb{P}_X(x) = (-1)^k\Phi_X^{2k}(0) < \infty,\end{aligned}$$

qui est le résultat attendu. □

Si Φ_X est analytique (développable en série entière), ce résultat montre que la loi de X est caractérisée par ses moments.

> **Exemple 2.6.2.** La fonction caractéristique $\Phi = (t) = e^{-t^2/2}$ de la loi normale centrée réduite est C^∞ et développable en série entière sur $\mathbb{R}$, donc la loi normale centrée réduite a des moments à tout ordre et
>
> $$\Phi(t) = \sum_{n=0}^{\infty}\frac{(-t^2/2)^n}{n!} = \sum_{n=0}^{\infty}\frac{(it)^{2n}}{2^n n!} = \sum_{n=0}^{\infty}\frac{\Phi^{(n)}(0)t^n}{n!}.$$

On en déduit que $\mathbb{E}[X^n] = 0$ si n est impair et $\mathbb{E}[X^{2n}] = -\Phi^{(2n)}(0) = (2n)!/(2^n n!)$.
On a de même accès à tous les moments de la loi normale $\mathcal{N}(m, \sigma^2)$. En effet, soit Y de loi $\mathcal{N}(m, \sigma^2)$. Alors $X = (Y - m)/\sigma$ est de loi $\mathcal{N}(0, 1)$, autrement dit $Y = \sigma X + m$ a donc des moments à tout ordre. De plus

$$\mathbb{E}[Y^n] = \mathbb{E}[(\sigma X + m)^n] = \sum_{k=0}^{n} \binom{n}{k} \sigma^k \mathbb{E}[X^k] m^{n-k}.$$

D'où

$$\begin{aligned} \mathbb{E}[Y^{2n}] &= \sum_{k=0}^{n} \binom{2n}{2k} \sigma^{2k} m^{2n-2k} \frac{(2k)!}{(2^k k!)} = \sum_{k=0}^{n} \frac{(2n)!}{(2n-2k)!(2^k k!)} \sigma^{2k} m^{2n-2k} \\ \mathbb{E}[Y^{2n+1}] &= \sum_{k=0}^{n} \binom{2n+1}{2k} \sigma^{2k} m^{2n+1-2k} \frac{(2k)!}{(2^k k!)} = \sum_{k=0}^{n} \frac{(2n+1)!}{(2n+1-2k)!(2^k k!)} \sigma^{2k} m^{2n+1-2k}. \end{aligned}$$

On a ainsi tous les moments.

Proposition 2.6.4. *Soit X et Y deux variables aléatoires indépendantes. Alors la loi de la somme $X + Y$ est donnée par le produit de convolution $\mathbb{P}_X * \mathbb{P}_Y$ défini pour toute fonction φ mesurable bornée par*

$$\begin{aligned} \int_{\mathbb{R}} \varphi d(\mathbb{P}_X * \mathbb{P}_Y) &= \int_{\mathbb{R}} \Big(\int_{\mathbb{R}} \varphi(x+y) d\mathbb{P}_Y(y) \Big) d\mathbb{P}_X(x) \\ &= \int_{\mathbb{R}} \Big(\int_{\mathbb{R}} \varphi(x+y) d\mathbb{P}_X(x) \Big) d\mathbb{P}_Y(y). \end{aligned}$$

Démonstration. Soit $U(x, y) = x + y$. On a, par le théorème de transfert et les propriétés de l'indépendance

$$\begin{aligned} \int_{\mathbb{R}} \varphi d\mathbb{P}_{X+Y} &= \int_{\mathbb{R}^2} \varphi(x+y) d\mathbb{P}_X(x) d\mathbb{P}_Y(y) \\ &= \int_{\mathbb{R}} \varphi d(\mathbb{P}_X * \mathbb{P}_Y), \end{aligned}$$

qui est le résultat attendu. □

Corollaire 2.6.2. *Si X et Y sont deux variables aléatoires indépendantes, alors la fonction caractéristique de leur somme est le produit des fonctions caractéristiques de X et Y*

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

Démonstration. Pour tout réel ou vecteur t on a

$$\begin{aligned} \Phi_{X+Y}(t) &= \mathbb{E}[e^{i\langle t, (X+Y) \rangle}] \\ &= \mathbb{E}[e^{i\langle t, X \rangle} e^{i\langle t, Y \rangle}] \\ &= \mathbb{E}[e^{i\langle t, X \rangle}] \mathbb{E}[e^{i\langle t, Y \rangle}] \\ &= \Phi_X(t)\Phi_Y(t), \end{aligned}$$

en utilisant l'indépendance de X et Y. □

Exemple 2.6.3. Soit X et Y indépendantes de loi de Poisson de paramètre respectif λ et μ. Alors $X+Y$ suit la loi de Poisson de paramètre $\lambda+\mu$. En effet, on a

$$\begin{aligned}\Phi_{X+Y}(t) &= \Phi_X(t)\Phi_Y(t) \\ &= e^{\lambda(e^{it}-1)}e^{\mu(e^{it}-1)} \\ &= e^{(\lambda+\mu)(e^{it}-1)},\end{aligned}$$

et on reconnaît la fonction caractéristique de la loi de Poisson.

La fonction génératrice des moments ou transformée de Laplace est une variante de la fonction génératrice qui a des propriétés analogues mais requiert des conditions d'intégralité. On l'utilise en général pour des lois à support dans $\mathbb{R}_+$. Elle a les mêmes propriétés que la fonction caractéristique.

Définition 2.6.2. *Soit X une variable (ou vecteur) aléatoire. On appelle* ***transformée de Laplace*** *(ou fonction génératrice des moments) de X (ou de la loi de X) la fonction $L_X(s) = \mathbb{E}[e^{\langle s,X\rangle}]$ définie pour les valeurs de s pour lesquelles $e^{\langle s,X\rangle}$ est intégrable.*

Chapitre 3

Loi des grands nombres

Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace de probabilité sur lequel sont définies toutes les variables aléatoires.

3.1 Loi du $0 - 1$

Proposition 3.1.1 Inégalité de Markov. *Soit X une variable aléatoire réelle intégrable et $t > 0$, alors*

$$\mathbb{P}(X \geq t) \leq \frac{\mathbb{E}[X^+]}{t} \leq \frac{\mathbb{E}[|X|]}{t}.$$

Démonstration. On a $\mathbb{1}_{[t,\infty[}(X) \leq (X/t)\mathbb{1}_{[t,\infty[}(X) \leq X^+/t \leq |X|/t$ et on intègre par rapport à $\mathbb{P}$. $\square$

Exemple 3.1.1. Soit X une variable aléatoire de loi exponentielle de paramètre λ. Alors l'inégalité de Markov donne $e^{-\lambda t} = \mathbb{P}(X \geq t) \leq \frac{\mathbb{E}[X]}{t} \leq \frac{1}{\lambda t}$.

Corollaire 3.1.1 Inégalité de Tchébychev. *Si X^p est intégrable, $p > 0$ alors*

$$\mathbb{P}(|X| \geq t) \leq \frac{\mathbb{E}[|X|^p]}{t^p}.$$

Si X est de carré intégrable alors

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

Démonstration. Utiliser l'inégalité de Markov pour $|X|^p$ car $(X \geq t) \subset (|X|^p \geq t^p)$ pour $t > 0$. Puis utiliser le cas particulier $p = 2$ et reconnaître $\mathsf{Var}(X) = \mathbb{E}[(X - \mathbb{E}[X])^2]$. $\square$

Exemple 3.1.2. Soit (X_n) une suite de variables aléatoires indépendantes de même loi de Bernoulli de paramètre p. On pose $S_n = \sum_{k=1}^n X_k$. S_n suit la loi binomiale de paramètres (n, p). Alors, pour tout $\varepsilon > 0$ on a

$$\mathbb{P}(|S_n/n - p| \geq \varepsilon) \leq \frac{\mathsf{Var}(S_n/n)}{\varepsilon^2} = \frac{p(1-p)}{n\varepsilon^2}.$$

Le ratio S_n/n correspond à la fréquence empirique des succès tandis que le paramètre p correspond à la probabilité théorique d'apparition d'un succès. D'après l'inégalité précédente, la probabilité que la fréquence empirique s'éloigne de la probabilité théorique tend vers 0.

Remarquons que l'inégalité reste valable pour n'importe quelle loi de carré intégrable en remplaçant p par son espérance et $p(1-p)$ par sa variance.

Définition 3.1.1. *Soit $(\mathcal{F}_n)_{n\in\mathbb{N}}$ une famille indépendante de tribus sur $(\Omega, \mathcal{F}, \mathbb{P})$. Soit $\overline{\mathcal{F}}_n$ la tribu engendrée par $(\mathcal{F}_k, k \geq n)$ et $\overline{\mathcal{F}}_\infty = \bigcap_{n\in\mathbb{N}} \overline{\mathcal{F}}_n$. La tribu $\overline{\mathcal{F}}_\infty$ s'appelle tribu des événements terminaux ou tribu terminale de la suite $(\mathcal{F}_n)_{n\in\mathbb{N}}$.*

Exemple 3.1.3. Par exemple, $\mathcal{F}_n = \sigma(X_n)$ pour une suite (X_n) de variables aléatoires indépendantes. La tribu terminale correspond aux événement qu'on peut définir à partir de la suite (X_n) mais qui ne dépendent pas des premiers termes (la limite par exemple).

La tribu terminale vérifie la loi du tout ou rien ou loi du $0-1$.

Proposition 3.1.2 Loi du $0-1$. *Si $\overline{\mathcal{F}}_\infty$ est une tribu terminale, alors pour tout $A \in \overline{\mathcal{F}}_\infty$, $\mathbb{P}(A) = 0$ ou $\mathbb{P}(A) = 1$.*

Démonstration. Soit $A \in \overline{\mathcal{F}}_\infty$ fixé. On considère l'ensemble des événements indépendants de A

$$\mathcal{C} = \{B \in \mathcal{F}; \mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)\}.$$

On veut montrer que $\overline{\mathcal{F}}_\infty \subset \mathcal{C}$. Alors A est indépendant de lui-même donc sa probabilité est 0 ou 1.
Soit les tribus $\underline{\mathcal{F}}_n = \sigma(\mathcal{F}_k, k \leq n)$. Les tribus $\underline{\mathcal{F}}_n$ et $\overline{\mathcal{F}}_{n+1}$ sont indépendantes, donc tout élément de $\underline{\mathcal{F}}_n$ est indépendant de A pour tout n car $\overline{\mathcal{F}}_\infty \subset \overline{\mathcal{F}}_{n+1}$. Ainsi, $\underline{\mathcal{F}}_\infty = \cup \underline{\mathcal{F}}_n \subset \mathcal{C}$. Or, $\Omega \in \underline{\mathcal{F}}_\infty$, $\underline{\mathcal{F}}_\infty$ est stable par complémentaire et union finie (car union croissante de tribus). Le théorème π-δ implique $\sigma(\underline{\mathcal{F}}_\infty) \subset \mathcal{C}$. De plus, pour tout k, on a $\mathcal{F}_k \subset \underline{\mathcal{F}}_k \subset \underline{\mathcal{F}}_\infty \subset \sigma(\underline{\mathcal{F}}_\infty)$, donc $\overline{\mathcal{F}}_n \subset \sigma(\underline{\mathcal{F}}_\infty)$ pour tout n et $\overline{\mathcal{F}}_\infty \subset \sigma(\underline{\mathcal{F}}_\infty) \subset \mathcal{C}$, ce qui prouve le résultat. □

Exemple 3.1.4. On dispose d'une infinité de pièces qu'on lance à tour de rôle. La n-ème pièce a une probabilité p_n de donner Pile. On suppose que les lancers sont indépendants. Soit A_n l'événement "*le lancer de la n-ème pièce a donné pile*". Alors l'événement

$$A = \bigcap_{n\in\mathbb{N}} \bigcup_{m\geq n} A_m = (A_n \text{ est réalisé pour une infinité de } n) = \limsup A_n$$

est un événement terminal pour la suite de tribus $\mathcal{F}_n = \sigma(A_n)$, donc $\mathbb{P}(A) = 0$ ou 1. L'événement A s'appelle aussi "*A_n a lieu infiniment souvent*", noté (A_n is). Ici il correspond à "*obtenir une infinité de pile*" (d'où son nom). Dire que $\mathbb{P}(A) = 0$ signifie que pour presque tout $\omega \in \Omega$, il existe $n(\omega)$ fini tel que pour tout $n \geq n(\omega)$, $\omega \notin A_n$, i.e. A_n n'a pas lieu. Le résultat suivant permet de savoir si on est dans le cas $\mathbb{P}(A) = 0$ ou $\mathbb{P}(A) = 1$.

Théorème 3.1.1 Lemme de Borel Cantelli. *Soit $(A_n)_{n\in\mathbb{N}}$ une suite d'événements.*
- *Si $\sum \mathbb{P}(A_n) < \infty$ alors $\mathbb{P}(A_n \text{ is}) = 0$.*
- *Si les événements $(A_n)_{n\in\mathbb{N}}$ sont indépendants et $\sum \mathbb{P}(A_n) = \infty$ alors $\mathbb{P}(A_n \text{ is}) = 1$.*

Démonstration. 1. Pour tout n, on a

$$A = \bigcap_{n\in\mathbb{N}} \bigcup_{m\geq n} A_m = (A_n \text{ is}) \subset \bigcup_{m\geq n} A_m,$$

donc $\mathbb{P}(A_n \text{ is}) = \mathbb{P}(A) \leq \sum_{m\geq n} \mathbb{P}(A_m)$ car intersection décroissante, et ce terme tend vers 0 avec n quand la série converge.

2. Par indépendance

$$\mathbb{P}(\bigcup_{m=n}^{N} A_m) = 1 - \mathbb{P}(\bigcap_{m=n}^{N} A_m^c)$$
$$= 1 - \prod_{m=n}^{N} (1 - \mathbb{P}(A_m)).$$

Comme $1 - x \leq \exp(-x)$ pour tout $x \geq 0$, on a

$$\mathbb{P}(\bigcup_{m=n}^{N} A_m) \geq 1 - \prod_{m=n}^{N} \exp(-\mathbb{P}(A_m)) = 1 - \exp(-\sum_{m=n}^{N} \mathbb{P}(A_m)).$$

Quand N tend vers l'infini, la série dans le terme de droite tend vers l'infini donc $\mathbb{P}(\bigcup_{m=n}^{N} A_m)$ tend vers 1, et on passe à la limite en n encore par intersection décroissante.

□

Exemple 3.1.5. On revient sur les pièces. Si $\sum p_n < \infty$ (par exemple, $p_n = 1/n^2$) alors $\mathbb{P}(A) = 0$ on aura ps un nombre fini de Pile ; si $\sum p_n = \infty$ (par exemple, $p_n = 1/n$, $p_n = p > 0$ fixé,...) alors $\mathbb{P}(A) = 1$, on aura ps une infinité de Pile.

Exemple 3.1.6 Le singe dactylo. Un singe tape au hasard sur un clavier d'ordinateur pendant un temps infini. Quelle est la probabilité que son texte contienne les oeuvres complètes de Molière ?
La suite de lettres (et espaces, ponctuation...) tapées par le singe est une réalisation d'une suite de variables aléatoires (X_n) indépendantes et de même loi (uniforme sur le clavier). Soit $(a_1, \cdots, a_N)$ la chaîne de caractères (finie) correspondent aux oeuvres complètes de Molière. On pose $B_k = (X_k = a_1, \cdots, X_{k+N-1} = a_N)$ pout $k \geq 0$ et $A_n = B_{(n-1)N+1}$, $n \geq 1$, autrement dit $A_1 = B_1$, $A_2 = B_{N+1}$, $A_3 = B_{2N+1}, \ldots$ de sorte que les événements (A_n) sont indépendants. Chaque événement A_n est de probabilité nb touches$^{-N} > 0$ donc $\sum \mathbb{P}(A_n) = \infty$. Le lemme de Borel Cantelli dit que $\mathbb{P}(A_n \text{ is}) = 1$ donc non seulement le singe tapera tout Molière, mais en plus il le fera une infinité de fois avec probabilité 1.

3.2 Convergence presque sure et en probabilité

Il existe différentes notions de convergence pour les suites de variables aléatoires. On va en définir quelques unes et explorer leur liens. Dans toute la suite du chapitre les variables aléatoires $(X_n)_{n \in \mathbb{N}^*}$ sont définies de $(\Omega, \mathcal{F}, \mathbb{P})$ dans $\mathbb{R}$ ou $\mathbb{R}^d$.

Définition 3.2.1. *On dit que la suite (X_n)* ***converge presque sûrement (ps)*** *vers la variable aléatoire X, noté $X_n \xrightarrow{ps} X$, si*

$$\mathbb{P}\Big(\omega \in \Omega : \lim_{n \to \infty} X_n(\omega) = X(\omega)\Big) = 1.$$

Autrement dit, il existe $A \subset \Omega$ de mesure pleine tel que pour tout $\omega \in A$, pour tout $\varepsilon > 0$, il existe $n_0(\omega)$ tel que pour tout $n \geq n_0$, $|X_n(\omega) - X(\omega)| \leq \varepsilon$.

Exemple 3.2.1. Soit $\Omega = [0, 1]$ muni de la tribu des boréliens et de la mesure de Lebesgue. Pour $n \geq 1$, on pose

$$X_n(\omega) = \mathbb{1}_{[0,1/n]}(\omega).$$

Alors la suite (X_n) converge ps vers 0 (même si $X_n(0) = 1$ pour tout n).

Définition 3.2.2. *On dit que la suite (X_n)* ***converge en probabilité*** *vers la variable aléatoire X, noté $X_n \xrightarrow{\mathbb{P}} X$, si pour tout $\varepsilon > 0$*

$$\lim_{n\to\infty} \mathbb{P}(|X_n - X| \geq \varepsilon) = 0.$$

Autrement dit, pour tout $\varepsilon, \eta > 0$, il existe n_0 tel que pour tout $n \geq n_0$, $\mathbb{P}(|X_n - X| \geq \varepsilon) \leq \eta$.

Exemple 3.2.2. Soit (X_n) une suite de variables aléatoires indépendantes de même loi de Bernoulli de paramètre p. On pose $S_n = \sum_{k=1}^n X_k$ qui suit donc une loi binomiale de paramètres (n, p). Alors, pour tout $\varepsilon > 0$ on a, par l'inégalité de Tchébychev

$$\mathbb{P}(|S_n/n - p| \geq \varepsilon) \leq \frac{\mathsf{Var}(S_n/n)}{\varepsilon^2} = \frac{p(1-p)}{n\varepsilon^2} \xrightarrow[n\to+\infty]{} 0.$$

Donc la fréquence S_n/n des succès converge en probabilité vers la probabilité de succès p lorsque n tend vers l'infini.

Proposition 3.2.1. *Si la suite (X_n) converge presque sûrement vers la variable aléatoire X alors la (X_n) converge en probabilité vers X.*

Démonstration. □

Contre-Exemple 3.2.1. La convergence ps entraîne la convergence en probabilité, mais la réciproque est fausse en général. Soit $\Omega = [0, 1]$ muni de la tribu des boréliens et de la mesure de Lebesgue. On définit une suite de variables aléatoires $(X_i)_{i\geq 1}$ par $X_i(\omega) = \mathbb{1}_{](k-1)/2^n, k/2^n]}(\omega)$ où $n = n(i) = \min\{m; i + 1 \leq 2^{m+1}\}$ et $k = k(i) = i + 1 - 2^n$ (la numérotation est en fait reliée au développement en base 2 de i). En particulier, on a

— $i = 1$, $n = 0$, $k = 1$ donc $X_1(\omega) = \mathbb{1}_{]0,1]}$
— $i = 2$, $n = 1$, $k = 1$ donc $X_2(\omega) = \mathbb{1}_{]0,1/2]}$
— $i = 3$, $n = 1$, $k = 2$ donc $X_3(\omega) = \mathbb{1}_{]1/2,1]}$, et ainsi de suite.

Rappel 5. La $\liminf$ d'une suite réelle est définie par

$$\liminf_{n\to\infty} x_n = \lim_{m\to\infty} \inf_{n\geq m} x_n.$$

C'est la plus petite limite d'une sous-suite convergente extraite de (x_n). De plus, une suite réelle est convergente si et seulement si elle a une unique valeur d'adhérence donc si et seulement si $\liminf = \limsup$.
Alors pour tout $\omega \in]0, 1]$, $\liminf X_i(\omega) = 0$ et $\limsup X_i(\omega) = 1$, donc la suite (X_n) ne converge pas ps. Or pour tout $0 < \varepsilon < 1$, $\mathbb{P}(|X_i| \geq \varepsilon) = \mathbb{P}(|X_i| = 1) = 2^{-n}$ si $i = 2^n + k - 1$ avec $1 \leq k \leq 2^n$. Donc la suite (X_i) converge en probabilité vers 0.

Proposition 3.2.2. *Soit (X_n) une suite de variables aléatoires convergeant ps vers X et h une fonction continue sur $\mathbb{R}$. Alors $h(X_n)$ converge presque sûrement vers $h(X)$. En particulier, si (X_n) et (Y_n) sont deux suites de variables aléatoires convergeant presque sûrement vers X et Y respectivement, alors pour tous réels a et b la suite $(aX_n + bY_n)$ converge presque sûrement vers $aX + bY$ et la suite (X_nY_n) converge presque sûrement vers XY.*

Démonstration. La convergence presque sûre correspond à la convergence simple des fonctions de Ω dans $\mathbb{R}$ (à un sous-ensemble de mesure nulle près). Les propriétés de la convergence simple sont donc également vraies pour la convergence presque sûre (en remarquant que l'union de deux ensembles de probabilité nulle est de probabilité nulle). □

Proposition 3.2.3. *Soit (X_n) et (Y_n) deux suites de variables aléatoires convergeant en probabilité vers X et Y respectivement, alors*

1. *pour toute fonction continue h, on a $h(X_n)$ converge en probabilité vers $h(X)$,*
2. *tous réels a et b la suite $(aX_n + bY_n)$ converge en probabilité vers $aX + bY$,*
3. *(X_nY_n) converge en probabilité vers XY*

Démonstration. Admis. □

3.3 Convergence dans L^p

On rappelle la définition générale des espaces L^p vue en cours de théorie de la mesure.

Définition 3.3.1. *Une variable aléatoire X est dans $L^p(\Omega, \mathcal{F}, \mathbb{P})$, $p > 0$, si $\mathbb{E}[|X|^p] < \infty$. L'espace $L^p(\Omega, \mathcal{F}, \mathbb{P})$ muni de la norme*

$$\|X\|_p = (\mathbb{E}[|X|^p])^{1/p}$$

est un espace complet.

On peut définir une notion de convergence associée à la norme L^p.

Définition 3.3.2. *Soit $0 < p < \infty$. On dit que (X_n)* ***converge vers*** *X* ***dans*** *L^p si pour tout n, $X_n \in L^p(\Omega, \mathcal{F}, \mathbb{P})$, $X \in L^p(\Omega, \mathcal{F}, \mathbb{P})$ et $\lim \|X_n - X\|_p = 0$.*

Exemple 3.3.1. Soit $\Omega =]0, 1]$ muni de la tribu des boréliens et de la mesure de Lebesgue. Soit $\alpha > 0$ et pour $n \geq 1$

$$X_n(\omega) = \omega^{-\alpha} \mathbb{1}_{]0,1/n]}(\omega).$$

On a que $X_n \in L^p$ dès que $\alpha p < 1$ puisque

$$\mathbb{E}[X_n^p] = \int_0^{1/n} \omega^{-\alpha p} d\omega = [\frac{\omega^{1-\alpha p}}{1 - \alpha p}]_0^{1/n} = \frac{1}{1 - \alpha p} n^{\alpha p - 1} \xrightarrow[n \to \infty]{} 0,$$

Donc la suite (X_n) converge vers 0 dans L^p dès que $\alpha p < 1$.

Proposition 3.3.1. *La convergence dans L^p implique la convergence en probabilité.*

Démonstration. C'est une conséquence directe de l'inégalité de Markov

$$\mathbb{P}(|X_n - X| \geq \varepsilon) = \mathbb{P}(|X_n - X|^p \geq \varepsilon^p) \leq \mathbb{E}[|X_n - X|^p]/\varepsilon^p = \|X_n - X\|_p^p/\varepsilon^p \to 0,$$

ce qui prouve le résultat. □

Contre-Exemple 3.3.1. La réciproque est fausse : en général, ni la convergence ps, ni la convergence en probabilité n'entrainent la convergence dans L^p.
Soit $\Omega =]0, 1]$ muni de la tribu des boréliens et de la mesure de Lebesgue. Soit $\alpha > 0$ et pour $n \geq 1$

$$X_n(\omega) = \omega^{-\alpha} \mathbb{1}_{]0,1/n]}(\omega).$$

Pour tout $\varepsilon \in]0,1[$, $\mathbb{P}(|X_n| \geq \varepsilon) = 1/n$. Donc la suite (X_n) converge en probabilité vers 0. Mais $X_n \notin L^p$ dès que $\alpha p \geq 1$ puisque

$$\mathbb{E}[X_n^p] = \int_0^{1/n} \omega^{-\alpha p} d\omega = +\infty.$$

La suite n'étant pas dans L^p ne peut pas converger dans L^p.
Soit maintenant $\Omega = \mathbb{R}$ muni de la tribu des boréliens. Pour tout $n \geq 1$, soit X_n une variable aléatoire de loi $(1 - n^{-p})\delta_0 + n^{-p}\delta_n$, ie

$$\mathbb{P}(X_n = n) = n^{-p} = 1 - \mathbb{P}(X_n = 0).$$

Soit $\varepsilon > 0$. Pour tout $n \geq 1$, $\mathbb{P}(|X_n| \geq \varepsilon) = n^{-p}$, donc (X_n) converge en probabilité vers 0. De plus, $\sum \mathbb{P}(|X_n| \geq \varepsilon) < +\infty$ si $p > 1$, donc par le lemme de Borel-Cantelli $\mathbb{P}(|X_n| \geq \varepsilon\ is) = 1$, autrement dit en passant au complémentaire ps pour tout $\varepsilon > 0$, il existe N tel que pour tout $n \geq N$, $|X_n| < \varepsilon$ donc on a aussi la convergence ps vers 0. Mais $\mathbb{E}[X_n^p] = n^{-1}n = 1$. La suite ne converge donc pas vers 0 dans L^p.

Pour passer de la convergence en probabilité à la convergence dans L^p, on introduit la notion d'équi-intégrabilité, ou intégrabilité uniforme.

Définition 3.3.3. *Une famille quelconque $(X_i)_{i\in I}$ de variables aléatoires intégrables est **équi-intégrable** ou **uniformément intégrable** si*

$$\lim_{c\to+\infty} \sup_{i\in I} \int_{(|X_i|>c)} |X_i| d\mathbb{P} = 0.$$

Proposition 3.3.2. *La famille $(X_i)_{i\in I}$ de variables aléatoires intégrables est équi-intégrable si et seulement si les deux conditions suivantes sont réalisées*

1. *pour tout $\varepsilon > 0$, il existe $\eta > 0$ tel que pour tout $A \in \mathcal{F}$, $\mathbb{P}(A) \leq \eta$ implique $\int_A |X_i| d\mathbb{P} \leq \varepsilon$ pour tout $i \in I$,*
2. *$\sup_{i\in I} \mathbb{E}[|X_i|] < \infty$.*

Démonstration. Supposons que la famille est équi-intégrable. Pour tout $\varepsilon > 0$, il existe $c > 0$ tel que

$$\sup_{i\in I} \int_{(|X_i|>c)} |X_i| d\mathbb{P} \leq \varepsilon/2.$$

Soit $A \in \mathcal{F}$. Alors, pour tout i on a

$$\int_A |X_i| d\mathbb{P} \leq \int_{A\cap(|X_i|>c)} |X_i| d\mathbb{P} + \int_{A\cap(|X_i|\leq c)} |X_i| d\mathbb{P} \leq \varepsilon/2 + c\mathbb{P}(A).$$

On a donc le premier point pour $\eta = \varepsilon/2c$ et le deuxième en prenant $A = \Omega$.
Réciproquement, si on a les deux points, soit $M = \sup_{i\in I} \mathbb{E}[|X_i|]$, ε et η donnés par le premier point. On pose $c_0 = M/\eta$. Alors pour tout $c \geq c_0$ et pour tout i, l'inégalité de Markov implique $\mathbb{P}(|X_i| > c) \leq M/c \leq \eta$. On applique donc le premier point à $A = (|X_i| > c)$ pour chaque i et on obtient $\sup_{i\in I} \int_{(|X_i|>c)} |X_i| d\mathbb{P} \leq \varepsilon$. D'où la limite. □

Théorème 3.3.1. *Soit (X_n) une suite de variables aléatoires intégrables et X une variable aléatoire. Alors les deux assertions suivantes sont équivalentes*

1. *$X_n \xrightarrow{\mathbb{P}} X$ et la famille (X_n) est équi-intégrable,*
2. *X est intégrable et (X_n) converge vers X dans L^1.*

Démonstration. 1 $\Rightarrow$ 2 Comme $X_n \xrightarrow{\mathbb{P}} X$, on peut en extraire une sous-suite (X_{n_k}) qui converge ps vers X.

Rappel 6. [Lemme de Fatou]
Soit (f_n) une suite de fonctions mesurables positives et μ une mesure. Alors

$$\int \liminf_{n\to\infty} f_n d\mu \leq \liminf_{n\to\infty} \int f_n d\mu.$$

Le lemme de Fatou et la propriété d'équi-intégrabilité donnent

$$\mathbb{E}[|X|] = \mathbb{E}[\liminf |X_{n_k}|] \leq \liminf \mathbb{E}[|X_{n_k}|] \leq \sup \mathbb{E}[|X_n|] < \infty.$$

Donc X est intégrable. De plus, pour tout $\varepsilon > 0$, on a

$$\begin{aligned}\mathbb{E}[|X_n - X|] &= \int_{(|X_n - X| < \varepsilon/3)} |X_n - X| d\mathbb{P} + \int_{(|X_n - X| \geq \varepsilon/3)} |X_n - X| d\mathbb{P}\\ &\leq \varepsilon/3 + \int_{(|X_n - X| \geq \varepsilon/3)} |X_n| d\mathbb{P} + \int_{(|X_n - X| \geq \varepsilon/3)} |X| d\mathbb{P}.\end{aligned}$$

Comme X est intégrable, la famille (X_n, X) est encore équi-intégrable. On applique la proposition précédente. Soit $\eta > 0$. Pour n assez grand, on a $\mathbb{P}(|X_n - X| \geq \varepsilon/3) \leq \eta$ car la suite converge en probabilité. Donc pour tout n assez grand, les deux intégrales ci-dessus sont inférieures à $\varepsilon/3$. On obtient $\mathbb{E}[|X_n - X|] \leq \varepsilon$ pour n assez grand, d'où la convergence de (X_n) vers X dans L^1.
Réciproque 2 $\Rightarrow$ 1. Soit $\varepsilon > 0$ et n_0 tel que pour tout $n \geq n_0$, $\|X_n - X\|_1 \leq \varepsilon/2$. Comme X et les X_n sont dans L^1, la famille finie $(X, X_n, n \leq n_0)$ est équi-intégrable. Donc il existe η tel que si $\mathbb{P}(A) \leq \eta$ alors

$$\int_A |X_n| d\mathbb{P} \leq \varepsilon/2, \quad \int_A |X| d\mathbb{P} \leq \varepsilon/2$$

pour $n \leq n_0$. Pour $n > n_0$, on a par inégalité triangulaire

$$\int_A |X_n| d\mathbb{P} \leq \int_A |X| d\mathbb{P} + \|X_n - X\|_1 \leq \varepsilon.$$

La suite (X_n) vérifie donc les deux points de la proposition (pour le deuxième, $\mathbb{E}[|X_n|] \leq \mathbb{E}[|X_n - X|] + \mathbb{E}[|X|]$), elle est uniformément intégrable. Et on a déjà vu que convergence L^1 implique convergence en proba. $\square$

3.4 Loi des grands nombres

Dans toute la suite du chapitre les variables aléatoires $(X_n)_{n\in\mathbb{N}^*}$ et X sont définies de $(\Omega, \mathcal{F}, \mathbb{P})$ dans $\mathbb{R}$ ou $\mathbb{R}^d$. On suppose de plus que les variables X_n sont indépendantes et de même loi que X. Pour tout $n \geq 1$, on pose $S_n = \sum_{k=1}^n X_k$. On s'intéresse maintenant aux propriétés asymptotiques de la suite (S_n).

> **Théorème 3.4.1 Loi forte des grands nombres.** *Si $\mathbb{E}[|X|] < \infty$, alors (S_n/n) converge ps et dans L^1 vers $\mathbb{E}[X]$ lorsque n tend vers l'infini.*

Démonstration. Remarquons d'abord que la famille (X_n) est équi-intégrable. En effet, elle est dans L^1 et

$$\lim_{c\to+\infty} \sup_n \mathbb{E}[\mathbb{1}_{(|X_n|>c)}|X_n|] = \lim_{c\to+\infty} \sup_n \mathbb{E}[\mathbb{1}_{(|X_1|>c)}|X_1|] = \lim_{c\to+\infty} \mathbb{E}[\mathbb{1}_{(|X_1|>c)}|X_1|] = 0$$

puisque toutes les variables ont même loi. Donc pour tout $\varepsilon > 0$, il existe $\eta > 0$ tel que pour tout $A \in \mathcal{F}$ et $n \in \mathbb{N}$, on ait $\mathbb{E}[\mathbb{1}_A|X_n|] \leq \varepsilon$. On en déduit que

$$\mathbb{E}[\mathbb{1}_A|S_n/n|] \leq n\mathbb{E}[\mathbb{1}_A|X_n|/n] = \mathbb{E}[\mathbb{1}_A|X_n|] \leq \varepsilon.$$

De plus, $\mathbb{E}[|S_n|/n] \leq n\mathbb{E}[|X_1|]/n \leq \mathbb{E}[|X_1|]$ est borné indépendamment de n. Donc la famille S_n/n est également équi-intégrable. Ainsi, si on montre la convergence ps, on aura la convergence en proba puis la convergence dans L^1 par équi-intégrabilité.

Quitte à remplacer X_k par $X_k - \mathbb{E}[X_k]$, on peut considérer que les X_k sont centrées. Quitte également à regarder la convergence coordonnée par coordonnée (ce qui est équivalent à la convergence du vecteur pour la convergence ps) on suppose qu'on est en dimension 1.

1. La première étape de la preuve consiste à prouver le résultat sous l'hypothèse plus forte $\mathbb{E}[|X|^4] < \infty$. Dans ce cas, pour tout $n \geq 1$ et $\delta > 0$, l'inégalité de Markov donne

$$\mathbb{P}(|S_n| \geq n\delta) \leq \frac{\mathbb{E}[S_n^4]}{\delta^4 n^4}.$$

   On décompose maintenant $\mathbb{E}[S_n^4]$. On a

$$\begin{aligned} S_n^4 &= \Big(\sum_{k=1}^n X_k\Big)^4 \\ &= \sum_{i=1}^n X_i \sum_{j=1}^n X_j \sum_{k=1}^n X_k \sum_{\ell=1}^n X_\ell \\ &= \sum_{i=1}^n X_i^4 + 4 \sum_{1\leq i\neq j\leq n} X_i^3 X_j + 3 \sum_{1\leq i\neq j\leq n} X_i^2 X_j^2 \\ &\quad + 6 \sum_{1\leq i,j,k \text{ distincts}\leq n} X_i X_j X_k^2 + \sum_{1\leq i,j,k,\ell \text{ distincts}\leq n} X_i X_j X_k X_\ell. \end{aligned}$$

   En utilisant la linéarité de l'espérance, et le fait que les X_i sont indépendants, centrés et de même loi, on obtient

$$\begin{aligned} \mathbb{E}[S_n^4] &= \sum_{i=1}^n \mathbb{E}[X_i^4] + 4 \sum_{1\leq i\neq j\leq n} \mathbb{E}[X_i^3]\mathbb{E}[X_j] + 3 \sum_{1\leq i\neq j\leq n} \mathbb{E}[X_i^2]\mathbb{E}[X_j^2] \\ &\quad + 6 \sum_{1\leq i,j,k \text{ distincts}\leq n} \mathbb{E}[X_i]\mathbb{E}[X_j]\mathbb{E}[X_k^2] + \sum_{1\leq i,j,k,\ell \text{ distincts}\leq n} \mathbb{E}[X_i]\mathbb{E}[X_j]\mathbb{E}[X_k]\mathbb{E}[X_\ell] \\ &= n\mathbb{E}[X^4] + 0 + 3n(n-1)\mathbb{E}[X^2]^2 + 0 + 0. \end{aligned}$$

   On obtient donc

$$\mathbb{P}(|S_n| \geq n\delta) \leq \frac{n\mathbb{E}[X^4] + 3n(n-1)\mathbb{E}[X^2]^2}{\delta^4 n^4},$$

   qui est le terme générique d'une série convergente. Le lemme de Borel Cantelli donne donc la convergence ps de (S_n/n) vers 0.

2. Deuxième étape : cas général. Soit $\varepsilon > 0$. Pour tout $i \geq 1$, il existe des variables aléatoires Y_i étagées, centrées, indépendantes et de même loi telles que $\mathbb{E}[|X_i - Y_i|] \leq \varepsilon$, par définition de l'intégrale de Lebesgue. Soit $T_n = \sum_{k=1}^n Y_k$. Alors on a

$$\frac{1}{n}|S_n| \leq \frac{1}{n}\sum_{i=1}^n |X_i - Y_i| + \frac{1}{n}|T_n|.$$

Les variables Y_i sont étagées donc bornées, elles vérifient donc le premier point. On a ainsi T_n/n qui tend vers 0. Il suffit donc de regarder la moyenne des différences $|X_i - Y_i|$. Soit $Z_i = |X_i - Y_i|$. On sait que le variables (Z_i) sont indépendantes, de même loi, intégrables, positives et vérifient $\mathbb{E}[Z_i] \leq \varepsilon$ pour tout i. On veut examiner $\limsup \sum_{k=1}^n Z_k/n$. On utilise un argument dit de bloc : on va découper selon les valeurs de n avec une partition en puissances de 2 et minorer différemment sur chaque bloc. Soit $k \geq 0$ et $\delta > 0$. On a

$$\begin{aligned}
&\mathbb{P}\Big(\max_{2^k<n\leq 2^{k+1}} \frac{1}{n}\sum_{i=1}^n Z_i \geq 2\mathbb{E}[Z_1]+\delta\Big)\\
&\qquad\leq \mathbb{P}(\exists i \leq 2^{k+1}, Z_i > 2^k) + \mathbb{P}(\max_{2^k<n\leq 2^{k+1}} \frac{1}{n}\sum_{i=1}^n Z_i \mathbb{1}_{[0,2^k]}(Z_i) \geq 2\mathbb{E}[Z_1]+\delta)\\
&\qquad= A_k + B_k,
\end{aligned}$$

car soit tous les Z_i sont inférieurs à 2^k, soit il y en a au moins un qui est supérieur. D'une part, en majorant l'union sur les $i \leq 2^{k+1}$ par la somme, on a

$$\begin{aligned}
A_k = \mathbb{P}(\exists i \leq 2^{k+1}, Z_i > 2^k) &\leq 2^{k+1}\mathbb{P}(Z_1 > 2^k)\\
&= 4 \times 2^{k-1}\mathbb{P}(Z_1 > 2^k)\\
&\leq 4\int_{2^{k-1}}^{2^k} \mathbb{P}(Z_1 > t)dt.
\end{aligned}$$

D'autre part, en utilisant $2^k < n \leq 2^{k+1}$ et la positivité des Z_i, on obtient

$$\begin{aligned}
B_k &\leq \mathbb{P}\Big(\sum_{i=1}^{2^{k+1}} Z_i \mathbb{1}_{[0,2^k]}(Z_i) \geq 2^{k+1}\mathbb{E}[Z_1] + \delta 2^k\Big)\\
&\leq \mathbb{P}\Big(\sum_{i=1}^{2^{k+1}} Z_i \mathbb{1}_{[0,2^k]}(Z_i) \geq 2^{k+1}\mathbb{E}[Z_1\mathbb{1}_{[0,2^k]}(Z_1)] + \delta 2^k\Big)\\
&\leq \mathbb{P}\Big(\sum_{i=1}^{2^{k+1}} Z_i \mathbb{1}_{[0,2^k]}(Z_i) - \mathbb{E}[Z_i\mathbb{1}_{[0,2^k]}(Z_i)] \geq \delta 2^k\Big)\\
&\leq \mathbb{P}\Big(|\sum_{i=1}^{2^{k+1}} Z_i \mathbb{1}_{[0,2^k]}(Z_i) - \mathbb{E}[Z_i\mathbb{1}_{[0,2^k]}(Z_i)]| \geq \delta 2^k\Big)
\end{aligned}$$

puisque $\mathbb{E}[Z_1] \geq \mathbb{E}[Z_1\mathbb{1}_{[0,2^k]}(Z_1)]$. On utilise maintenant l'inégalité de Tchébychev

$$\begin{aligned}
B_k &\leq \frac{1}{\delta^2 2^{2k}} 2^{k+1}\,\mathsf{Var}[Z_1^2\mathbb{1}_{[0,2^k]}(Z_1)]\\
&\leq \frac{1}{\delta^2 2^k} 2\mathbb{E}[Z_1^2\mathbb{1}_{[0,2^k]}(Z_1)].
\end{aligned}$$

On a ainsi obtenu

$$\mathbb{P}\Big(\max_{2^k<n\leq 2^{k+1}} \frac{1}{n}\sum_{i=1}^n Z_i \geq 2\mathbb{E}[Z_1]+\delta\Big) \leq 4\int_{2^{k-1}}^{2^k} \mathbb{P}(Z_1 > t)dt + \frac{1}{\delta^2 2^k} 2\mathbb{E}[Z^2\mathbb{1}_{[0,2^k]}(Z_1)].$$

On somme maintenant ces inégalités sur k. D'une part

$$\sum_{k\geq 0}\int_{2^{k-1}}^{2^k} \mathbb{P}(Z_1 > t)dt \leq \int_0^\infty \mathbb{P}(Z_1 > t)dt = \mathbb{E}[Z_1],$$

d'autre part

$$\begin{aligned}\sum_{k\geq 0} 2^{-k}\mathbb{E}\Big[Z_1^2 \mathbb{1}_{[0,2^k]}(Z_1)\Big] &= \mathbb{E}\Big[Z_1^2 \sum_{k\geq 0} 2^{-k}\mathbb{1}_{[0,2^k]}(Z_1)\Big] \\ &= \mathbb{E}\Big[\sum_{\ell\geq 0}(Z_1^2 \sum_{k\geq 0} 2^{-k}\mathbb{1}_{[0,2^k]}(Z_1))\mathbb{1}_{Z_1\in]2^\ell,2^{\ell+1}]} \\ &\qquad + (Z_1^2 \sum_{k\geq 0} 2^{-k}\mathbb{1}_{[0,2^k]}(Z_1))\mathbb{1}_{Z_1\in[0,1]}\Big].\end{aligned}$$

Pour le premier terme, on a

$$\begin{aligned}\sum_{\ell\geq 0}(Z_1^2 \sum_{k\geq 0} 2^{-k}\mathbb{1}_{[0,2^k]}(Z_1))\mathbb{1}_{Z_1\in]2^\ell,2^{\ell+1}]} &\leq \sum_{\ell\geq 0} \mathbb{1}_{Z_1\in]2^\ell,2^{\ell+1}]}(2^{2\ell+2} \sum_{k\geq \ell+1} 2^{-k}\mathbb{1}_{[0,2^k]}(Z_1)) \\ &\leq \sum_{\ell\geq 0} \mathbb{1}_{Z_1\in]2^\ell,2^{\ell+1}]}(2^{2\ell+2}2^{-\ell}) \\ &= 4\sum_{\ell\geq 0} \mathbb{1}_{Z_1\in]2^\ell,2^{\ell+1}]}(2^{\ell}) \\ &\leq 4Z_1 \sum_{\ell\geq 0} \mathbb{1}_{Z_1\in]2^\ell,2^{\ell+1}]}\end{aligned}$$

Pour le deuxième terme

$$\begin{aligned}Z_1^2 \sum_{k\geq 0} 2^{-k}\mathbb{1}_{[0,2^k]}(Z_1))\mathbb{1}_{Z_1\in[0,1]} &= Z_1^2(\sum_{k\geq 0} 2^{-k})\mathbb{1}_{Z_1\in[0,1]} \\ &= 2Z_1^2\mathbb{1}_{Z_1\in[0,1]} \\ &\leq 2Z_1\mathbb{1}_{Z_1\in[0,1]} \\ &\leq 4Z_1\mathbb{1}_{Z_1\in[0,1]}.\end{aligned}$$

Donc en recollant les deux termes

$$\sum_{k\geq 0} 2^{-k}\mathbb{E}[Z_1^2 \mathbb{1}_{[0,2^k]}(Z_1)] \leq 4\mathbb{E}[Z_1].$$

Ainsi, on obtient

$$\sum_k \mathbb{P}\Big(\max_{2^k<n\leq 2^{k+1}} \frac{1}{n}\sum_{i=I}^n Z_i \geq 2\mathbb{E}[Z_1]+\delta\Big) \leq 4(1+2\delta^{-2})\mathbb{E}[Z].$$

La série converge, donc le lemme de Borel Cantelli donne presque sûrement pour k assez grand

$$\max_{2^k<n\leq 2^{k+1}} \frac{1}{n}\sum_{i=I}^n Z_i < 2\mathbb{E}[Z]+\delta,$$

et puisque δ est arbitraire, on en déduit la limite ps

$$\limsup_{n\to\infty} \frac{1}{n}\sum_{i=I}^n Z_i \leq 2\mathbb{E}[Z].$$

Si on revient maintenant à notre démonstration, on a

$$\begin{aligned}\limsup_{n\to\infty}\frac{1}{n}|S_n| &\leq \limsup_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n}|X_i-Y_i| + \limsup_{n\to\infty}\frac{1}{n}|T_n|\\ &\leq 2\mathbb{E}[|X-Y|]\leq 2\varepsilon.\end{aligned}$$

On a la conclusion voulue puisque ε est arbitraire. □

Exemple 3.4.1. Soit (X_n) une suite de variables aléatoires indépendantes de même loi de Bernoulli de paramètre p. On pose $S_n = \sum_{k=1}^n X_k$. On a déjà vu que S_n/n converge en probabilité vers p, on a maintenant la convergence presque sure puisque les variables de Bernouilli sont intégrables. Une façon de trouver une valeur approchée du paramètre p est donc de calculer S_n/n pour n assez grand.

3.5 Application de la loi des grands nombres à l'estimation ponctuelle

Pour étudier un certain caractère d'une population donnée (taille des Français, poids des plaques de chocolat d'une usine, intentions de vote), on fait l'hypothèse que ce caractère suit une certaine loi de probabilité $\mathbb{P}_X$.

Définition 3.5.1. *On appelle **échantillon aléatoire** de taille n de la loi $\mathbb{P}_X$ une suite $(X_1,\ldots,X_n)$ de n variables aléatoires indépendantes identiquement distribuées de loi $\mathbb{P}_X$. On appelle **échantillon** une réalisation particulière d'un échantillon aléatoire : $(x_1,\ldots,x_n) = (X_1(\omega),\ldots,X_n(\omega))$.*

Le travail du statisticien ou de la statisticienne est de retrouver la loi $\mathbb{P}_X$ ou certaines de ses caractéristiques, à partir d'un ou plusieurs échantillons. En estimation paramétrique, étant donné un paramètre inconnu θ (proportion, moyenne, variance, ...) de la loi $\mathbb{P}_X$, on cherche à donner une valeur numérique pour θ à partir d'un échantillon.

Définition 3.5.2. *Un **estimateur** T_n du paramètre θ associé à un échantillon aléatoire $(X_1,\ldots,X_n)$ de taille n est une fonction du vecteur aléatoire $(X_1,\ldots,X_n)$*

$$T_n = h(X_1,\ldots,X_n).$$

*L'**erreur d'estimation** est la différence entre l'estimateur T_n et le paramètre à estimer θ. C'est la variable aléatoire $T_n-\theta$. L'**erreur quadratique moyenne** est le moment d'ordre 2 de l'erreur d'estimation : $\mathbb{E}[(T_n-\theta)^2]$.*

Proposition 3.5.1. *Si T_n est de carré intégrable, l'erreur quadratique se décompose en un terme de **biais** et un terme de variance :*

$$\mathbb{E}[(T_n-\theta)^2] = (\mathbb{E}[T_n]-\theta)^2 + \mathsf{Var}(T_n).$$

Démonstration. On a

$$\begin{aligned}\mathbb{E}[(T_n-\theta)^2] &= \mathbb{E}[(T_n-\mathbb{E}[T_n]+\mathbb{E}[T_n]-\theta)^2]\\ &= \mathbb{E}[(T_n-\mathbb{E}[T_n])^2] + (\mathbb{E}[T_n]-\theta)^2 + 2(\mathbb{E}[T_n]-\theta)\mathbb{E}[T_n-\mathbb{E}[T_n]]\\ &= \mathsf{Var}(T_n) + (\mathbb{E}[T_n]-\theta)^2 + 0,\end{aligned}$$

d'où le résultat. □

Plus l'erreur quadratique sera faible, plus l'estimateur sera considéré comme satisfaisant.

Définition 3.5.3. *Un estimateur est dit* ***centré*** *ou* ***sans biais*** *si* $\mathbb{E}[T_n] = \theta$. *Un estimateur est* ***asymptotiquement sans biais*** *si*

$$\lim_{n\to\infty} \mathbb{E}[T_n] = \theta.$$

Un estimateur est dit ***convergent*** *s'il converge en probabilité vers* θ *:*

$$\forall \varepsilon > 0, \quad \lim_{n\to+\infty} \mathbb{P}(|T_n - \theta| \geq \varepsilon) = 0.$$

Exemple 3.5.1. Soit $(X_1, \ldots, X_n)$ un échantillon aléatoire de taille n. La moyenne empirique

$$\overline{X}_n = \frac{X_1 + \cdots + X_n}{n}$$

est un estimateur sans biais de l'espérance car $\mathbb{E}[\overline{X}_n] = \mathbb{E}[X_1]$. C'est un estimateur convergent d'après la loi des grands nombres.
La variance empirique

$$\overline{V}_n = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2$$

est un estimateur biaisé de la variance car $\mathbb{E}[\overline{V}_n] = \frac{n}{n-1}\,\mathsf{Var}(X_1)$. En effet, on a

$$\begin{aligned}
\mathbb{E}[\overline{V}_n] &= \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[(X_i - \overline{X}_n)^2] \\
&= \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[(X_i - \mathbb{E}[X_i] + \mathbb{E}[X_i] - \overline{X}_n)^2] \\
&= \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[(X_i - \mathbb{E}[X_i])^2] + \mathbb{E}[(\mathbb{E}[X_i] - \overline{X}_n)^2] + 2\mathbb{E}[(X_i - \mathbb{E}[X_i])(\mathbb{E}[X_i] - \overline{X}_n)] \\
&= \mathsf{Var}(X_1) + \mathsf{Var}(\overline{X}_n) - \frac{2}{n}\sum_{i=1}^{n}\mathsf{Cov}(X_i, \overline{X}_n).
\end{aligned}$$

Or on a par indépendance

$$\mathsf{Var}(\overline{X}_n) = \mathsf{Var}(\frac{X_1 + \cdots + X_n}{n}) = \frac{n}{n^2}\,\mathsf{Var}(X_1),$$

et

$$\begin{aligned}
\mathsf{Cov}(X_i, \overline{X}_n) &= \mathbb{E}[X_i\overline{X}_n] - \mathbb{E}[X_1]^2 \\
&= \frac{n-1}{n}\mathbb{E}[X_1]^2 + \frac{1}{n}\mathbb{E}[X_1^2] - \mathbb{E}[X_1]^2 \\
&= \frac{1}{n}\,\mathsf{Var}(X_1),
\end{aligned}$$

Donc finalement on a bien

$$\begin{aligned}
\mathbb{E}[\overline{V}_n] &= \mathsf{Var}(X_1) + \mathsf{Var}(\overline{X}_n) - \frac{2}{n}\sum_{i=1}^{n}\mathsf{Cov}(X_i, \overline{X}_n) \\
&= \mathsf{Var}(X_1) + \frac{1}{n}\,\mathsf{Var}(X_1) - \frac{2}{n}\,\mathsf{Var}(X_1)
\end{aligned}$$

$$= \frac{n}{n-1} \mathsf{Var}(X_1)$$

Cependant cet estimateur est asymptotiquement sans biais et convergent. On utilise plutôt sa version sans biais

$$S_n^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \overline{X}_n)^2$$

On va voir dans le chapitre suivant comment évaluer la précision de ces estimateurs.

Chapitre 4

Théorème Central Limite

Attention, dans ce chapitre on ne suppose plus que les variables aléatoires sont définies sur la même espace probabilisé. Les variables sont toujours à valeurs dans $\mathbb{R}$ ou $\mathbb{R}^d$.

4.1 Convergence en loi

On commence par introduire un nouveau type de convergence, la convergence en loi.

Rappel 7. Une suite de lois de probabilités $(\mu_n)_n$ définies sur un même espace $(\Omega, \mathcal{F})$ converge

— **simplement** vers la mesure μ si

$$\mu_n(A) \xrightarrow[n\to+\infty]{} \mu(A), \qquad \text{pour tout } A \text{ dans } \mathcal{F}.$$

— **étroitement** vers la mesure μ si

$$\int h d\mu_n \xrightarrow[n\to+\infty]{} \int h d\mu \text{pour toute fonction } h \text{ continue bornée.}$$

— **vaguement** vers la mesure μ si

$$\int h d\mu_n \xrightarrow[n\to+\infty]{} \int h d\mu \text{pour toute fonction } h \text{ continue à support compact.}$$

Définition 4.1.1. *Pour tout $n \geq 1$, soit X_n une variable aléatoire définie sur l'espace probabilisé $(\Omega_n, \mathcal{F}_n, \mathbb{P}_n)$ à valeurs dans $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ (ou $(\mathbb{R}^d, \mathcal{B}(\mathbb{R})^d)$), et μ une loi de probabilité sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ (ou $(\mathbb{R}^d, \mathcal{B}(\mathbb{R})^d)$). On dit que la suite (X_n)* ***converge en loi*** *vers μ si la suite $(\mathbb{P}_{X_n})$ des lois de (X_n) converge étroitement vers la loi μ, c'est-à-dire que pour toute fonction ϕ continue bornée sur $\mathbb{R}$ (ou $\mathbb{R}^d$) on a*

$$\lim_{n\to\infty} \int \phi d\mathbb{P}_{X_n} = \int \phi d\mu.$$

On note $X_n \xrightarrow{\mathcal{L}} \mu$.

Cette notion n'est pas relative aux variables aléatoires comme fonctions mais concerne uniquement leur loi. Par abus de langage, on pourra dire aussi que (X_n) converge en loi vers la variable aléatoire X pour toute variable aléatoire de loi μ. On notera alors $X_n \xrightarrow{\mathcal{L}} X$ ou $X_n \xrightarrow{\mathcal{L}} \mathbb{P}_X$ qui correspond alors à $\lim_{n\to\infty} \mathbb{E}[\phi(X_n)] = \mathbb{E}[\phi(X)]$ pour toute fonction ϕ continue bornée.

Exemple 4.1.1. Soit X une variable aléatoire de loi normale centrée réduite et $X_n = (-1)^n X$. Alors la loi de X_n est encore la loi $\mathcal{N}(0, 1)$ puisqu'elle est symétrique, donc la suite (X_n)

converge en loi vers la loi $\mathcal{N}(0,1)$.

Théorème 4.1.1 Théorème de Lévy. *Pour tout $n \geq 1$, soit X_n une variable aléatoire définie sur l'espace probabilisé $(\Omega_n, \mathcal{F}_n, \mathbb{P}_n)$ de fonction caractéristique Φ_{X_n}.*

1. *Si (X_n) converge en loi vers une loi μ, alors la suite des fonctions caractéristiques (Φ_{X_n}) converge simplement vers la fonction caractéristique Φ_μ de μ.*
2. *Réciproquement, si la suite des fonctions caractéristiques (Φ_{X_n}) converge simplement vers une fonction Φ qui est la fonction caractéristique d'une loi de probabilité μ, alors la suite (X_n) converge en loi vers μ.*

Démonstration.

1. La partie réelle et la partie imaginaire de $x \mapsto \exp(i\langle t, x\rangle)$ sont $x \mapsto \cos(\langle t, x\rangle)$ et $x \mapsto \sin(\langle t, x\rangle)$ respectivement. Ces deux fonctions sont continues bornées, donc si (X_n) converge en loi vers μ, on a pour t fixé
$$\lim_{n\to\infty} \int \cos(\langle t, x\rangle) d\mathbb{P}_{X_n}(x) = \int \cos(\langle t, x\rangle) d\mu(x),$$
$$\lim_{n\to\infty} \int \sin(\langle t, x\rangle) d\mathbb{P}_{X_n}(x) = \int \sin(\langle t, x\rangle) d\mu(x),$$
d'où la convergence simple des fonctions caractéristiques.
2. On fait la démonstration en 3 étapes détaillées dans les lemmes suivants.

□

Définition 4.1.2. *Une suite (μ_n) de mesures est* ***tendue*** *si pour tout $\varepsilon > 0$ il existe un compact $K = [-a,a]^d$ de $\mathbb{R}^d$ tel que*
$$\sup_{n\in\mathbb{N}} \mu_n(K^c) \leq \varepsilon.$$

Lemme 4.1.1. *Soit $(\mathbb{P}_n)$ une suite de lois de probabilités et μ une mesure bornée. Si $\int h d\mathbb{P}_n \to \int h d\mu$ pour toute fonction h continue à support compact (convergence vague) et la suite $(\mathbb{P}_n)$ est tendue, alors $\int \phi d\mathbb{P}_n \to \int \phi d\mu$ pour toute fonction ϕ continue bornée (convergence étroite).*

Démonstration. Soit ϕ une fonction continue bornée et g_a la fonction réelle définie par
$$g_a(t) = \begin{cases} 0 & \text{si } |t| \geq a+1, \\ 1 & \text{si } |t| \leq a, \end{cases}$$
et prolongée linéairement entre $-(a+1)$ et $-a$ et entre a et $a+1$, et soit h_a la fonction de $\mathbb{R}^d$ dans $\mathbb{R}$ qui a $(t_1,\dots,t_d)$ associe $g_a(t_1)\times\cdots\times g_a(t_d)$, voir figure 4.1. On a
$$\begin{aligned}
&\Big|\int \phi d\mathbb{P}_n - \int \phi d\mu\Big| \\
\leq\ & \Big|\int \phi - \phi\times h_a d\mathbb{P}_n\Big| + \Big|\int \phi\times h_a d\mathbb{P}_n - \int \phi\times h_a d\mu\Big| + \Big|\int \phi\times h_a - \phi d\mu\Big| \\
\leq\ & \int |\phi|(1-h_a) d\mathbb{P}_n\Big| + \Big|\int \phi\times h_a d\mathbb{P}_n - \int \phi\times h_a d\mu\Big| + \Big|\int |\phi|(1-h_a) d\mu\Big| \\
\leq\ & \sup|\phi|\mathbb{P}_n(([-a,a]^d)^c) + \Big|\int \phi\times h_a d\mathbb{P}_n - \int \phi\times h_a d\mu\Big| + \sup|\phi|\mu(([-a,a]^d)^c).
\end{aligned}$$

Comme $(\mathbb{P}_n)$ est tendue, et μ est bornée donc tendue, pour tout $\varepsilon > 0$ il existe a tel que $\sup|\phi|\mathbb{P}_n(([-a,a]^d)^c) + \sup|\phi|\mu(([-a,a]^d)^c) \leq \varepsilon/2$. Comme la suite est vaguement convergente, il existe N tel que pour tout $n \geq N$ on ait $\left|\int \phi\times h_a d\mathbb{P}_n - \int \phi\times h_a d\mu\right| \leq \varepsilon/2$. On a donc pour tout $n \geq N$, $\left|\int \phi d\mathbb{P}_n - \int \phi d\mu\right| \leq \varepsilon$ et la convergence voulue. □

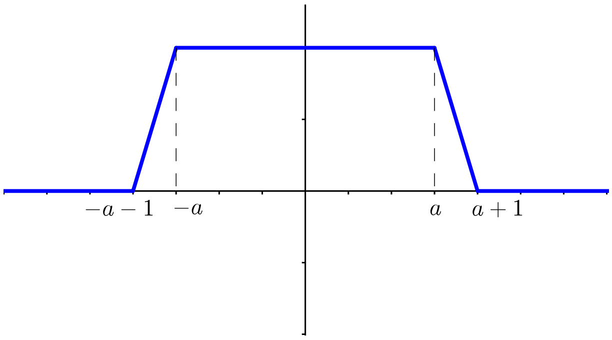

FIGURE 4.1 – Fonction g_a.

Il nous suffit donc de montrer que notre suite de probabilités converge vaguement et qu'elle est tendue. Pour montrer la tension, on a d'abord besoin du lemme de Cramer.

Lemme 4.1.2 Lemme de Cramer. *Soit μ une loi de probabilité sur $\mathbb{R}$ ou $\mathbb{R}^d$. Alors il existe une constante universelle C (qui ne dépend pas de μ) telle que pour tout réel a strictement positif,*

$$\mu(([-a,a]^d)^c) \leq C \int_{[0,1]^d} \Big(1 - \Re(\Phi_\mu(t/a))\Big) dt,$$

où $\Re$ désigne la partie réelle.

Démonstration. Comme μ est une probabilité et $1 - \Re(\Phi_\mu(t) \geq 0$, on peut utiliser le théorème de Fubini Tonelli pour obtenir

$$\begin{aligned}
&\int_{[0,1]^d} \Big(1 - \Re \int_{\mathbb{R}^d} \exp(i\langle t/a, x\rangle\, d\mu(x)\Big) dt \\
&= \Re \int_{[0,1]^d} \int_{\mathbb{R}^d} (1 - \exp(i\langle t/a, x\rangle)) d\mu(x) dt \\
&= \Re \int_{\mathbb{R}^d} \Big(\int_{[0,1]^d} (1 - \exp(i\langle t/a, x\rangle)) dt \Big) d\mu(x).
\end{aligned}$$

L'intégrande est positive puisque la partie réelle de l'exponentielle est un cosinus donc inférieur ou égale à 1. On peut donc minorer par l'intégrale sur $\mathbb{R}^d$ par la même intégrale sur un domaine plus petit de la forme $([-a,a]^d)^c$

$$\begin{aligned}
&\int_{[0,1]^d} \Big(1 - \Re \int_{\mathbb{R}^d} \exp(i\langle t/a, x\rangle\, d\mu(x)\Big) dt \\
&\geq \Re \int_{([-a,a]^d)^c} \Big(\int_{[0,1]^d} (1 - \exp(i\langle t/a, x\rangle)) dt \Big) d\mu(x) \\
&= \Re \int_{([-a,a]^d)^c} \Big(\int_{[0,1]^d} (1 - \exp(it_1x_1/a + \cdots + it_dx_d/a)) dt_1 \cdots dt_d \Big) d\mu(x).
\end{aligned}$$

On peut maintenant calculer explicitement l'intégrale sur $[0,1]^d$ par récurrence

$$\begin{aligned}
&\int_{[0,1]^d} \Big(1 - \Re \int_{\mathbb{R}^d} \exp(i\langle t/a, x\rangle\, d\mu(x)\Big) dt \\
&\geq \Re \int_{([-a,a]^d)^c} \int_{[0,1]^{d-1}} \Big[t_1 - \frac{a}{ix_1} \exp(it_1x_1/a) \prod_{k=2}^d \exp(it_kx_k/a)\Big]_0^1 dt_2 \cdots dt_d d\mu(x)
\end{aligned}$$

$$
\begin{aligned}
&= \Re \int_{([-a,a]^d)^c} \int_{[0,1]^{d-1}} \Big(1 - \frac{a}{ix_1}(\exp(x_1/a) - 1) \prod_{k=2}^d \exp(it_k x_k/a)\Big) dt_2 \cdots dt_d d\mu(x) \\
&= \Re \int_{([-a,a]^d)^c} \int_{[0,1]^{d-2}} \Big[t_2 - \frac{a}{ix_1}(\exp(ix_1/a) - 1)\frac{a}{ix_2}\exp(it_2x_2/a) \prod_{k=3}^d \exp(it_k x_k/a)\Big]_0^1 dt_3 \cdots dt_d d\mu(x) \\
&= \Re \int_{([-a,a]^d)^c} \int_{[0,1]^{d-2}} \Big(1 - \prod_{k=1}^2 \frac{a}{ix_k}(\exp(ix_k/a) - 1) \prod_{k=3}^d \exp(it_k x_k/a)\Big) dt_3 \cdots dt_d d\mu(x) \\
&= \dots \\
&= \Re \int_{([-a,a]^d)^c} \Big(1 - \prod_{k=1}^d \frac{a}{ix_k}(\exp(ix_k/a) - 1)\Big) d\mu(x).
\end{aligned}
$$

Au passage, $(\exp(ix) - 1)/ix$ est bien défini par continuité en 0 et vaut 1 en 0. On prend maintenant une minoration grossière par la borne inférieure

$$
\begin{aligned}
&\int_{[0,1]^d} \Big(1 - \Re \int_{\mathbb{R}^d} \exp(i \langle t/a, x\rangle \, d\mu(x)\Big) dt \\
&\geq \int_{([-a,a]^d)^c} \inf_{x\in([-a,a]^d)^c} \Re\Big(1 - \prod_{k=1}^d \frac{a}{ix_k}(\exp(ix_k/a) - 1)\Big) d\mu(x) \\
&= \inf_{x\in([-a,a]^d)^c} \Re\Big(1 - \prod_{k=1}^d \frac{a}{ix_k}(\exp(ix_k/a) - 1)\Big) \mu([-a,a]^d)^c) \\
&= \inf_{y\in([-1,1]^d)^c} \Re\Big(1 - \prod_{k=1}^d (\exp(iy_k) - 1)/(iy_k)\Big) \mu([-a,a]^d)^c),
\end{aligned}
$$

en posant le changement de variable $y_k = x_k/a$. Il reste à montrer que la borne inférieure est strictement positive. Or

$$
\begin{aligned}
\Big|\prod_{k=1}^d (\exp(iy_k) - 1)/(iy_k)\Big| &= \prod_{k=1}^d |(\exp(iy_k) - 1)/(iy_k)| \\
&= \prod_{k=1}^d \big((\cos y_k - 1)^2 + \sin^2 y_k\big)^{1/2}/|y_k| \\
&= \prod_{k=1}^d (2 - 2\cos y_k)^{1/2}/|y_k|.
\end{aligned}
$$

Or, en étudiant la fonction $y \mapsto 2 - 2\cos(y) - y^2$, on peut montrer que $(2 - 2\cos y)/y^2 \leq 2 - 2\cos 1 \simeq 0.9194 < 1$ pour tout tout $y \geq 1$ (de même pour $y \leq -1$ par parité, voir Figure 4.2). Donc on a

$$
\Big|\prod_{k=1}^d (\exp(iy_k) - 1)/(iy_k)\Big| \leq (2 - 2\cos 1)^{d/2} = \alpha < 1
$$

pour $y \in ([-1,1]^d)^c$ d'où $\Re\Big(1 - \prod_{k=1}^d (\exp(iy_k) - 1)/(iy_k)\Big) \geq 1 - \alpha > 0$ pour $y \in ([-1,1]^d)^c$. La borne inférieure est donc strictement positive. Elle ne dépend ni de a ni de μ. On pose C égale à son inverse et on a le résultat. □

On peut maintenant prouver que si on a convergence des fonctions caractéristiques alors la suite est tendue.

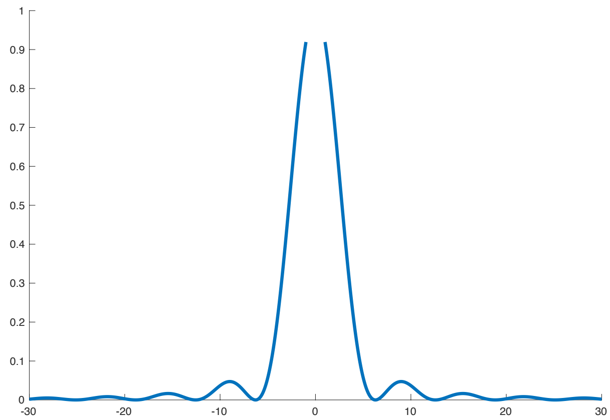

FIGURE 4.2 – Fonction $y \mapsto (2-2\cos y)/y^2$ sur $[-1,1]^c$.

Lemme 4.1.3. *Pour tout $n \geq 1$, soit X_n une variable aléatoire définie sur l'espace probabilisé $(\Omega_n, \mathcal{F}_n, \mathbb{P}_n)$ de fonction caractéristique Φ_{X_n}. Si la suite des fonctions caractéristiques (Φ_{X_n}) converge simplement vers une fonction Φ continue en 0, alors la suite de mesures de probabilités $(\mathbb{P}_{X_n})$ est tendue.*

Notons que par convergence dominée, toute fonction caractéristiques est continue en 0.

Démonstration. Soit la famille de fonctions $t \mapsto 1 - \Re\Phi(t/a)$ indexée par $a > 0$. Par passage à la limite des Φ_{X_n}, on a $|\Phi| \leq 1$, donc $1 - \Re\Phi(t/a)$ est majoré en module indépendamment de a pour $t \in [0,1]^d$. De plus, par continuité de Φ en 0, on a $1 - \Re\Phi(t/a)$ tend vers $1 - \Re\Phi(0) = 0$ lorsque a tend vers l'infini (car $\Phi_{X_n}(0) = 0$ + convergence simple). On peut donc appliquer le théorème de convergence dominée pour obtenir

$$\lim_{a \to +\infty} \int_{[0,1]^d} \Big(1 - \Re(\Phi(t/a))\Big) dt = 0.$$

Soit $\varepsilon > 0$, il existe donc a_1 tel que pour tout $a \geq a_1$ on ait $\int_{[0,1]^d} \Big(1 - \Re(\Phi(t/a))\Big) dt \leq \varepsilon/2$. La suite (Φ_{X_n}) est bornée par 1 en module et convergente, donc par le théorème de convergence dominée, on a

$$\lim_{n \to +\infty} \int_{[0,1]^d} \Big(1 - \Re(\Phi_{X_n}(t/a))\Big) dt = \int_{[0,1]^d} \Big(1 - \Re(\Phi(t/a))\Big) dt.$$

Donc pour tout a fixé, il existe N tel que pour tout $n \geq N$

$$\Big| \int_{[0,1]^d} \Big(1 - \Re(\Phi_{X_n}(t/a))\Big) dt - \int_{[0,1]^d} \Big(1 - \Re(\Phi(t/a))\Big) dt \Big| \leq \varepsilon/2.$$

Soit $a > a_1$ et $n \geq N(a)$. Alors

$$\int_{[0,1]^d} \Big(1 - \Re(\Phi_{X_n}(t/a))\Big) dt \leq \varepsilon/2 + \varepsilon/2.$$

Par le lemme de Cramer, on en déduit que $\mathbb{P}_{X_n}(([-a,a]^d)^c) \leq C\varepsilon$ pour $n \geq N(a)$. Comme les autres sont en nombre fini, on peut trouver $a_2 \geq a_1$ tel que pour tout $a \geq a_2$, $\mathbb{P}_{X_n}(([-a,a]^d)^c) \leq C\varepsilon$ pour tout n. Donc le sup sur n est inférieur à $C\varepsilon$ et la suite est tendue. $\square$

On recolle maintenant les morceaux pour démontrer la réciproque du théorème de Lévy.

Démonstration. du théorème de Lévy
Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace probabilisé sur lequel on définit des variables aléatoires (Z_n), Z et Y telles que Z_n a la même loi que X_n pour tout n, Y suit la loi $\mathcal{N}(0, I_d)$, Z suit la loi μ, et Y est indépendante de (Z_n) et de Z. Soit σ un réel strictement positif. Pour toute fonction h continue à support compact, on a

$$\begin{aligned}
&|\mathbb{E}_n[h(X_n)] - \mathbb{E}[h(Z)]| \\
&= |\mathbb{E}[h(Z_n)] - \mathbb{E}[h(Z)]| \\
&= |\mathbb{E}[h(Z_n)] - \mathbb{E}[h(Z_n + \sigma Y)] + \mathbb{E}[h(Z_n + \sigma Y)] - \mathbb{E}[h(Z + \sigma Y)] + \mathbb{E}[h(Z + \sigma Y)] - \mathbb{E}[h(Z)]| \\
&\leq \mathbb{E}[|h(Z_n) - h(Z_n + \sigma Y)|] + |\mathbb{E}[h(Z_n + \sigma Y)] - \mathbb{E}[h(Z + \sigma Y)]| + \mathbb{E}[|h(Z + \sigma Y)] - h(Z)|].
\end{aligned}$$

Comme h est continue à support compact, elle est en particulier bornée par $\|h\|$ et uniformément continue. Soit donc $\varepsilon > 0$ et δ_ε tel que pour tous (x, y) tels que $|x-y| \leq \delta_\varepsilon$ on ait $|h(x)-h(y)| \leq \varepsilon$. On a alors

$$\begin{aligned}
&\mathbb{E}[|h(Z_n) - h(Z_n + \sigma Y)|] \\
&= \mathbb{E}[|h(Z_n) - h(Z_n + \sigma Y)|\mathbb{1}_{(|\sigma Y| \leq \delta_\varepsilon)}] + \mathbb{E}[|h(Z_n) - h(Z_n + \sigma Y)|\mathbb{1}_{(|\sigma Y| \leq \delta_\varepsilon)}] \\
&\leq \varepsilon \mathbb{P}(|\sigma Y| \leq \delta_\varepsilon) + 2\|h\|\mathbb{P}(|\sigma Y| > \delta_\varepsilon) \\
&\leq \varepsilon + 2\|h\| \frac{\mathbb{E}[(\sigma Y)^2]}{\delta_\varepsilon^2} \\
&\leq \varepsilon + 2\|h\| \frac{\sigma^{2d}}{\delta_\varepsilon},
\end{aligned}$$

par l'inégalité de Markov et en utilisant que $\mathbb{E}[Y^2] = 1$. De la même façon, on a

$$\mathbb{E}[|h(Z + \sigma Y) - h(Z)|] \leq \varepsilon + 2\|h\| \frac{\sigma^{2d}}{\delta_\varepsilon}.$$

Finalement, on obtient

$$\begin{aligned}
&|\mathbb{E}_n[h(X_n)] - \mathbb{E}[h(Z)]| \\
&\leq 2\varepsilon + 4\|h\| \frac{\sigma^{2d}}{\delta_\varepsilon} + |\mathbb{E}[h(Z_n + \sigma Y)] - \mathbb{E}[h(Z + \sigma Y)]|.
\end{aligned}$$

Dans la preuve de la Proposition 2.6.2 on a vu que

Rappel 8. Pour toute variable aléatoire X et toute fonction h continue bornée

$$\mathbb{E}[h(X + \sigma Y)] = \int_{\mathbb{R}^d} h(v) \Big(\int_{\mathbb{R}^d} \frac{1}{(2\pi)^d} e^{-i\langle z,u \rangle - \frac{\sigma^2}{2}|u|^2} \Phi_X(u) du \Big) dv.$$

Donc on a

$$\mathbb{E}[h(Z_n + \sigma Y)] = \int_{\mathbb{R}^d} h(v) \Big(\int_{\mathbb{R}^d} \frac{1}{(2\pi)^d} e^{-i\langle z,u \rangle - \frac{\sigma^2}{2}|u|^2} \Phi_{Z_n}(u) du \Big) dv,$$

D'une part, on a $\Phi_{Z_n}(u)$ qui converge simplement vers $\Phi_Z(u)$ lorsque n tend vers l'infini. De plus,

$$\Big| \frac{1}{(2\pi)^d} e^{-i\langle z,u \rangle - \frac{\sigma^2}{2}|u|^2} \Phi_{Z_n}(u) \Big| \leq \frac{1}{(2\pi)^d} e^{-\frac{\sigma^2}{2}|u|^2}$$

qui est d'intégrale 1 contre la mesure de Lebesgue. Une première application du théorème de convergence dominée donne la convergence simple de $\int_{\mathbb{R}^d} \frac{1}{(2\pi)^d} e^{-i\langle z,u\rangle - \frac{\sigma^2}{2}|u|^2} \Phi_{Z_n}(u)du$ vers $\int_{\mathbb{R}^d} \frac{1}{(2\pi)^d} e^{-i\langle z,u\rangle - \frac{\sigma^2}{2}|u|^2} \Phi_{Z}(u)du$. D'autre part

$$\Big|h(v)\Big(\int_{\mathbb{R}^d} \frac{1}{(2\pi)^d} e^{-i\langle z,u\rangle - \frac{\sigma^2}{2}|u|^2} \Phi_{Z_n}(u)du\Big)\Big| \leq |h(v)| \int_{\mathbb{R}^d} \frac{1}{(2\pi)^d} e^{-\frac{\sigma^2}{2}|u|^2} du = |h(v)|,$$

qui est à nouveau intégrable contre la mesure de Lebesgue puisque h est continue à support compact. Une deuxième application du théorème de convergence dominée donne alors

$$\begin{aligned}\lim_{n\to\infty} \mathbb{E}[h(Z_n + \sigma Y)] &= \int_{\mathbb{R}^d} h(v)\Big(\int_{\mathbb{R}^d} \frac{1}{(2\pi)^d} e^{-i\langle z,u\rangle - \frac{\sigma^2}{2}|u|^2} \Phi_{Z}(u)du\Big)dv \\ &= \mathbb{E}[h(Z + \sigma Y)].\end{aligned}$$

On obtient ainsi

$$\begin{aligned}\limsup_{n\to\infty} |\mathbb{E}_n[h(X_n)] - \mathbb{E}[h(Z)]| &\leq 2\varepsilon + 4\|h\|\frac{\sigma^{2d}}{\delta_\varepsilon} + \lim_{n\to\infty} |\mathbb{E}[h(Z_n + \sigma Y)] - \mathbb{E}[h(Z+\sigma Y)]| \\ &\leq 2\varepsilon + 4\|h\|\frac{\sigma^{2d}}{\delta_\varepsilon},\end{aligned}$$

qui est valable pour tout σ et pour tout ε. On fait ensuite tendre σ vers 0 pour obtenir

$$\limsup_{n\to\infty} |\mathbb{E}_n[h(X_n)] - \mathbb{E}[h(Z)]| \leq 2\varepsilon$$

et enfin on fait tendre ε vers 0 pour obtenir le résultat voulu $\limsup_{n\to\infty} |\mathbb{E}_n[h(X_n)] - \mathbb{E}[h(Z)]| = 0$. On a ainsi montré la convergence sur les fonctions continues à support compact et la tension, donc on a bien la convergence en loi. □

Exemple 4.1.2. Soit $(X_n)_{n\geq 1}$ une suite de variables aléatoires indépendantes de même loi de Bernoulli de paramètre p et $S_n = \sum_{k=1}^n X_k$. On sait que S_n suit la loi binomiale (n,p) et $\Phi_{S_n}(t) = (1-p+pe^{it})^n$. Donc

$$\begin{aligned}\Phi_{\frac{S_n - np}{\sqrt{n}}}(t) &= \mathbb{E}[e^{it\frac{S_n-np}{\sqrt{n}}}] \\ &= \mathbb{E}[e^{it\frac{S_n}{\sqrt{n}}}]e^{-it\frac{np}{\sqrt{n}}} \\ &= e^{-it\frac{np}{\sqrt{n}}}\Phi_{S_n}(\frac{t}{\sqrt{n}}) \\ &= e^{-it\frac{np}{\sqrt{n}}}(1-p+pe^{i\frac{t}{\sqrt{n}}})^n \\ &= e^{-it\frac{np}{\sqrt{n}}}\exp(n\log(1-p+pe^{i\frac{t}{\sqrt{n}}})) \\ &= e^{-it\frac{np}{\sqrt{n}}}\exp(n(-p+pe^{i\frac{t}{\sqrt{n}}} - \frac{(p+pe^{i\frac{t}{\sqrt{n}}})^2}{2}) + o(1)) \\ &= e^{-it\frac{np}{\sqrt{n}}}\exp(n(-p+pe^{i\frac{t}{\sqrt{n}}} - \frac{p^2}{2} - \frac{p^2}{2}e^{2i\frac{t}{\sqrt{n}}} + p^2e^{i\frac{t}{\sqrt{n}}}) + o(1)) \\ &= e^{-it\frac{np}{\sqrt{n}}}\exp(n(-p - \frac{p^2}{2} + (p-p^2)e^{i\frac{t}{\sqrt{n}}} - \frac{p^2}{2}e^{2i\frac{t}{\sqrt{n}}} + o(1)) \\ &= e^{-it\frac{np}{\sqrt{n}}}\exp(n(-p - \frac{p^2}{2} + (p-p^2)(1+i\frac{t}{\sqrt{n}} + \frac{(it)^2}{2n}) - \frac{p^2}{2}(1+2i\frac{t}{\sqrt{n}} + \frac{(2it)^2}{2n}) + o(1)) \\ &= e^{-it\frac{np}{\sqrt{n}}}\exp(npi\frac{t}{\sqrt{n}} - p(1-p)\frac{t^2}{2} + o(1))\end{aligned}$$

$$= \exp(-(p-p^2)\frac{t^2}{2} + o(1)),$$

donc $(S_n - np)/\sqrt{n}$ converge en loi vers $\mathcal{N}(0, p(1-p))$.

Exemple 4.1.3. Soit X_n et X des variables aléatoires à valeurs dans $\mathbb{R}^d$. Alors $X_n \xrightarrow{\mathcal{L}} X$ si et seulement si pour tout $u \in \mathbb{R}^d$, $\langle u, X_n\rangle \xrightarrow{\mathcal{L}} \langle u, X\rangle$. En particulier en prenant pour u un vecteur de la base canonique, si $X_n \xrightarrow{\mathcal{L}} X$ alors les coordonnées de X_n convergent en loi vers les coordonnées de X.

Contre-Exemple 4.1.1. La convergence des coordonnées n'implique pas la convergence en loi du vecteur (de même que la loi des coordonnées ne suffit pas à définir la loi du vecteur). Soit X une variable aléatoire de loi normale centrée réduite et $X_n = (-1)^n X$. On a vu que la loi de X_n est encore la loi $\mathcal{N}(0,1)$ puisqu'elle est symétrique, donc la suite (X_n) converge en loi vers X, la suite constante X converge en loi vers X. Or $(X_n, X) = ((-1)^n X, X) = ((-1)^n, 1)X$ ne converge pas en loi vers (X, X). En effet, sinon, par l'exemple précédent on aurait pour $u = (1,1)$, $X_n - X = ((-1)^n + 1)X$ qui converge en loi vers $2X$. Or $X_{2n} + X = 2X$ mais $X_{2n+1} - X = 0$ donc la suite ne converge pas. On constate aussi au passage (et pour les mêmes raisons) que si deux suites convergent en loi, alors la suite somme ne converge pas nécessairement.

4.2 Critères de convergence en loi

On donne maintenant des critères de convergence en loi dans les cas particuliers : dimension 1, variables à densité et variables discrètes.

Proposition 4.2.1. *Pour tout $n \geq 1$, soit X_n une variable aléatoire définie sur l'espace probabilisé $(\Omega_n, \mathcal{F}_n, \mathbb{P}_n)$ et X une variable aléatoire définie sur $(\Omega, \mathcal{F}, \mathbb{P})$. Si ce sont des variables aléatoires réelles, alors les assertions suivantes sont équivalentes.*

1. *La suite (X_n) converge en loi vers X.*
2. *La suite des fonctions de répartition F_{X_n} converge simplement vers F_X en tout point de continuité de F_X.*
3. *Il existe un espace probabilisé $(\Omega', \mathcal{F}', \mathbb{P}')$ sur lequel sont définies des variables aléatoires X'_n et X' telles que pour tout n X'_n a la même loi que X_n, X' a la même loi que X et $X'_n \xrightarrow{ps} X'$.*

Démonstration. $1 \Rightarrow 2$ Supposons que la suite converge en loi. Pout tout $\varepsilon > 0$, $t \in \mathbb{R}$ et $x \in \mathbb{R}$, on a (voir Figure 4.3)

$$\mathbb{1}_{]-\infty, t-\varepsilon]}(x) \leq \frac{(t-x)^+}{\varepsilon} \wedge 1 \leq \mathbb{1}_{]-\infty,t]}(x) \leq \frac{(t+\varepsilon-x)^+}{\varepsilon} \wedge 1 \leq \mathbb{1}_{]-\infty,t+\varepsilon]}(x).$$

On a ainsi

$$\begin{aligned} F(t-\varepsilon) &= \mathbb{E}[\mathbb{1}_{]-\infty,t-\varepsilon]}(X)] \\ &\leq \mathbb{E}\Big[\frac{(t-X)^+}{\varepsilon} \wedge 1\Big] \end{aligned}$$

La fonction $t \mapsto \frac{(t-x)^+}{\varepsilon} \wedge 1$ est continue bornée, donc la convergence en loi donne

$$F(t-\varepsilon) \leq \mathbb{E}\Big[\frac{(t-X)^+}{\varepsilon} \wedge 1\Big]$$

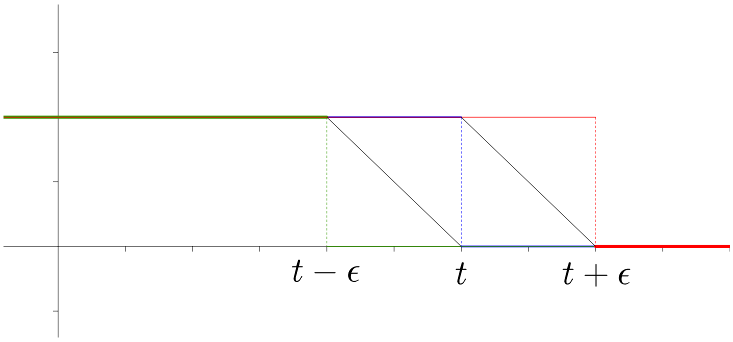

FIGURE 4.3 – Encadrement de $x \mapsto \mathbb{1}_{]-\infty,t]}(x)$.

$$\begin{aligned} &= \lim_{n\to\infty} \mathbb{E}\Big[\frac{(t-X_n)^+}{\varepsilon} \wedge 1\Big] \\ &\leq \liminf_{n\to\infty} \mathbb{E}[\mathbb{1}_{]-\infty,t]}(X_n)] \\ &= \liminf_{n\to\infty} F_{X_n}(t). \end{aligned}$$

De même, on a

$$\begin{aligned} F(t+\varepsilon) &= F(t+\varepsilon) \\ &= \mathbb{E}[\mathbb{1}_{]-\infty,t+\varepsilon]}(X)] \\ &\geq \mathbb{E}\Big[\frac{(t+\varepsilon-X)^+}{\varepsilon} \wedge 1] \\ &= \limsup_{n\to\infty} \mathbb{E}\Big[\frac{(t+\varepsilon-X_n)^+}{\varepsilon} \wedge 1\Big] \\ &\geq \limsup_{n\to\infty} \mathbb{E}[\mathbb{1}_{]-\infty,t]}(X_n)] \\ &= \limsup_{n\to\infty} F_{X_n}(t). \end{aligned}$$

D'où la convergence des fonctions de répartition aux points de continuité de F puisque ε est arbitraire.
$2 \Rightarrow 3$ On définit l'espace $(\Omega', \mathcal{F}', \mathbb{P}')$ par $(]0,1[, \mathcal{B}(]0,1[), \lambda)$ où λ est la mesure de Lebesgue. Soit U de loi uniforme sur $]0,1[$ (on peut choisir la fonction identité sur Ω'). On pose $X'_n = F^{\leftarrow}_{X_n}(U)$ et $X' = F^{\leftarrow}_X(U)$. Alors X'_n a la même loi que X_n et X' a la même loi que X. Il suffit donc de montrer que $F^{\leftarrow}_{X_n}(u)$ converge vers $F^{\leftarrow}_X(u)$ pour un sous-ensemble de $]0,1[$ de mesure de Lebesgue 1.

Rappel 9.

$$F^{\leftarrow}(u) = \inf\{t; F(t) > u\}.$$

Pour tout $u \in]0,1[$, si $F^{\leftarrow}(u) \leq t$ alors $F(t) \geq u$. De plus, si $F(t) > u$, alors $F^{\leftarrow}(u) \leq t$. Soit $u \in]0,1[$ et $t = F^{\leftarrow}_X(u)$, donc $F_X(t) \geq u$. Soit $\varepsilon > 0$ et t^+_ε, t^-_ε des points de continuité de F_X tels que $t^-_\varepsilon < t < t^+_\varepsilon$ et $|t^+_\varepsilon - t^-_\varepsilon| \leq \varepsilon$ (possible car il y a au plus un nombre dénombrable de points de discontinuité). Comme $t^-_\varepsilon < t$, par définition de la fonction quantile, on a $F_X(t^-_\varepsilon) \leq u$. Par croissance d'une fonction de répartition, on a aussi $F_X(t^+_\varepsilon) \geq F_X(t) \geq u$. Comme les fonctions de répartition convergent pour les points de continuité de F_X, pour tout $\eta > 0$ tel que $0 < u - \eta < u + \eta < 1$ on a $F_{X_n}(t^-_\varepsilon) < u + \eta$ et $F_{X_n}(t^+_\varepsilon) > u - \eta$ pour n assez grand. Comme $F_{X_n}(t^-_\varepsilon) < u + \eta$, on a par définition de la fonction quantile $F^{\leftarrow}_{X_n}(u+\eta) \geq t^-_\varepsilon \geq t - \varepsilon$. Comme $F_{X_n}(t^+_\varepsilon) > u - \eta$, de même $t + \varepsilon \geq t^+_\varepsilon \geq F^{\leftarrow}_{X_n}(u - \eta)$. Comme ε est arbitraire, on obtient

$\liminf F_{X_n}^{\leftarrow}(u+\eta) \geq t = F_X^{\leftarrow}(u)$ et $\limsup F_{X_n}^{\leftarrow}(u-\eta) \leq t = F_X^{\leftarrow}(u)$. Les points de discontinuité sont de mesure de Lebesgue nulle car dénombrables, on a donc bien la convergence ps.
$3 \Rightarrow 1$ Soit ϕ une fonction continue bornée. Comme elle est bornée, on peut utiliser le théorème de convergence dominée. On a

$$\begin{aligned}
\int \phi(X_n) d\mathbb{P}_n &= \int \phi(X'_n) d\lambda \\
&\to \int \phi(X') d\lambda \\
&= \int \phi(X) d\mathbb{P}.
\end{aligned}$$

d'où la convergence en loi. □

Attention : on a changé d'espace et de variables aléatoires pour obtenir la convergence ps. La convergence en loi n'implique pas la convergence ps des variables a aléatoires d'origine. Celle-ci n'a même pas de sens si elles sont définies sur des espaces différents.

Proposition 4.2.2. *Pour tout $n \geq 1$, soit X_n une variable aléatoire définies sur $(\Omega_n, \mathcal{F}_n, \mathbb{P}_n)$ et X définie sur $(\Omega, \mathcal{F}, \mathbb{P})$. On suppose que toutes ces variables aléatoires sont à valeurs dans $\mathbb{Z}$. Alors (X_n) converge en loi vers X si et seulement si pour tout $k \in \mathbb{Z}$, $\mathbb{P}(X_n = k) \to \mathbb{P}(X = k)$.*

Démonstration. Comme on est en dimension 1, on peut utiliser la caractérisation de la convergence en loi par les fonctions de répartition. Pour tout $0 < \varepsilon < 1$ et $k \in \mathbb{Z}$, $k + \varepsilon$ est un point de continuité de F_X et $F_X(k+\varepsilon) = F_X(k)$. On a donc

$$\begin{aligned}
\lim \mathbb{P}(X_n = k) &= \lim_{n\to\infty} F_{X_n}(k) - F_{X_n}(k-1) \\
&= \lim_{n\to\infty} F_{X_n}(k+\varepsilon) - F_{X_n}(k-1+\varepsilon) \\
&= F_X(k+\varepsilon) - F_X(k-1+\varepsilon) \\
&= F_X(k) - F_X(k-1) \\
&= \mathbb{P}(X = k),
\end{aligned}$$

d'où la convergence. Pour la réciproque, tout point de continuité de F_X est de la forme $k + \varepsilon$ avec $0 < \varepsilon < 1$. On a donc

$$\begin{aligned}
\lim_{n\to\infty} F_{X_n}(k+\varepsilon) &= \lim_{n\to\infty} \sum_{j \leq k} \mathbb{P}(X_n = j) \\
&= \sum_{j \leq k} \lim_{n\to\infty} \mathbb{P}(X_n = j) \\
&= \sum_{j \leq k} \mathbb{P}(X = j) \\
&= F_X(k) \\
&= F_X(k+\varepsilon)
\end{aligned}$$

On a pu intervertir limite et séries par convergence monotone. □

Exemple 4.2.1. La loi Binomiale de paramètres $(n, \lambda/n)$ converge en loi vers la loi de Poisson de paramètre λ.

4.3 Comparaison de la convergence en loi avec les autres convergences

Cette comparaison n'est possible que si les variables aléatoires sont toutes définies sur le même espace probabilisé $(\Omega, \mathcal{F}, \mathbb{P})$.

Proposition 4.3.1. *Soit (X_n) une suite de variables aléatoires définies sur le même espace probabilisé $(\Omega, \mathcal{F}, \mathbb{P})$ et X définie sur $(\Omega, \mathcal{F}, \mathbb{P})$. Si la suite (X_n) converge ps vers X alors elle converge aussi en loi vers X.*

Démonstration. Soit ϕ une fonction continue bornée. Pour tout $\varepsilon > 0$, on a

$$\int \phi d\mathbb{P}_{X_n} \quad = \quad \mathbb{E}[\phi(X_n)] \to \mathbb{E}[\phi(X)]$$

par convergence dominée. □

Contre-Exemple 4.3.1. Soit X une variable aléatoire de loi normale centrée réduite et $X_n = (-1)^n X$. Alors la loi de X_n est encore la loi $\mathcal{N}(0,1)$ puisqu'elle est symétrique, donc la suite (X_n) converge en loi vers X, mais (X_n) ne converge pas ps vers X.

Proposition 4.3.2. *Soit (X_n) une suite de variables aléatoires définies sur le même espace probabilisé $(\Omega, \mathcal{F}, \mathbb{P})$ et X définie sur $(\Omega, \mathcal{F}, \mathbb{P})$. Si la suite (X_n) converge en probabilité vers X alors elle converge aussi en loi vers X.*

Démonstration. Soit ϕ une fonction continue bornée. Pour tout $\varepsilon > 0$, on a

$$\begin{aligned}
\Big| \int \phi d\mathbb{P}_{X_n} - \int \phi d\mathbb{P}_X \Big| &= |\mathbb{E}[\phi(X_n)] - \mathbb{E}[\phi(X)]| \\
&= |\mathbb{E}[(\phi(X_n) - \phi(X))(\mathbb{1}_{|\phi(X_n)-\phi(X)|\leq\varepsilon} + \mathbb{1}_{|\phi(X_n)-\phi(X)|>\varepsilon})]| \\
&\leq \varepsilon + 2|\phi|\mathbb{P}(|\phi(X_n) - \phi(X)| > \varepsilon).
\end{aligned}$$

Comme $X_n \xrightarrow{\mathbb{P}} X$ et que ϕ est continue, $\phi(X_n) \xrightarrow{\mathbb{P}} \phi(X)$ et le dernier terme de l'inégalité est plus petit que 2ε pour n assez grand. Comme ε est arbitraire, on a bien la convergence en loi de (X_n) vers X. □

Contre-Exemple 4.3.2. La réciproque est fausse. Soit X une variable aléatoire de loi de Bernoulli de paramètre $1/2$. Pour tout n, on pose $X_n = X$. On a donc $X_n \xrightarrow{\mathcal{L}} X$. La variable aléatoire $Y = 1 - X$ suit encore une loi de Bernoulli de paramètre $1/2$, donc on a aussi $X_n \xrightarrow{\mathcal{L}} Y$. Or $|X_n - Y| = |2X - 1| = 1$ ps, donc pour tout $0 < \varepsilon < 1$, $\mathbb{P}(|X_n - Y| > \varepsilon) = 1$ et la suite ne converge pas en proba vers Y.

On a une réciproque partielle dans le cas particulier où la limite n'est pas aléatoire.

Proposition 4.3.3. *Soit (X_n) une suite de variables aléatoires définies sur le même espace probabilisé $(\Omega, \mathcal{F}, \mathbb{P})$. Si la suite (X_n) converge en loi vers une ($\mathbb{P}$-ps) constante a, alors elle converge aussi en probabilité vers a.*

Démonstration. Soit $B_f(a, \varepsilon)$ la boule fermé de centre a et de rayon ε. On a $\delta_a(\partial B_f(a, \varepsilon)) = 0$, donc d'après le lemme $\lim \mathbb{P}_{X_n}(B_f(a, \varepsilon)) = \delta_a(B_f(a, \varepsilon)) = 1$, donc $\lim \mathbb{P}(|X_n - a| > \varepsilon) = 0$, d'où la convergence en proba. □

4.4 Théorème central limite

Théorème 4.4.1. *Soit* (X_n) *une suite de variables aléatoires définies sur le même espace probabilisé* $(\Omega, \mathcal{F}, \mathbb{P})$*, indépendantes, de même loi et de carré intégrable. Soit*

$$Y_n = \frac{1}{\sqrt{n}} \sum_{k=1}^{n} (X_k - \mathbb{E}[X_k]).$$

Alors la suite (Y_n) *converge en loi vers* $\mathsf{Var}(X_1)^{1/2} Y$ *où* Y *suit la loi normale* $\mathcal{N}(0, I_d)$*. Dans le cas particulier des variables réelles,*

$$Y'_n = \frac{1}{\sqrt{n\,\mathsf{Var}(X_1)}} \sum_{k=1}^{n} (X_k - \mathbb{E}[X_k])$$

converge en loi vers $\mathcal{N}(0,1)$*.*

Démonstration. Commençons par calculer la fonction caractéristique de $\mathsf{Var}(X_1)^{1/2}Y$ en utilisant la symétrie de $\mathsf{Var}(X_1)$

$$\begin{aligned}\Phi_{\mathsf{Var}(X_1)^{1/2}Y}(u) &= \mathbb{E}[e^{i\langle u, \mathsf{Var}(X_1)^{1/2}Y\rangle}] = \mathbb{E}[e^{i\langle \mathsf{Var}(X_1)^{1/2}u, Y\rangle}] \\ &= e^{-\frac{1}{2}\langle \mathsf{Var}(X_1)^{1/2}u, \mathsf{Var}(X_1)^{1/2}u\rangle} = e^{-\frac{1}{2}\langle \mathsf{Var}(X_1)u, u\rangle}.\end{aligned}$$

On utilise le théorème de Lévy et le développement limité à l'ordre 2 des fonctions caractéristiques de loi de carré intégrable. Comme les X_n sont indépendantes et de même loi, la fonction caractéristique de Y_n est

$$\Phi_{Y_n}(t) = \prod_{k=1}^{n} \Phi_{X_k - \mathbb{E}[X_k]}(t/\sqrt{n}) = \left(\Phi_{X_1 - \mathbb{E}[X_1]}(t/\sqrt{n})\right)^n = \left(\Phi_{\langle X_1 - \mathbb{E}[X_1], t\rangle}(1/\sqrt{n})\right)^n$$

La variable réelle centrée $Z = \langle X_1 - \mathbb{E}[X_1], t\rangle$ a un moment d'ordre 2. On peut donc faire un développement limité de sa fonction caractéristique. On a

$$\begin{aligned}\Phi_{Y_n}(t) &= \left(\Phi_Z(1/\sqrt{n})\right)^n \\ &= \left(1 + i/\sqrt{n}\mathbb{E}[Z] - 1/2n\mathbb{E}[Z^2] + o(1/n)\right)^n \\ &= \left(1 - 1/2n\mathbb{E}[Z^2] + o(1/n)\right)^n \\ &\to \exp(-1/2\mathbb{E}[Z^2]).\end{aligned}$$

Or $\mathbb{E}[Z^2] = \mathbb{E}[\langle X_1 - \mathbb{E}[X_1], t\rangle^2] = \langle t, \mathsf{Var}(X_1)t\rangle$. On reconnaît la fonction caractéristique de $\mathsf{Var}(X_1)^{1/2}Y$. D'où la convergence en loi par le théorème de Lévy. $\square$

Exemple 4.4.1. Ce résultat donne un ordre de grandeur de la vitesse de convergence de la loi des grands nombres, même s'il s'agit ici d'une convergence plus faible puisque seulement en loi. En effet, la loi de grands nombres dit que S_n/n converge vers $\mathbb{E}[X_1]$ et on peut récrire Y_n comme

$$Y_n = \frac{1}{\sqrt{n}}\left(\sum_{k=1}^{n} X_k\right) - \frac{n}{\sqrt{n}}\mathbb{E}[X_1] = \frac{1}{\sqrt{n}}(S_n) - \sqrt{n}\mathbb{E}[X_1] = \sqrt{n}(S_n/n - \mathbb{E}[X_1]).$$

La vitesse de convergence est donc de l'ordre de $\sqrt{n}$, avec une constante proportionnelle à l'écart-type.

Exemple 4.4.2. On peut donc dire que Y'_n suit approximativement la loi $\mathcal{N}(0,1)$ dans le cas réel, ou bien que S_n/n suit approximativement la loi $\mathcal{N}(\mathbb{E}[X_1], 1/n)$. Comme la loi $\mathcal{N}(0,1)$ est tabulée, ce résultat permet de faire de nombreux calculs approximatifs.

4.5 Application du théorème central limite à l'estimation par intervalles

On se replace dans le cadre d'une estimation paramétrique. On cherche à estimer le paramètre θ d'une loi μ sur $\mathbb{R}$. Maintenant, on ne cherche plus à donner une valeur à θ, mais à donner une fourchette qui contient θ avec une grande probabilité.

Définition 4.5.1. *Un **intervalle de confiance** de probabilité de confiance α construit à partir de l'échantillon aléatoire $(X_1, \ldots, X_n)$ est un intervalle de la forme $[f(X_1, \ldots, X_n); g(X_1, \ldots, X_n)]$ tel que*

$$\mathbb{P}(f(X_1, \ldots, X_n) \leq \mathbb{E}[X_1] \leq g(X_1, \ldots, X_n)) = \alpha.$$

Proposition 4.5.1. *Pour un échantillon de taille élevée ($n \geq 30$), un intervalle de confiance de probabilité de confiance (asymptotique) α est*

$$\left[\overline{X}_n - F^{\leftarrow}_{\mathcal{N}(0,1)}\left(\frac{1+\alpha}{2}\right)\sqrt{\frac{V(X_1)}{n}}; \overline{X}_n + F^{\leftarrow}_{\mathcal{N}(0,1)}\left(\frac{1+\alpha}{2}\right)\sqrt{\frac{V(X_1)}{n}}\right],$$

où $F^{\leftarrow}_{\mathcal{N}(0,1)}$ est la fonction quantile de la loi $\mathcal{N}(0;1)$. Si la variance $V(X_1)$ est inconnue et si $n \geq 50$, on remplace $V(X_1)$ par S_n^2 dans la formule.

Démonstration. Il suffit d'appliquer le TCL pour obtenir que $\overline{X}_n - \theta$ suit approximativement la loi $\mathcal{N}(0, \mathsf{Var}(X_1)/n)$, et le lemme de Slutsky pour remplacer la variance par un estimateur convergent. □

Exemple 4.5.1. Un sondage Elabe pour **Le Figaro** réalisé du 28 au 31 mars 2019 auprès d'un échantillon représentatif des résident · e · s parisien · ne · sde 999 personnes donnait les intentions de votes suivantes pour le premier tour des élections municipales de 2020, pour un des scénarios proposés

candidat · e	intention de vote
Anne Hidalgo (PS, PCF, RDG)	25%
Cédric Villani (LREM, MODEM)	21%
Florence Berthout (LR)	14,5%
Danielle Simonnet (FI)	8,5%
Julien Bayou (EELV)	8,5%

Soit A un candidat et p la probabilité de déclarer vouloir voter pour A. Le nombre X de candidats se déclarant pour A suit une loi binomial de paramètres (n,p) où n est le nombre total d'individus interrogés. Comme une loi binomiale est une somme de variables aléatoires indépendantes et de même loi de Bernoulli, le théorème central limite donne $\sqrt{n/p(1-p)}(X/n - p)$ suit approximativement une loi normal centrée réduite, donc, pour

$Z \sim \mathcal{N}(0,1)$,

$$\begin{aligned}
\mathbb{P}(|X/n-p|<\beta) &= \mathbb{P}(\sqrt{n/p(1-p)}|X/n-p|<\sqrt{n/p(1-p)}\beta)\\
&\simeq \mathbb{P}(|Z|<\sqrt{n/p(1-p)}\beta)\\
&= \mathbb{P}(Z<\sqrt{n/p(1-p)}\beta)-\mathbb{P}(Z<-\sqrt{n/p(1-p)}\beta)\\
&= 2\Pi(\sqrt{n/p(1-p)}\beta)-1
\end{aligned}$$

D'où $\mathbb{P}(|X_n-p|<\beta)=\alpha \Leftrightarrow 2F_{\mathcal{N}(0,1)}(\sqrt{n/p(1-p)}\beta)-1=\alpha \Leftrightarrow \beta = F^{\leftarrow}_{\mathcal{N}(0,1)}((1+\alpha)/2)\sqrt{p(1-p)/n}$. On a donc

$$\mathbb{P}\Big(p\in\Big[X/n-\Pi^{-1}\Big(\frac{1+\alpha}{2}\Big)\sqrt{\frac{X(n-X)/n^2}{n}};X/n+\Pi^{-1}\Big(\frac{1+\alpha}{2}\Big)\sqrt{\frac{X(n-X)/n^2}{n}}\Big]\Big)=\alpha,$$

en remplaçant p dans la variance par son estimation on a $\frac{X}{n}(1-\frac{X}{n})$. Les intervalles de confiance à 95% du sondage sont ($F^{\leftarrow}_{\mathcal{N}(0,1)}(0.975)=1.96$)

parti	intention de vote	Intervalle de confiance
Anne Hidalgo (PS, PCF, RDG)	25%	[22,3% ; 27,7%]
Cédric Villani (LREM, MODEM)	21%	[18,4% ; 23,5%]
Florence Berthout (LR)	14,5%	[12,3% ; 16,7%]
Danielle Simonnet (FI)	8,5%	[6,7% ; 10,2%]
Julien Bayou (EELV)	8,5%	[6,7% ; 10,2%]

A 95% de chance, le sondage ne permet pas ici de départager les deux candidat · e · s de tête, malgré un écart de 4 points de l'estimation ponctuelle.

Bibliographie

[Bil99] Patrick Billingsley. *Convergence of probability measures.* Wiley Series in Probability and Statistics : Probability and Statistics. John Wiley & Sons, Inc., New York, second edition, 1999. A Wiley-Interscience Publication.

[Bil12] Patrick Billingsley. *Probability and measure.* Wiley Series in Probability and Statistics. John Wiley & Sons, Inc., Hoboken, NJ, 2012. Anniversary edition [of MR1324786], With a foreword by Steve Lalley and a brief biography of Billingsley by Steve Koppes.

[BL07] Philippe Barbe and Michel Ledoux. *Probabilités.* Collection enseignement sup, Mathématiques. EDP Science, 2007.

[Chu01] Kai Lai Chung. *A course in probability theory.* Academic Press, Inc., San Diego, CA, third edition, 2001.

[Fel68] William Feller. *An introduction to probability theory and its applications. Vol. I.* Third edition. John Wiley & Sons Inc., New York, 1968.

[FF03] Dominique Foata and André Fuchs. *Calcul des probabilités.* Dunod, second edition, 2003.

[Ouv08] Jean-Yves Ouvrard. *Probabilités,Tomes 1 et 2.* Cassini, 2008.