

Chapitre 1

Notions fondamentales de la Théorie des Probabilités

Dans ce chapitre, nous allons donner les définitions de base concernant la Théorie des Probabilités.

Il faut plutôt voir ce cours plus comme un cours de techniques probabilistes que comme un cours de Probabilités : on ne cherchera pas à donner d'interprétation "concrète" des notions et des résultats présentés dans ce cours. C'est en quelque sorte le "minimum vital de Probabilités" que doit posséder tout mathématicien.

1 Espace de probabilité

Une **probabilité** (ou **mesure de probabilité**) est une *mesure positive* \mathbb{P} sur un espace mesurable (Ω, \mathcal{A}) telle que $\mathbb{P}(\Omega) = 1$. On dit aussi **loi de probabilité**. Le triplet $(\Omega, \mathcal{A}, \mathbb{P})$ est appelé **espace de probabilité**.

La Théorie des Probabilités utilise la Théorie de la mesure, mais, pour des raisons historiques, a un langage qui lui est propre (ce qui est d'ailleurs peut-être à l'origine de la réticence, voire de l'ostracisme, dont la majorité des mathématiciens ont fait preuve par le passé, et encore parfois aujourd'hui, envers les Probabilités). On a le dictionnaire suivant :

notation	point de vue ensembliste	point de vue probabiliste
$\omega \in \Omega$	point de Ω	<i>observation</i>
$A \in \mathcal{A}$	partie mesurable	événement
$\omega \in A$	ω est dans la partie A	l'événement A est réalisé par ω
$A \subseteq B$	A est contenu dans B	A implique B
\emptyset	ensemble vide	événement impossible
Ω	ensemble plein	événement certain
$A \cup B$	réunion de A et B	A ou B
$A \cap B$	intersection de A et B	A et B

Si l'on considère qu'un espace de probabilité est la modélisation d'une expérience, un point ω de Ω peut aussi être vu comme le *résultat*, ou la *réalisation*, de l'expérience modélisée par $(\Omega, \mathcal{A}, \mathbb{P})$. On dit aussi que ω est une *éventualité*. Par rapport au point de vue ensembliste, dans lequel la notion de base est le point, qui peut ou non appartenir à une partie donnée, dans le point de vue probabiliste, il faut plutôt inverser les rôles : la notion de base est l'ensemble des événements, et chaque événement peut être réalisé ou non par une observation.

Pour deux événements A et B , dire que $\omega \in A \cup B$ s'exprime en disant que ω réalise l'événement A ou l'événement B , et $\omega \in A \cap B$ en disant que ω réalise l'événement A et l'événement B .

Comme \mathcal{A} est une tribu, si l'on a une famille dénombrable d'événements $A_n, n \geq 1$, alors les ensembles suivants sont des événements :

$\bigcup_{n \geq 1} A_n$: on a $\omega \in \bigcup_{n \geq 1} A_n$ si l'**un au moins** des événements A_n est réalisé par ω .

$\bigcap_{n \geq 1} A_n$: on a $\omega \in \bigcap_{n \geq 1} A_n$ si **tous** les événements A_n sont réalisés par ω .

$\overline{\lim}_{n \rightarrow \infty} A_n$: on a $\omega \in \overline{\lim}_{n \rightarrow \infty} A_n = \bigcap_{n \geq 1} \bigcup_{k \geq n} A_k$ si **une infinité** des événements A_n sont réalisés par ω .

$\underline{\lim}_{n \rightarrow \infty} A_n$: on a $\omega \in \underline{\lim}_{n \rightarrow \infty} A_n = \bigcup_{n \geq 1} \bigcap_{k \geq n} A_k$ si **tous** les événements A_n , **sauf un nombre fini**, sont réalisés par ω .

On a bien sûr $\underline{\lim}_{n \rightarrow \infty} A_n \subseteq \overline{\lim}_{n \rightarrow \infty} A_n$.

Le résultat suivant, bien que très facile, est extrêmement utile.

Théorème 1.1 (Lemme de Borel-Cantelli)

Si $\sum_{n=1}^{\infty} \mathbb{P}(A_n) < +\infty$, alors on a : $\mathbb{P}(\overline{\lim}_{n \rightarrow \infty} A_n) = 0$.

Définition 1.2 Pour tout événement A , on dit que :

- A est **presque impossible** si $\mathbb{P}(A) = 0$;
- A est **presque sûr**, ou **presque certain**, si $\mathbb{P}(A) = 1$.

La notion de "presque sûr" est la version probabiliste de la notion de "presque partout". Lorsque l'événement A est l'ensemble des ω pour lesquels une certaine propriété est vraie, on dit que cette **propriété** est **presque sûrement vraie** si A est presque sûr. On notera en abrégé *p.s.* pour "presque sûrement".

Comme $(\overline{\lim}_{n \rightarrow \infty} A_n)^c = \underline{\lim}_{n \rightarrow \infty} (A_n^c)$, le lemme de Borel-Cantelli dit que, sous l'hypothèse $\sum_{n \geq 1} \mathbb{P}(A_n) < +\infty$, on a $\underline{\lim}_{n \rightarrow \infty} (A_n^c)$ presque sûrement ; par conséquent, le lemme de Borel-Cantelli s'exprime de la façon plus utile suivante :

Théorème 1.3 (Lemme de Borel-Cantelli) Si $\sum_{n=1}^{\infty} \mathbb{P}(B_n^c) < +\infty$, alors presque sûrement, tous les événements B_n sont réalisés, à l'exception d'un nombre fini.

Preuve 1. Pour tout $n \geq 1$, on a $\mathbb{P}\left(\bigcup_{k \geq n} A_k\right) \leq \sum_{k \geq n} \mathbb{P}(A_k)$; donc, puisque la suite $\left(\bigcup_{k \geq n} A_k\right)_{n \geq 1}$ est décroissante :

$$\mathbb{P}\left(\overline{\lim}_{n \rightarrow \infty} A_n\right) = \lim_{n \rightarrow \infty} \downarrow \mathbb{P}\left(\bigcup_{k \geq n} A_k\right) \leq \lim_{n \rightarrow \infty} \sum_{k \geq n} \mathbb{P}(A_k) = 0,$$

puisque la série converge. \square

Preuve 2. La condition $\sum_{n \geq 1} \mathbb{P}(A_n) < +\infty$ signifie que la fonction $\sum_{n \geq 1} \mathbb{1}_{A_n}$ est \mathbb{P} -intégrable; elle est donc finie \mathbb{P} -presque partout, ce qui est le résultat annoncé. \square

2 Variables aléatoires

Définition 2.1 On appelle **variable aléatoire**, en abrégé $\boxed{v.a.}$ (vectorielle, à valeurs dans \mathbb{R}^d), toute application mesurable $X : (\Omega, \mathcal{A}) \rightarrow (\mathbb{R}^d, \mathcal{B}or(\mathbb{R}^d))$.

Pour $d = 1$, on autorise X à prendre ses valeurs dans $\overline{\mathbb{R}}$; elle doit donc être mesurable pour $\mathcal{B}or(\overline{\mathbb{R}})$. On dit que c'est une **variable aléatoire réelle**; en abrégé $\boxed{v.a.r.}$.

On conviendra par la suite que $\mathbb{R}^1 = \overline{\mathbb{R}}$.

Si $d > 1$, on dira aussi **vecteur aléatoire**.

Si $X = (X_1, \dots, X_d)$ est un vecteur aléatoire, alors les X_k sont des *v.a.r.*; réciproquement, si X_1, \dots, X_d sont des *v.a.r.*, à valeurs dans \mathbb{R} , alors (X_1, \dots, X_d) est un vecteur aléatoire.

Remarque. Il faut noter que, contrairement à ce que peut laisser croire le nom, une variable aléatoire est une fonction, et non pas une variable. Cette terminologie trompeuse est, là encore, due à des raisons historiques, datant de l'époque où l'on avait pas conscience de la distinction entre une fonction et sa valeur en un point.

Terminologie. On dit que la *v.a.r.* X a un *moment d'ordre* p , $1 \leq p < \infty$, si $X \in L^p(\Omega, \mathcal{A}, \mathbb{P})$. Lorsque X est intégrable (on utilisera plutôt ce terme que celui de moment d'ordre 1), son intégrale est appelée **espérance** (*mathématique*) de X ; on dit aussi que c'est la **moyenne** de X . On la note :

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

On dit que la **variable aléatoire** X est **centrée** lorsque $\mathbb{E}(X) = 0$. La *v.a.r.* $X - \mathbb{E}(X)$ (mis pour $X - \mathbb{E}(X)\mathbf{1}$) est appelée la *v.a. centrée* associée à X ; elle est évidemment centrée.

Lorsque $X \in L^2(\Omega, \mathcal{A}, \mathbb{P})$, on définit la **variance** de X par :

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

Ce nombre (qui est positif) mesure les variations de X autour de sa moyenne. On la calcule habituellement par la formule suivante :

Proposition 2.2 $\boxed{\text{Var}(X) = \mathbb{E}(X^2) - [\mathbb{E}(X)]^2}$.

Preuve. Il suffit de développer :

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

car $\mathbb{E}(\mathbf{1}) = 1$. □

Notons que l'on a $\boxed{\text{Var}(aX) = a^2\text{Var}(X)}$ pour tout $a \in \mathbb{R}$. Pour revenir à l'ordre de grandeur initial, on utilise l'**écart-type** :

$$\boxed{\sigma_X = \sqrt{\text{Var}(X)}}.$$

Autrement dit : $\sigma_X = \|X - \mathbb{E}(X)\|_{L^2(\Omega, \mathcal{A}, \mathbb{P})}$.

Le nombre $\|X\|_{L^2(\mathbb{P})} = [\mathbb{E}(X^2)]^{1/2}$ est appelé *moyenne quadratique* de X .

Lorsque l'on mesure les résultats d'une expérience aléatoire par une variable aléatoire X , l'espace de probabilité $(\Omega, \mathcal{A}, \mathbb{P})$ est une *modélisation* de la situation liée à cette expérience. Cette modélisation peut être faite de diverses façons, ce qui veut dire que l'espace $(\Omega, \mathcal{A}, \mathbb{P})$, et en particulier la probabilité \mathbb{P} , n'a pas de sens intrinsèque. Ce qui est important, c'est la répartition des valeurs que peut prendre X , et leur fréquence. C'est pourquoi le concept suivant, qui précise ces fréquences d'apparitions, est **fondamental** ; d'ailleurs avant l'axiomatisation de la Théorie des Probabilités par Kolmogorov en 1933, les Probabilistes ne voyaient les variables aléatoires qu'à travers leur loi.

Définition 2.3 *On appelle loi de la variable aléatoire X la mesure-image de la probabilité \mathbb{P} par X . On la note \mathbb{P}_X . On dit aussi loi de probabilité de X . On l'appelle encore la distribution de la variable aléatoire X .*

Cela signifie que pour tout borélien B de \mathbb{R}^d , on a :

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

$\mathbb{P}(X \in B)$ étant la notation usuelle pour :

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

\mathbb{P}_X est donc une mesure de probabilité sur \mathbb{R}^d . Si Q est une mesure de probabilité sur \mathbb{R}^d , on note $X \sim Q$ pour exprimer que X a la loi Q , c'est-à-dire que $Q = \mathbb{P}_X$; on a donc $X \sim \mathbb{P}_X$.

Si la loi de X a une densité f_X par rapport à la mesure de Lebesgue, on dit que X possède la densité f_X .

Rappelons que pour toute fonction $f: \mathbb{R}^d \rightarrow \overline{\mathbb{R}}$ borélienne positive ou \mathbb{P}_X -intégrable, on a :

$$\int_{\mathbb{R}^d} f(x) d\mathbb{P}_X(x) = \int_{\Omega} f[X(\omega)] d\mathbb{P}(\omega),$$

soit :

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

où $f(X)$ est la notation habituelle pour la composée $f \circ X$.

Ainsi, lorsque X est une *v.a.r.*, X est intégrable si et seulement si $\mathbb{P}(\{X = -\infty\}) = \mathbb{P}(\{X = +\infty\}) = 0$ et :

$$\int_{\mathbb{R}} |x| d\mathbb{P}_X(x) < +\infty;$$

et alors :

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

On notera que, X étant intégrable, elle est presque sûrement finie, c'est-à-dire presque sûrement à valeurs dans \mathbb{R} (soit $\mathbb{P}_X(\{-\infty, +\infty\}) = 0$); on peut donc intégrer sur \mathbb{R} au lieu d'intégrer sur $\overline{\mathbb{R}}$.

Pour une expérience mesurée par une *v.a.* X , il est important de savoir quels sont les événements liés à cette expérience.

Définition 2.4 On appelle **tribu engendrée par la variable aléatoire X** la plus petite sous-tribu \mathcal{A}_X de \mathcal{A} pour laquelle X reste mesurable.

On la note aussi $\sigma(X)$ (à ne pas confondre avec l'écart-type!).

Proposition 2.5 On a : $\mathcal{A}_X = \{X^{-1}(B); B \in \mathcal{B}or(\mathbb{R}^d)\}$.

Preuve. L'ensemble écrit à droite doit évidemment être contenu dans \mathcal{A}_X ; c'est d'autre part une tribu, et X est mesurable pour cette tribu. \square

Théorème 2.6 (Doob) Une fonction $Y : \Omega \rightarrow \mathbb{R}^d$ est une variable aléatoire \mathcal{A}_X -mesurable si et seulement s'il existe une fonction borélienne $f : \mathbb{R}^d \rightarrow \mathbb{R}^d$ telle que $Y = f(X)$.

Preuve. La condition suffisante est évidente. Pour la condition nécessaire, en regardant coordonnée par coordonnée, on peut supposer Y réelle ; et en écrivant $Y = Y^+ - Y^-$, la supposer *positive*. Mais alors, comme elle est \mathcal{A}_X -mesurable, elle est limite d'une suite croissante de fonctions Y_n \mathcal{A}_X -étagées. D'après la proposition précédente, ces fonctions \mathcal{A}_X -étagées peuvent s'écrire :

$$Y_n = \sum_{k=1}^{k_n} na_{k,n} \mathbb{1}_{X^{-1}(B_{k,n})} = \sum_{k=1}^{k_n} na_{k,n} (\mathbb{1}_{B_{k,n}} \circ X) = \left(\sum_{k=1}^{k_n} a_{k,n} \mathbb{1}_{B_{k,n}} \right) \circ X,$$

avec $a_{k,n} \geq 0$ et les $B_{k,n}$ boréliens de \mathbb{R}^d .

Posons :

$$f_n = \sum_{k=1}^{k_n} a_{k,n} \mathbb{1}_{B_{k,n}} \quad \text{et} \quad f = \overline{\lim}_{n \rightarrow \infty} f_n.$$

On a bien

$$Y = \lim_{n \rightarrow \infty} \uparrow Y_n = \lim_{n \rightarrow \infty} \uparrow f_n(X) = \overline{\lim}_{n \rightarrow \infty} f_n(X) = f(X). \quad \square$$

Lorsque X est une *v.a.r.*, à valeurs dans \mathbb{R} , on utilise souvent la notion suivante.

Définition 2.7 La fonction de répartition de X est définie, pour tout $x \in \mathbb{R}$, par :

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

On dit aussi que F_X est la fonction de répartition de \mathbb{P}_X .

C'est une fonction $F_X : \mathbb{R} \rightarrow [0, 1]$, qui est croissante, continue à droite, et telle que $\lim_{x \rightarrow -\infty} F_X(x) = 0$ et $\lim_{x \rightarrow +\infty} F_X(x) = 1$.

Le Théorème d'unicité des mesures dit qu'il n'y a qu'une seule mesure de probabilité sur \mathbb{R} dont F_X soit la fonction de répartition : F_X détermine \mathbb{P}_X .

3 Exemples usuels de lois de probabilité sur \mathbb{R}

Dans ce cours, on n'utilisera essentiellement que deux lois de probabilité : la loi de Bernoulli et la loi gaussienne, avec en plus, la loi uniforme.

On se contentera donc de parler de celles-ci, renvoyant, pour les autres lois usuelles, aux exercices.

3.1 Loi de Bernoulli

C'est la loi d'une *v.a.r.* qui ne prend (presque sûrement) que les deux valeurs 0 et 1. Si X est une telle *v.a.r.*, sa loi est déterminée par le nombre $p = \mathbb{P}(X = 1)$; on a $0 < p < 1$, et p est appelé le *paramètre* de la loi. On a :

$$\boxed{\mathbb{P}(X = 1) = p \quad \text{et} \quad \mathbb{P}(X = 0) = 1 - p = q},$$

ou $\mathbb{P}_X(\{1\}) = p$, $\mathbb{P}_X(\{0\}) = 1 - p$; la loi de X est donc $\mathbb{P}_X = p\delta_1 + (1 - p)\delta_0$, où δ_1 et δ_0 sont les mesures de Dirac en 1 et en 0. C'est la loi du "pile ou face" : on a la probabilité p de réussir.

On a $\boxed{\mathbb{E}(X) = p}$ et $\boxed{\text{Var}(X) = pq = p(1 - p)}$; en effet, $\mathbb{E}(X) = 0 \cdot (1 - p) + 1 \cdot p = p$ et $\mathbb{E}(X^2) = 0^2 \cdot (1 - p) + 1^2 \cdot p = p$, donc $\text{Var}(X) = p - p^2$.

3.2 Loi de Gauss

On dit aussi que c'est la **loi de Laplace** ou la **loi normale**.

Elle est caractérisée par deux paramètres $m \in \mathbb{R}$ et $\sigma > 0$. On la note :

$$\boxed{\mathcal{N}(m, \sigma^2)}.$$

C'est la loi de probabilité sur \mathbb{R} qui a pour densité (par rapport à la mesure de Lebesgue) :

$$\boxed{\gamma_{m, \sigma^2} = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{x - m}{\sigma}\right)^2\right]}.$$

Rappelons que $\int_{\mathbb{R}} e^{-u^2/2} du = \sqrt{2\pi}$, et donc γ_{m, σ^2} est bien la densité d'une loi de probabilité

Si $X \sim \mathcal{N}(m, \sigma^2)$, on dit que X est une **variable aléatoire gaussienne**.

On a donc, pour tout borélien B de \mathbb{R} :

$$\mathbb{P}(X \in B) = \frac{1}{\sigma\sqrt{2\pi}} \int_B \exp\left[-\frac{1}{2}\left(\frac{x - m}{\sigma}\right)^2\right] dx.$$

En particulier, sa fonction de répartition est :

$$F_X(x) = \mathbb{P}(X \leq x) = \int_{-\infty}^x \gamma_{m, \sigma^2}(x) dx.$$

On a :

$$\boxed{\mathbb{E}(X) = m} \quad \text{et} \quad \boxed{\text{Var}(X) = \sigma^2}.$$

En effet :

$$\begin{aligned} \mathbb{E}(X) &= \frac{1}{\sigma\sqrt{2\pi}} \int_{\mathbb{R}} x \exp\left[-\frac{1}{2}\left(\frac{x - m}{\sigma}\right)^2\right] dx \\ &= \frac{1}{\sqrt{2\pi}} \left[\sigma \int_{\mathbb{R}} u e^{-u^2/2} du + m \int_{\mathbb{R}} e^{-u^2/2} du \right] \\ &= m, \end{aligned}$$

et :

$$\begin{aligned}\text{Var}(X) &= \frac{1}{\sigma\sqrt{2\pi}} \int_{\mathbb{R}} (x-m)^2 \exp\left[-\frac{1}{2}\left(\frac{x-m}{\sigma}\right)^2\right] dx \\ &= \frac{\sigma^2}{\sqrt{2\pi}} \int_{\mathbb{R}} u^2 e^{-u^2/2} du = \sigma^2,\end{aligned}$$

en écrivant $u^2 e^{-u^2/2} = u \cdot u 2e^{-u^2/2}$ et en intégrant par parties.

Ainsi donc la loi d'une gaussienne est déterminée par sa moyenne m et sa variance σ^2 .

La loi $\mathcal{N}(0, 1)$ est appelée **loi normale centrée réduite**; elle a pour densité :

$$\gamma(x) = \gamma_{0,1}(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2};$$

son espérance est nulle et sa variance est 1.

On a :

$$X \sim \mathcal{N}(m, \sigma^2) \iff Z = \frac{X-m}{\sigma} \sim \mathcal{N}(0, 1);$$

ou, inversement :

$$Z \sim \mathcal{N}(0, 1) \iff X = \sigma Z + m \sim \mathcal{N}(m, \sigma^2).$$

Cette loi est *centrale* en Probabilités, car on verra que, en un certain sens, *toutes* les lois tendent vers la loi de Gauss.

3.3 Loi uniforme

La **loi uniforme** sur le segment $[a, b]$ est la loi ayant pour densité la fonction :

$$\frac{1}{b-a} \mathbb{1}_{(a,b]}$$

par rapport à la mesure de Lebesgue sur \mathbb{R} .

Pour tout intervalle $I \subseteq \mathbb{R}$, on a, pour une *v.ar.* ayant cette loi :

$$\mathbb{P}_X(I) = \mathbb{P}(X \in I) = \int_I \frac{1}{b-a} \mathbb{1}_{(a,b]} d\lambda = \frac{\lambda(I \cap [a, b])}{b-a}.$$

4 Indépendance

C'est avec la notion d'indépendance que la Théorie des Probabilités prend son autonomie par rapport à la Théorie de la mesure.

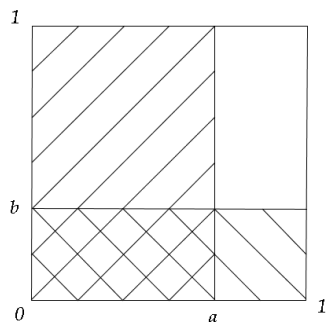
Définition 4.1 Deux événements A et B sont dits **indépendants** si :

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

On note $A \perp\!\!\!\perp B$.

L'exemple suivant est fondamental, car, comme nous le verrons, l'indépendance peut toujours se ramener, en un certain sens, à une mesure de probabilité qui est une mesure-produit (voir Théorème 4.7).

Exemple fondamental. Prenons $\Omega = [0, 1] \times [0, 1]$, muni de sa tribu borélienne, et prenons pour probabilité $\mathbb{P} = \lambda_2 = \lambda \otimes \lambda$, la mesure de Lebesgue 2-dimensionnelle, qui est aussi le produit de la mesure de Lebesgue 1-dimensionnelle par elle-même. Soit :



$$A = [0, a] \times [0, 1] \text{ et } B = [0, 1] \times [0, b].$$

$$\text{On a } \mathbb{P}(A) = a \text{ et } \mathbb{P}(B) = b;$$

$$\text{et, comme } (A \cap B = [0, a] \times [0, b]) :$$

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

Autrement dit, l'événement " $0 \leq x \leq a$ " est indépendant de l'événement " $0 \leq y \leq b$ ", ou : choisir au hasard un x dans $[0, a]$ est indépendant du choix au hasard d'un y dans $[0, b]$.

Définition 4.2 Soit A_1, \dots, A_n n événements; on dit qu'ils sont **indépendants** si :

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

pour toute partie $\{i_1, \dots, i_k\} \subseteq \{1, \dots, n\}$.

Cela entraîne bien sûr que A_1, \dots, A_n n sont deux-à-deux indépendants.

Définition 4.3 Si $\mathcal{A}_1, \dots, \mathcal{A}_n$ sont n sous-tribus de \mathcal{A} , on dit qu'elles sont **indépendantes** si les événements A_1, \dots, A_n sont indépendants, pour tout choix de $A_1 \in \mathcal{A}_1, \dots, A_n \in \mathcal{A}_n$.

Cela revient à dire que :

$$\mathbb{P}(A_1 \cap \dots \cap A_n) = \mathbb{P}(A_1) \times \dots \times \mathbb{P}(A_n), \quad \forall A_1 \in \mathcal{A}_1, \dots, \forall A_n \in \mathcal{A}_n;$$

en effet, il est clair que c'est entraîné par l'indépendance; réciproquement, si l'on a cette propriété, prenons $J \subseteq \{1, \dots, n\}$ et $A_j \in \mathcal{A}_j$ arbitraire pour $j \in J$; alors, en prenant $A_j = \Omega \in \mathcal{A}_j$ pour tout $j \notin J$, on a $\bigcap_{j=1}^n A_j = \bigcap_{j \in J} A_j$ et l'on obtient $\mathbb{P}(\bigcap_{j \in J} A_j) = \prod_{j \in J} \mathbb{P}(A_j)$, puisque $\mathbb{P}(\Omega) = 1$.

Notons que, étant donnés n événements A_1, \dots, A_n , ils sont indépendants si et seulement si les tribus $\mathcal{A}_k = \{\emptyset, A_k, A_k^c, \Omega\}$ que chacun engendre sont indépendantes.

Définition 4.4 *Etant données n variables aléatoires $X_k: \Omega \rightarrow \mathbb{R}^{d_k}$, $1 \leq k \leq n$, on dit qu'elles sont **indépendantes** si les tribus $\mathcal{A}_{X_1}, \dots, \mathcal{A}_{X_n}$ qu'elles engendrent sont indépendantes. On notera $X \perp\!\!\!\perp Y$ lorsque les v.a. X et Y sont indépendantes.*

Comme :

$$\mathcal{A}_{X_k} = \{X_k^{-1}(B); B \in \mathcal{B}or(\mathbb{R}^{d_k})\},$$

cela revient à dire que, pour tous $B_k \in \mathcal{B}or(\mathbb{R}^{d_k})$, $1 \leq k \leq n$:

$$\mathbb{P}(X_k \in B_k, \forall k \leq n) = \mathbb{P}(X_1 \in B_1) \times \dots \times \mathbb{P}(X_n \in B_n)$$

On notera que n événements A_1, \dots, A_n sont indépendants si et seulement si les v.a.r. $\mathbb{1}_{A_1}, \dots, \mathbb{1}_{A_n}$ le sont, car $\mathcal{A}_{\mathbb{1}_{A_k}} = \{\emptyset, A_k, A_k^c, \Omega\}$.

Remarque. Une v.a. est indépendante d'elle-même si et seulement si elle est constante (presque sûrement), et toute v.a. p.s. constante est indépendante de toutes les autres v.a., comme on pourra le vérifier à titre d'exercice.

Proposition 4.5 *Si les v.a. $X_k: \Omega \rightarrow \mathbb{R}^{d_k}$, $1 \leq k \leq n$, sont indépendantes, alors, pour toutes fonctions boréliennes $f_k: \mathbb{R}^{d_k} \rightarrow \mathbb{R}^{d'_k}$, les v.a. $f_k(X_k)$, $1 \leq k \leq n$, sont encore indépendantes.*

En effet, $f_k(X_k)$ est \mathcal{A}_{X_k} -mesurable; donc $\sigma[f_k(X_k)] \subseteq \sigma(X_k)$. □

Définition 4.6 *Soit $X_1, \dots, X_n: \Omega \rightarrow \mathbb{R}$ des v.a.r.. La loi du n -uplet $X = (X_1, \dots, X_n)$, définissant un vecteur aléatoire $X: \Omega \rightarrow \mathbb{R}^n$, est appelée la **loi conjointe** des v.a.r. X_1, \dots, X_n . Les lois de X_1, \dots, X_n s'appellent les **lois marginales** du v.a. X .*

Le théorème suivant est essentiel, car c'est la plupart du temps en l'utilisant que l'on démontre l'indépendance de v.a.r. car il est très facile d'utilisation.

Théorème 4.7 *Les v.a.r. $X_1, \dots, X_n: \Omega \rightarrow \mathbb{R}$ sont indépendantes si et seulement si leur loi conjointe $\mathbb{P}_{(X_1, \dots, X_n)}$ est égale au **produit** des lois marginales \mathbb{P}_{X_k} :*

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

Bien sûr, on aura de même :

Théorème 4.8 Les v.a. $X_k: \Omega \rightarrow \mathbb{R}^{d_k}$, $1 \leq k \leq n$, sont indépendantes si et seulement si $\mathbb{P}_{(X_1, \dots, X_n)} = \mathbb{P}_{X_1} \otimes \dots \otimes \mathbb{P}_{X_n}$.

Preuve. Pour tous boréliens B_1, \dots, B_n , on a :

$$\begin{aligned} \mathbb{P}(X_k \in B_k, \forall k \leq n) &= \mathbb{P}((X_1, \dots, X_n) \in B_1 \times \dots \times B_n) \\ &= \mathbb{P}_{(X_1, \dots, X_n)}(B_1 \times \dots \times B_n). \end{aligned}$$

Les v.a. X_1, \dots, X_n sont donc indépendantes si et seulement si, pour tous boréliens B_1, \dots, B_n , on a :

$$\mathbb{P}_{(X_1, \dots, X_n)}(B_1 \times \dots \times B_n) = \mathbb{P}_{X_1}(B_1) \times \dots \times \mathbb{P}_{X_n}(B_n).$$

Par définition (et l'unicité) de la mesure-produit, cela signifie que $\mathbb{P}_{(X_1, \dots, X_n)}$ est égale à $\mathbb{P}_{X_1} \otimes \dots \otimes \mathbb{P}_{X_n}$. \square

Corollaire 4.9 Si les variables aléatoires réelles X_1, \dots, X_n sont intégrables et indépendantes, leur produit est intégrable et l'espérance de leur produit est égal au produit de leurs espérances :

$$\mathbb{E}(X_1 \cdots X_n) = \mathbb{E}(X_1) \cdots \mathbb{E}(X_n).$$

Bien sûr, cette égalité est loin de suffire pour avoir l'indépendance.

Preuve. On a :

$$\begin{aligned} \int_{\Omega} |X_1 \cdots X_n| d\mathbb{P} &= \int_{\mathbb{R}^n} |x_1 \cdots x_n| d\mathbb{P}_{(X_1, \dots, X_n)}(x_1, \dots, x_n) \\ &= \int_{\mathbb{R}^n} |x_1 \cdots x_n| d(\mathbb{P}_{X_1} \otimes \dots \otimes \mathbb{P}_{X_n})(x_1, \dots, x_n) \\ &= \left(\int_{\mathbb{R}} |x_1| d\mathbb{P}_{X_1}(x_1) \right) \cdots \left(\int_{\mathbb{R}} |x_n| d\mathbb{P}_{X_n}(x_n) \right) \\ &\quad \text{par le Théorème de Fubini} \\ &= \mathbb{E}(|X_1|) \cdots \mathbb{E}(|X_n|) < +\infty, \end{aligned}$$

ce qui donne l'intégrabilité du produit. Ensuite, le même calcul que ci-dessus, mais sans les valeurs absolues, et en utilisant cette fois-ci le Théorème de Fubini, donne l'égalité $\mathbb{E}(X_1 \cdots X_n) = \mathbb{E}(X_1) \cdots \mathbb{E}(X_n)$. \square

Corollaire 4.10 Si les v.a.r. X_1, \dots, X_n sont de carré intégrable et sont indépendantes, alors :

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

Bien sûr, là aussi, cette égalité est loin de suffire pour avoir indépendance.

Preuve. Il suffit de le montrer pour deux *v.a.r.* indépendantes X et Y . On peut de plus supposer que $\mathbb{E}(X) = \mathbb{E}(Y) = 0$, puisque la variance est la même pour une *v.a.r.* et sa *v.a.r.* centrée. Or, dans ce cas, en utilisant le Corollaire 4.9 :

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

puisque $\mathbb{E}(X) = \mathbb{E}(Y) = 0$. □

Remarque. Comme $\text{Var}(X) = \|X\|_{L^2(\mathbb{P})}^2$ lorsque X est centrée, le Corollaire 4.10 dit que si X et Y sont **centrées et indépendantes**, alors elles sont **orthogonales** dans l'espace de Hilbert $L^2(\mathbb{P})$; ainsi, pour des *v.a.r.* centrées :

$$X \perp\!\!\!\perp Y \implies X \perp Y.$$

Corollaire 4.11 *Si les v.a. X_1, \dots, X_n sont indépendantes et ont des densités f_{X_1}, \dots, f_{X_n} , alors $X = (X_1, \dots, X_n)$ a une densité, définie, pour tout (x_1, \dots, x_n) , par :*

$$f_X(x_1, \dots, x_n) = f_{X_1}(x_1) \cdots f_{X_n}(x_n).$$

Autrement dit, $f_X = f_{X_1} \otimes \cdots \otimes f_{X_n}$.

Il est pratique de pouvoir tester l'indépendance sur une classe plus petite que la tribu entière.

Proposition 4.12 *Si, pour $1 \leq k \leq n$, on a une famille d'événements $\mathcal{C}_k \subseteq \mathcal{A}_k$, qui est stable par intersection finie, contient une suite croissante dont la réunion est Ω , engendrant la tribu $\mathcal{A}_k : \sigma(\mathcal{C}_k) = \mathcal{A}_k$, et si l'on a :*

$$\mathbb{P}\left(\bigcap_{k=1}^n C_k\right) = \prod_{k=1}^n \mathbb{P}(C_k), \quad \forall C_k \in \mathcal{C}_k, 1 \leq k \leq n,$$

alors les tribus $\mathcal{A}_1, \dots, \mathcal{A}_n$ sont indépendantes.

Preuve. Par itération : fixons $C_2 \in \mathcal{C}_2, \dots, C_n \in \mathcal{C}_n$. Les applications :

$$\begin{array}{ccc} \mathcal{A}_1 & \longrightarrow & [0, 1] & \text{et} & \mathcal{A}_1 & \longrightarrow & [0, 1] \\ A & \longmapsto & \mathbb{P}(A \cap \bigcap_{k=2}^n C_k) & & A & \longmapsto & [\prod_{k=2}^n \mathbb{P}(C_k)] \mathbb{P}(A) \end{array}$$

sont des mesures positives bornées qui sont égales sur \mathcal{C}_1 . Par le Théorème d'unicité des mesures, elles sont égales sur \mathcal{A}_1 .

On recommence ensuite, en fixant $A_1 \in \mathcal{A}_1$, et $C_3 \in \mathcal{C}_3, \dots, C_n \in \mathcal{C}_n$, et ainsi de suite. □

Corollaire 4.13 *Les v.a. $X_k : \Omega \rightarrow \mathbb{R}^{d_k}$, $1 \leq k \leq n$, sont indépendantes dès que :*

$$\mathbb{P}(X_k \in D_k, \forall k \leq n) = \prod_{k=1}^n \mathbb{P}(X_k \in D_k), \quad \forall D_k \in \mathcal{D}_k, 1 \leq k \leq n,$$

où chaque $\mathcal{D}_k \subseteq \mathcal{Bor}(\mathbb{R}^{d_k})$ est stable par intersection finie, contient une suite croissante dont la réunion est \mathbb{R}^{d_k} et qui engendre $\mathcal{Bor}(\mathbb{R}^{d_k})$.

Preuve. Il suffit de prendre $\mathcal{C}_k = \{X_k^{-1}(D_k); D_k \in \mathcal{D}_k\} = X_k^{-1}(\mathcal{D}_k)$, et d'utiliser la formule de transfert :

$$\sigma[X_k^{-1}(\mathcal{D}_k)] = X_k^{-1}[\sigma(\mathcal{D}_k)];$$

on a donc $\sigma(\mathcal{C}_k) = \mathcal{A}_{X_k}$. □

Corollaire 4.14 Les v.a.r. $X_1, \dots, X_n: \Omega \rightarrow \mathbb{R}$ sont indépendantes si et seulement si :

$$\mathbb{P}(X_k \leq x_k, \forall k \leq n) = \prod_{k=1}^n F_{X_k}(x_k), \quad \forall x_1, \dots, x_n \in \mathbb{R}.$$

On a une variante de la Proposition 4.12 :

Proposition 4.15 Soit, pour $1 \leq k \leq n$, $\mathcal{C}_k \subseteq \mathcal{A}_k$ stable par intersection finie et engendrant \mathcal{A}_k . Si C_1, \dots, C_n sont indépendants pour tous les choix de $C_1 \in \mathcal{C}_1, \dots, C_n \in \mathcal{C}_n$, alors les tribus $\mathcal{A}_1, \dots, \mathcal{A}_n$ sont indépendantes.

Preuve. Si l'on pose $\mathcal{C}'_k = \mathcal{C}_k \cup \{\Omega\}$, alors \mathcal{C}'_k vérifie les conditions de la Proposition 4.12, puisque, si $C_k \in \mathcal{C}'_k$ et si J est l'ensemble des $k \leq n$ pour lesquels $C_k = \Omega$, on a, grâce à l'indépendance des C_k pour $k \notin J$:

$$\mathbb{P}\left(\bigcap_{k=1}^n C_k\right) = \mathbb{P}\left(\bigcap_{k \notin J} C_k\right) = \prod_{k \notin J} \mathbb{P}(C_k) = \prod_{k=1}^n \mathbb{P}(C_k). \quad \square$$

5 Fonction caractéristique

5.1 Généralités

Définition 5.1 La fonction caractéristique (en abrégé f.c.) d'une v.a.r. $X: \Omega \rightarrow \mathbb{R}$ est définie par :

$$\Phi_X(t) = \mathbb{E}(e^{itX}), \quad \forall t \in \mathbb{R}.$$

En d'autres termes :

$$\Phi_X(t) = \int_{\mathbb{R}} e^{itx} d\mathbb{P}_X(x).$$

Remarquons que lorsque X a une densité f_X , cela s'écrit :

$$\Phi_X(t) = \int_{\mathbb{R}} e^{itx} f_X(x) dx;$$

on reconnaît la définition de la *transformée de Fourier* de f_X :

$$\Phi_X(t) = (\mathcal{F} f_X)\left(-\frac{t}{2\pi}\right).$$

Plus généralement, on peut définir la transformée de Fourier d'une mesure.

Définition 5.2 Soit m une mesure complexe sur \mathbb{R} muni de sa tribu borélienne. La **transformée de Fourier** de m est définie par :

$$(\mathcal{F}m)(t) = \hat{m}(t) = \int_{\mathbb{R}} e^{-2\pi itx} dm(x), \quad \forall x \in \mathbb{R}.$$

Cela s'applique en particulier lorsque m est une mesure positive **bornée**, donc lorsque m est une probabilité. Pour toute *v.a.r.* X , on a donc :

$$\Phi_X(t) = (\mathcal{F} \mathbb{P}_X)\left(-\frac{t}{2\pi}\right).$$

Le Théorème de convergence dominée donne immédiatement :

Proposition 5.3 $\mathcal{F}m = \hat{m}$ est une fonction continue bornée sur \mathbb{R} (c'est-à-dire que $\mathcal{F} : \mathcal{M}(\mathbb{R}) \rightarrow \mathcal{C}_b(\mathbb{R})$) et $|\hat{m}(t)| \leq |m|(\mathbb{R}) = \widehat{|m|}(0)$.

Corollaire 5.4 Pour toute *v.a.r.* X , on a $\Phi_X \in \mathcal{C}_b(\mathbb{R})$; de plus $\boxed{\Phi_X(0) = 1}$ et $\boxed{|\Phi_X(t)| \leq 1}$ pour tout $t \in \mathbb{R}$.

Théorème 5.5 \mathcal{F} est injective sur $\mathcal{M}(\mathbb{R})$.

Corollaire 5.6 Si deux *v.a.r.* ont la même fonction caractéristique, alors ces deux *v.a.r.* ont la même loi.

Autrement dit : $\boxed{\text{si } \Phi_X = \Phi_Y, \text{ alors } \mathbb{P}_X = \mathbb{P}_Y}$: la fonction caractéristique détermine la loi.

Notation. On écrira $\boxed{X \stackrel{\mathcal{L}}{=} Y}$ lorsque X et Y ont la même loi.

Bien sûr, on ne peut rien dire des *v.a.r.* elle-mêmes ; elles pourraient d'ailleurs ne même pas être définies sur le même espace de probabilité.

Preuve. Soit $m \in \mathcal{M}(\mathbb{R})$, et soit $m = h \cdot |m|$ sa décomposition polaire (on a

$|h| = 1$ $|m|$ -p.p.). Pour toute $f \in L^1(\mathbb{R})$, on a, par le Théorème de Fubini :

$$\begin{aligned} \int_{\mathbb{R}} \hat{f}(y) dm(y) &= \int_{\mathbb{R}} \left[\int_{\mathbb{R}} f(x) e^{-2\pi ixy} dx \right] dm(y) \\ &= \int_{\mathbb{R}} \left[\int_{\mathbb{R}} f(x) e^{-2\pi ixy} dx \right] h(y) d|m|(y) \\ &= \int_{\mathbb{R}} f(x) \left[\int_{\mathbb{R}} e^{-2\pi ixy} h(y) d|m|(y) \right] dx \\ &= \int_{\mathbb{R}} f(x) \left[\int_{\mathbb{R}} e^{-2\pi ixy} dm(y) \right] dx \\ &= \int_{\mathbb{R}} f(x) \hat{m}(x) dx. \end{aligned}$$

Donc, si $\hat{m} = 0$, on a $\int_{\mathbb{R}} \hat{f}(y) dm(y) = 0$ pour toute $f \in L^1(\mathbb{R})$. Comme $\mathcal{FL}^1(\mathbb{R})$ est dense dans $\mathcal{C}_0(\mathbb{R})$, on a aussi $\int_{\mathbb{R}} g(y) dm(y) = 0$ pour toute $g \in \mathcal{C}_0(\mathbb{R})$; donc $m = 0$ (car, par exemple, on aura $m([a, b]) = 0$ pour tout intervalle $[a, b]$ de \mathbb{R} , en approchant $\mathbb{1}_{[a, b]}$ par une suite de fonctions de $\mathcal{C}_0(\mathbb{R})$). \square

5.2 Fonction caractéristique de la loi normale

Théorème 5.7 Si $X \sim \mathcal{N}(m, \sigma^2)$, alors :

$$\Phi_X(t) = \exp\left(imt - \frac{\sigma^2 t^2}{2}\right).$$

En particulier, si $Z \sim \mathcal{N}(0, 1)$ est centrée réduite, alors $\Phi_Z(t) = e^{-t^2/2}$.

Le calcul, dans le cas centrée réduite a été fait dans le cours d'Analyse Fonctionnelle. Dans le cas général, il suffit d'utiliser le résultat immédiat suivant, puisque si $X \sim \mathcal{N}(m, \sigma^2)$, alors $X = \sigma Z + m$ avec $Z \sim \mathcal{N}(0, 1)$.

Lemme 5.8 Pour toute v.a.r. Y , on a $\Phi_{aY+b}(t) = e^{ibt} \Phi_Y(at)$, pour tous $a, b \in \mathbb{R}$.

Une des propriétés importantes des f.c. est la suivante.

Théorème 5.9 Si X et Y sont deux v.a.r. indépendantes, alors :

$$\Phi_{X+Y}(t) = \Phi_X(t) \Phi_Y(t), \quad \forall t \in \mathbb{R}.$$

Preuve.

$$\begin{aligned} \Phi_{X+Y}(t) &= \mathbb{E}(e^{it(X+Y)}) = \mathbb{E}(e^{itX} e^{itY}) \\ &= \mathbb{E}(e^{itX}) \mathbb{E}(e^{itY}) \quad (\text{Proposition 4.9 et Proposition 4.5}) \\ &= \Phi_X(t) \Phi_Y(t). \end{aligned} \quad \square$$

On en déduit l'important résultat suivant.

Corollaire 5.10 *Si $X_1 \sim \mathcal{N}(m_1, \sigma_1^2)$ et $X_2 \sim \mathcal{N}(m_2, \sigma_2^2)$ sont deux v..a.r. gaussiennes indépendantes, alors leur somme $X_1 + X_2$ est encore gaussienne, et suit la loi $\mathcal{N}(m_1 + m_2, \sigma_1^2 + \sigma_2^2)$.*

Notons que, de toute façon, dès que l'on sait que $X_1 + X_2$ est gaussienne $\mathcal{N}(m, \sigma^2)$, on sait que $m = \mathbb{E}(X_1 + X_2) = \mathbb{E}(X_1) + \mathbb{E}(X_2) = m_1 + m_2$ et que, grâce à l'indépendance (Proposition 4.10), $\sigma^2 = \text{Var}(X_1 + X_2) = \text{Var}(X_1) + \text{Var}(X_2) = \sigma_1^2 + \sigma_2^2$; mais on va tout obtenir d'un coup.

Preuve. Il suffit d'appliquer le Théorème 5.9 :

$$\begin{aligned}\Phi_{X_1+X_2}(t) &= \Phi_{X_1}(t) \Phi_{X_2}(t) = \exp\left(im_1t - \frac{\sigma_1^2 t^2}{2}\right) \exp\left(im_2t - \frac{\sigma_2^2 t^2}{2}\right) \\ &= \exp\left(i(m_1 + m_2)t - \frac{(\sigma_1^2 + \sigma_2^2)t^2}{2}\right),\end{aligned}$$

et d'utiliser le Théorème 5.7 et le Théorème 5.6. □

5.3 Dérivabilité

Pour terminer, signalons juste le résultat suivant, conséquence immédiate du Théorème de dérivation sous les intégrales, qui permet de calculer les moments d'une v.a.r. lorsque l'on connaît sa f.c..

Proposition 5.11 *Soit r un entier ≥ 1 . Si X a un moment d'ordre r , alors Φ_X est r fois continûment dérivable et $\Phi_X^{(r)}(0) = i^r \mathbb{E}(X^r)$.*