1 Mesures, probabilités

Tribus
Soit Ω un ensemble quelconque. Un ensemble de parties $\mathcal{A} \subseteq \mathscr{P}(\Omega)$ est appelé *tribu sur* Ω ou *σ-algèbre* si :

(i) $\Omega \in \mathcal{A}$;

(ii) $\mathcal{A}$ est stable par passage au complémentaire : si $A \in \mathcal{A}$, alors $A^c \in \mathcal{A}$;

(iii) $\mathcal{A}$ est stable par union dénombrable : si $(A_n)_{n\in\mathbb{N}} \in \mathcal{A}^{\mathbb{N}}$, alors $\bigcup_{n\in\mathbb{N}} A_n \in \mathcal{A}$.

On dit que $(\Omega, \mathcal{A})$ est un espace *mesurable* ou *probabilisable*. Les éléments de $\mathcal{A}$ sont appelés *ensembles mesurables* ou *événements*.

(Rapprocher cela de la définition d'une topologie)

Conséquences immédiates
Soit $(\Omega, \mathcal{A})$ un espace probabilisable. On a :

(i) $\varnothing \in \mathcal{A}$;

(ii) $\mathcal{A}$ est stable par union finie ;

(iii) $\mathcal{A}$ est stable par intersection finie ou dénombrable.

Intersection de tribus
Une intersection quelconque de tribus sur un même ensemble est une tribu sur cet ensemble.

Tribu engendrée
Soient un ensemble Ω et une famille de parties $(A_i)_{i\in I} \subseteq \Omega^I$. On appelle *tribu engendrée par* (A_i) la plus petite tribu contenant tous les A_i, c'est-à-dire l'intersection de toutes les tribus contenant les A_i.
On la note $\sigma((A_i)_{i\in I})$.

Tribu engendrée par une partie
Si $A \subseteq \Omega$, alors $\sigma(A) = \{\varnothing, A, A^c, \Omega\}$.

Tribu borélienne
On appelle *tribu borélienne de* $\mathbb{R}$ la tribu engendrée par les ouverts de $\mathbb{R}$, ou encore par les $]-\infty, a]$ pour $a \in \mathbb{Q}$. On la note $\mathscr{B}(\mathbb{R})$. On remarque que :

(i) tous les ouverts et les fermés sont dans $\mathscr{B}(\mathbb{R})$;

(ii) tous les intervalles sont dans $\mathscr{B}(\mathbb{R})$.

Mais $\mathscr{B}(\mathbb{R}) \subsetneq \mathscr{P}(\mathbb{R})$.

Mesure
Soit $(\Omega, \mathcal{A})$ un espace mesurable. On appelle *mesure* sur Ω toute application $\mu : \Omega \to [0, \infty]$ telle que :

(i) $\mu(\varnothing) = 0$;

(ii) pour toute famille dénombrable $(A_n)_{n\in\mathbb{N}}$ de parties mesurables *deux à deux disjointes*,
$$\mu\Big(\bigcup_{n\in\mathbb{N}} A_n\Big) = \sum_{n\in\mathbb{N}} \mu(A_n)$$
dans $[0, \infty]$.

On dit que $(\Omega, \mathcal{A}, \mu)$ est un espace mesuré.

Mesures σ-finies, finies et probabilités
Soit $(\Omega, \mathcal{A}, \mu)$ un espace mesuré.

(i) on dit que μ est σ-finie si Ω peut être recouvert par un nombre dénombrable de parties de mesure finie. ;

(ii) on dit que μ est finie si $\mu(\Omega) < \infty$;

(iii) on dit que μ est une probabilité si $\mu(\Omega) = 1$.

Exemples de mesures
Quel que soit Ω, on peut définir :

(i) la *mesure de comptage* sur $(\Omega, \mathscr{P}(\Omega))$ par $\mu(A) = \operatorname{Card} A$;

(ii) la *mesure de Dirac* sur $(\Omega, \mathcal{A})$, où $\mathcal{A}$ est n'importe quelle tribu, par
$$\delta_{\omega_0} : A \in \mathcal{A} \mapsto \begin{cases} 1 & \text{si } \omega_0 \in A \\ 0 & \text{sinon} \end{cases}.$$

Propriétés des mesures σ-finies (1)
Soit $(\Omega, \mathcal{A}, \mu)$ un espace mesuré avec μ σ-finie. Dans ce qui suit, les ensembles considérés sont implicitement dans $\mathcal{A}$:

(i) si $A \subseteq B$, alors $\mu(A) \leqslant \mu(B)$ dans $[0, \infty]$;

(ii) $\mu(A \cup B) = \mu(A) + \mu(B) - \mu(A \cap B)$;

(iii) dans $[0, \infty]$ on a
$$\mu\Big(\bigcup_{n\in\mathbb{N}} A_n\Big) \leqslant \sum_{n\in\mathbb{N}} \mu(A_n).$$

Propriétés des mesures σ-finies (2)

Soit $(\Omega, \mathcal{A}, \mu)$ un espace mesuré avec μ σ-finie. Soit $(A_n) \in \mathcal{A}^{\mathbb{N}}$:

(i) si pour tout n, $A_n \subseteq A_{n+1}$, alors

$$\mu\Big(\bigcup_{n\in\mathbb{N}} A_n\Big) = \lim_{n\to\infty} \nearrow \mu(A_n)\,;$$

(ii) si pour tout n, $A_{n+1} \subseteq A_n$ et que $\mu(A_n) < \infty$ à partir d'un certain rang, alors

$$\mu\Big(\bigcap_{n\in\mathbb{N}} A_n\Big) = \lim_{n\to\infty} \searrow \mu(A_n)\,;$$

(iii) si $\big(\sum \mu(A_n)\big)_{n\in\mathbb{N}}$ converge, alors $\mu(\limsup_{n\to\infty} A_n) = 0$.

Lemme de Borel-Cantelli

Soit $(\Omega, \mathcal{A}, \mu)$ un espace mesuré. Soit $(A_n) \in \mathcal{A}^{\mathbb{N}}$:

(i) si $\big(\sum \mu(A_n)\big)_{n\in\mathbb{N}}$ converge, alors $\mu(\limsup A_n) = 0$;

(ii) si les A_n sont indépendants deux à deux et que $\big(\sum \mu(A_n)\big)_{n\in\mathbb{N}}$ diverge, alors $\mu(\limsup A_n) = 1$.

Mesure de Lebesgue

(i) il n'existe pas de probabilité $\mathbb{P}$ sur $([0,\,1], \mathscr{P}([0,\,1]))$ telle que pour tous $0 \leqslant a \leqslant b \leqslant 1$, $\mathbb{P}([a,\,b]) = b - a$;

(ii) il existe une unique probabilité $\mathbb{P}$ sur $([0,\,1], \mathscr{B}([0,\,1]))$ telle que pour tous $0 \leqslant a \leqslant b \leqslant 1$, $\mathbb{P}([a,\,b]) = b - a$;

(iii) il existe une unique mesure λ sur $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$ telle que pour tous $a \leqslant b$, $\lambda([a,\,b]) = b - a$.

On appelle λ la *mesure de Lebesgue sur* $\mathbb{R}$.

Propriétés de la mesure de Lebesgue

(i) pour tous $a \leqslant b$,

$$\begin{aligned}\lambda([a,\,b]) = \lambda([a,\,b[) = \lambda(]a,\,b]) = \lambda(]a,\,b[)\\ = b - a\,;\end{aligned}$$

(ii) pour tout $x \in \mathbb{R}$, $\lambda(\{x\}) = 0$;

(iii) λ est σ-finie ;

(iv) λ est invariante par translation ;

(v) si μ est une mesure sur $(\mathbb{R}, \mathscr{P}(\mathbb{R}))$ finie sur les parties bornées et invariante par translation, alors il existe $c \geqslant 0$ telle que $\mu = c\lambda$.

2 Variables aléatoires, lois

Espace de probabilité

Un espace mesuré $(\Omega, \mathcal{A}, \mathbb{P})$ dont la mesure $\mathbb{P}$ est une probabilité est appelé *espace de probabilité*.

Dans ce qui suit, on fixe un espace de probabilité $(\Omega, \mathcal{A}, \mathbb{P})$.

Variable aléatoire

Une fonction $X : \Omega \to \mathbb{R}$ telle que pour tout $B \in \mathscr{B}(\mathbb{R})$, $X^{-1}(B) \overset{\text{not.}}{=} (X \in B) \in \mathcal{A}$ est appelée *variable aléatoire* (ou *fonction mesurable*).

Un critère un peu plus simple

Une fonction $X : \Omega \to \mathbb{R}$ est une variable aléatoire si et seulement si $(X \leqslant x) \in \mathcal{A}$ pour tout $x \in \mathbb{Q}$.

Propriétés des variables aléatoires

(i) si X et Y sont deux v.a., alors $X + Y$, XY et $\inf(X, Y)$ sont des variables aléatoires ;

(ii) plus généralement, si $f : \mathbb{R} \to \mathbb{R}$ est une application mesurable (au sens des boréliens), alors $f(X)$ est une variable aléatoire ;

(iii) si (X_n) est une suite de v.a., alors $\inf X_n$, $\sup X_n$, $\liminf X_n$, $\limsup X_n$ sont des v.a. ;

(iv) si (X_n) converge simplement, alors $\lim X_n$ est une v.a.

Exemples de variables aléatoires

(i) si $A \subseteq \Omega$, alors $\mathbf{1}_A$ est une v.a. si et seulement si A est mesurable ;

(ii) une fonction continue $\mathbb{R} \to \mathbb{R}$ est une variable aléatoire sur $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$.

Loi de X

Soit X une v.a. réelle. La *loi de* X est l'application

$$\mathbb{P}_X : \begin{cases} \mathscr{B}(\mathbb{R}) & \to & [0,\,1] \\ B & \mapsto & \mathbb{P}_X(B) = \mathbb{P}(X \in B) \end{cases}.$$

Il s'agit d'une probabilité

Soit X une v.a. réelle. La loi de X, $\mathbb{P}_X$, est une probabilité sur $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$.

(À partir d'une probabilité sur Ω, on en a construit une sur $\mathbb{R}$)

3 Fonction de répartition

Fonction de répartition
Soit X une variable aléatoire de loi $\mathbb{P}_X$. On appelle *fonction de répartition de X* la fonction

$$F_X : \begin{cases} \mathbb{R} & \to & [0,1] \\ x & \mapsto & \mathbb{P}_X(]-\infty, x]) = \mathbb{P}(X \leqslant x) \end{cases}.$$

Propriétés de F_X
Quelle que soit la v.a. X, on a :

(i) F_X est croissante ;

(ii) F_X est continue à droite, c'est-à-dire que

$$(\forall x_0 \in \mathbb{R}) \quad F_X(x) \xrightarrow[\substack{x \to x_0 \\ x \geqslant x_0}]{} F_X(x_0)\,;$$

(iii) $F_X \xrightarrow[-\infty]{} 0$ et $F_X \xrightarrow[+\infty]{} 1$.

La fonction de répartition caractérise la loi
Si deux v.a. ont même fonction de répartition, alors elles ont même loi.

Une propriété d'existence
Toute fonction $\mathbb{R} \to \mathbb{R}$ qui vérifie les propriétés (i), (ii) et (iii) ci-dessus est la fonction de répartition d'une loi de probabilité sur $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$.

Exemples d'utilisation de F_X

(i) $\mathbb{P}(X \leqslant x) = \mathbb{P}_X(]-\infty, x]) = F_X(x)$;

(ii) $\mathbb{P}(X < x) = \mathbb{P}_X(]-\infty, x[) = F_X(x^-)$;

(iii) $\mathbb{P}(a < X < b) = \mathbb{P}_X(]a, b[) = F_X(b^-) - F_X(a)$.

Probabilité d'un singleton

(i) si $x \in \mathbb{R}$, alors
$$\mathbb{P}(X = x) = \mathbb{P}_X(\{x\}) = F_X(x) - F_X(x^-)\,;$$

(ii) F_X est continue en $x \in \mathbb{R}$ si et seulement si $(X = x)$ est négligeable.

Fonction de répartition d'une variable discrète
Si $X(\Omega) = \{x_n \,:\, n \in \mathbb{N}\}$ est dénombrable, alors

$$F_X(X) = \sum_{\substack{n \in \mathbb{N} \\ x_n \leqslant x}} \mathbb{P}(X = x_n).$$

4 Intégrale et espérance

Intégrale
Soit $(\Omega, \mathcal{A}, \mu)$ un espace mesuré. Soit f une fonction intégrable positive. On pose $f_n(x) = \min(n, \lfloor 2^n f(x) \rfloor / 2^n)$. f_n est une suite croissante de fonctions étagées.
L'intégrale de f_n est

$$\int_\Omega f_n \, \mathrm{d}\mu = \sum_{k=0}^{n2^n} \frac{k}{2^n} \cdot \mu\Big(f_n = \frac{k}{2^n}\Big) + n\mu(f_n = n).$$

Comme il s'agit d'une suite croissante de réels, on définit

$$\int_\Omega f \, \mathrm{d}\mu = \lim_{n \to \infty} \int_\Omega f_n \, \mathrm{d}\mu \in [0, \infty].$$

Espérance
Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé. Soit X une v.a. positive. On pose $X_n(\omega) = \min(n, \lfloor 2^n X(\omega) \rfloor / 2^n)$. X_n est une suite croissante de v.a. étagées.
L'espérance de X_n est

$$\mathbb{E}[X_n] = \sum_{k=0}^{n2^n} \frac{k}{2^n} \cdot \mathbb{P}\Big(X_n = \frac{k}{2^n}\Big) + n\mathbb{P}(X_n = n).$$

Comme il s'agit d'une suite croissante de réels, on définit

$$\mathbb{E}[X] = \lim_{n \to \infty} \mathbb{E}[X_n] \in [0, \infty].$$

Espérance d'une v.a. de signe quelconque
On décompose $X = X^+ - X^-$. Alors $|X| = X^+ + X^-$.
On dit que X est *intégrable* si X^+ et X^- le sont, ou équivalemment si $|X|$ l'est. On pose dans ce cas

$$\mathbb{E}[X] = \mathbb{E}[X^+] - \mathbb{E}[X^-]$$

qu'on note aussi

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

Propriétés de l'espérance

(i) l'espérance est linéaire (L^1 est un espace vectoriel) ;

(ii) si X est constante égale à a, $\mathbb{E}[X] = a$;

(iii) si X est presque sûrement égale à a, $\mathbb{E}[X] = a$;

(iv) l'espérance est croissante ;

(v) si $A \in \mathcal{A}$, alors $\mathbb{E}[\mathbf{1}_A] = \mathbb{P}(A)$;

Variance
On appelle L^2 l'espace vectoriel des v.a. réelles de carré intégrable. On a $L^2 \subseteq L^1$. Si $X \in L^2$, alors sa variance est définie comme

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

Meilleure approximation
Si $X \in L^2$, alors :

$$\mathbb{E}[X] = \arg\min_{b\in\mathbb{R}} \mathbb{E}[(X-b)^2]$$
$$\mathbb{V}[X] = \min_{b\in\mathbb{R}} \mathbb{E}[(X-b)^2]$$

Propriété de transport
Soient $X : \Omega \to \mathbb{R}$ une v.a. et $g : \mathbb{R} \to \mathbb{R}$ mesurable avec $g \geqslant 0$ ou $g(X)$ intégrable. Alors

$$\mathbb{E}[g(X)] = \int_\Omega g(X(\omega))\,\mathbb{P}(\mathrm{d}\omega) = \int_\mathbb{R} g(x)\,\mathbb{P}_X(\mathrm{d}x).$$

À quoi ça sert
On se ramène donc à une intégrale sur $\mathbb{R}$, par rapport à la mesure $\mathbb{P}_X$. En particulier, on a

$$\mathbb{E}[X] = \int_{-\infty}^{+\infty} x \,\mathrm{d}\mathbb{P}_X.$$

Markov et Bienaymé-Tchebychev

(i) si X est une v.a. intégrable et $a > 0$, alors

$$\mathbb{P}(|X| \geqslant a) \leqslant \frac{\mathbb{E}[X]}{a}\ ;$$

(ii) si X est une v.a. de carré intégrable et $a > 0$, alors

$$\mathbb{P}(|X - \mathbb{E}[X]| \geqslant a) \leqslant \frac{\mathbb{V}[X]}{a^2}.$$

Jensen
Soient X intégrable et g mesurable telle que $g(X)$ intégrable. Si g est convexe, alors

$$g(\mathbb{E}[X]) \leqslant \mathbb{E}[g(X)].$$

Covariance
Soit X et Y deux v.a. de carré intégrable. On appelle covariance de X et Y la quantité

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

Variance et covariance
La covariance est bilinéaire et $\mathrm{Cov}(X,X) = \mathbb{V}[X]$.

5 Variables particulières : les v.a. à densité

Densité de probabilité
Une fonction $f : \mathbb{R} \to \mathbb{R}$ est appelée *densité de probabilité* si :

(i) elle est positive ;

(ii) elle est intégrable et $\displaystyle\int_\mathbb{R} f = 1$.

Loi à densité

(i) soit $f : \mathbb{R} \to \mathbb{R}$ une densité de probabilité. La fonction

$$\mathbb{P}^f = \lambda \cdot f : B \in \mathscr{B}(\mathbb{R}) \mapsto \int_B f$$

est une probabilité sur $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$.

(ii) soit μ une probabilité sur $(\Omega, \mathscr{B}(\mathbb{R}))$. On dit que μ est une *loi à densité* s'il existe une densité de probabilité $f : \mathbb{R} \to \mathbb{R}$ telle que $\mu = \mathbb{P}^f$. Il suffit pour cela que pour tout x,

$$\mu(]-\infty, x]) = \int_{]-\infty,x]} f.$$

Variable aléatoire à densité
Soit $X : \Omega \to \mathbb{R}$ une v.a. réelle. On dit que X est *à densité* si $\mathbb{P}_X$ est une loi à densité, c'est-à-dire s'il existe $f : \mathbb{R} \to \mathbb{R}$ telle que $\mathbb{P}_X = \mathbb{P}^f$.

Fonction de répartition d'une v.a. à densité
Soit X une v.a. à densité f. On a :

$$F_X(x) = \int_{-\infty}^x f(y)\,\mathrm{d}y.$$

Dérivabilité de F_X

(i) si X est à densité f, alors F_X est continue : $\mathbb{P}(X = x) = 0$;

(ii) si X est à densité f, alors F_X est dérivable partout où f est continue et en un tel point x, on a $F_X'(x) = f(x)$;

(iii) réciproquement, si F_X est continue, et dérivable par morceau, alors X admet la densité F_X'.

Interprétation de la densité
Soit X une variable à densité f. On suppose f continue en x. Alors

$$f(x) = F_X'(x) = \lim_{\Delta x \to 0} \frac{F_X(x+\Delta x) - F_X(x)}{\Delta x} = \lim_{\Delta x \to 0} \frac{\mathbb{P}_X([x,\, x+\Delta x])}{\Delta x} = \lim_{\Delta x \to 0} \frac{\mathbb{P}(x \leqslant X \leqslant x+\Delta x)}{\Delta x}.$$

Moments d'une variable à densité
Soit X une v.a. à densité f.

(i) X admet une espérance finie si et seulement si $x \mapsto xf(x)$ est intégrable, et on a dans ce cas
$$\mathbb{E}[X] = \int_{-\infty}^{+\infty} xf(x)\,\mathrm{d}x\,;$$

(ii) X admet une variance finie si et seulement si $x \mapsto x^2 f(x)$ est intégrable, et on a dans ce cas
$$\mathbb{V}[X] = \int_{-\infty}^{+\infty} (x - \mathbb{E}[X])^2\, f(x)\,\mathrm{d}x\,;$$

(iii) plus généralement, X admet un moment d'ordre k si et seulement si $x \mapsto x^k f(x)$ est intégrable, et on a dans ce cas
$$\mathbb{E}[X^k] = \int_{-\infty}^{+\infty} x^k f(x)\,\mathrm{d}x.$$

Généralisation : transport
Soit X une v.a. à densité f. Soit $g : \mathbb{R} \to \mathbb{R}$ telle que gf soit intégrable. Alors $g(X)$ admet une espérance finie et dans ce cas
$$\mathbb{E}[g(X)] = \int_{-\infty}^{+\infty} g(x)f(x)\,\mathrm{d}x.$$

Méthode de la fonction muette
Soit X une v.a. réelle.

(i) s'il existe une probabilité μ sur $\mathbb{R}$ telle que pour toute fonction h continue bornée,
$$\mathbb{E}[h(X)] = \int_{-\infty}^{+\infty} h(x)\mu(\mathrm{d}x)$$
alors la loi de X est égale à μ (c'est-à-dire $\mathbb{P}_X = \mu$).

(ii) par conséquent, s'il existe une densité f telle que pour toute fonction h continue bornée,
$$\mathbb{E}[h(X)] = \int_{-\infty}^{+\infty} h(x)f(x)\,\mathrm{d}x$$
alors X admet comme densité f.

Une remarque
Si X a pour densité f et $a \neq 0$, alors aX a pour densité $x \mapsto f(x/a)/a$.

6 Vecteurs aléatoires

Vecteur aléatoire
$\vec{Z} : (\Omega, \mathcal{A}) \to \mathbb{R}^n$ est un vecteur aléatoire si pour tout borélien $B \in \mathscr{B}(\mathbb{R}^n)$, $Z^{-1}(B) \in \mathcal{A}$.

C'est simple en fait
$\vec{Z} = (Z_1, \ldots, Z_n)$ est un vecteur aléatoire si et seulement $Z_1, \ldots, Z_n$ sont des variables aléatoires.

Fonction de répartition
La fonction de répartition de $\vec{Z}$ est $F_{\vec{Z}}$ définie par
$$(z_1, \ldots, z_n) \mapsto \mathbb{P}(\vec{Z} \in \left]-\infty,\, z_1\right] \times \cdots \times \left]-\infty,\, z_n\right]).$$

Densité
$\vec{Z}$ admet une densité f si f est positive, intégrable sur $\mathbb{R}^n$ et d'intégrale 1, et si pour tout $(z_1, \ldots, z_n) \in \mathbb{R}^n$,
$$F_{\vec{Z}}(z_1, \ldots, z_n) = \int_{-\infty}^{z_1} \cdots \int_{-\infty}^{z_n} f(y_1, \ldots, y_n)\,\mathrm{d}y_n \cdots \mathrm{d}y_1.$$

Formule des marginales
Soit $\vec{Z}$ un $\vec{\text{v}}$.a. à densité f. Z_1 admet une densité donnée par
$$f_{Z_1}(z) = \int_{\mathbb{R}^{n-1}} f(z, y_2, \ldots, y_n)\,\mathrm{d}y_2 \cdots \mathrm{d}y_n.$$

Fonction muette
Soit $\vec{X}$ un $\vec{v}$.a. et f une densité. Si pour toute fonction continue bornée h,

$$\mathbb{E}[h(X)] = \int_{\mathbb{R}^n} h(\vec{v})f(\vec{v})\,\mathrm{d}\vec{v}$$

alors X admet f comme densité.

7 Indépendance

Vecteurs aléatoires indépendants
Deux $\vec{v}$.a. $\vec{Y}$ (à valeurs dans $\mathbb{R}^m$) et $\vec{Z}$ (à valeurs dans $\mathbb{R}^n$) sont indépendants si pour tous boréliens $B \subseteq \mathbb{R}^m$ et $C \subseteq \mathbb{R}^n$

$$\mathbb{P}(\vec{Y} \in B, \vec{Z} \in C) = \mathbb{P}(\vec{Y} \in B)\mathbb{P}(\vec{Z} \in C)$$

Indépendance et densité
Soient deux $\vec{v}$.a. $\vec{Y}$ (à valeurs dans $\mathbb{R}^m$) et $\vec{Z}$ (à valeurs dans $\mathbb{R}^n$), de densités respectives $f_{\vec{Y}}$ et $f_{\vec{Z}}$.
On a équivalence entre :
(i) $\vec{Y}$ et $\vec{Z}$ sont indépendants ;
(ii) $(\vec{Y}, \vec{Z})$ admet une densité $f_{(\vec{Y},\vec{Z})}$ et

$$(\text{p.p.t. } \vec{y} \in \mathbb{R}^m) \quad (\text{p.p.t. } \vec{z} \in \mathbb{R}^n) f_{(\vec{Y},\vec{Z})}(\vec{y}, \vec{z}) = f_{\vec{Y}}(\vec{y})f_{\vec{Z}}(\vec{z}).$$

Indépendance et espérance
Soient $\vec{Y}$ et $\vec{Z}$ des $\vec{v}$.a. indépendants. Soient $g : \mathbb{R}^m \to \mathbb{R}$ et $h : \mathbb{R}^n \to \mathbb{R}$ deux fonctions mesurables positives ou bornées. On a

$$\mathbb{E}[g(\vec{Y})h(\vec{Z})] = \mathbb{E}[g(\vec{Y})]\mathbb{E}[h(\vec{Z})].$$

Indépendance et covariance
Soit X et Y deux v.a. indépendantes. On a $\mathrm{Cov}(X, Y) = 0$.

Indépendance et variance
Soit X et Y deux v.a. indépendantes. On a $\mathbb{V}[X + Y] = \mathbb{V}[X] + \mathbb{V}[Y]$.

Indépendance et convolution
Soient X et Y deux v.a. indépendantes de densités respectives f_X et f_Y. Alors $X + Y$ admet une densité f_{X+Y} donnée par

$$f_{X+Y}(x) = (f_X * f_Y)(x) = \int_{\mathbb{R}} f_X(x - y)f_Y(y)\,\mathrm{d}y.$$

8 Fonction caractéristique

Fonction caractéristique
Soit $\vec{X}$ un $\vec{v}$.a. de dimension d. La fonction caractéristique de $\vec{X}$ est

$$\phi_{\vec{X}} \;:\; \vec{u} \in \mathbb{R}^d \mapsto \mathbb{E}[\mathrm{e}^{\mathrm{i}\vec{u}\cdot\vec{X}}] = \mathbb{E}[\cos(\vec{u} \cdot \vec{X})] + \mathrm{i}\,\mathbb{E}[\sin(\vec{u} \cdot \vec{X})]$$

Variable à densité
Si $\vec{X}$ est à densité f, alors $\phi_{\vec{X}}$ est la transformée de Fourier $\hat{f}$ de f.

Fonction caractéristique et continuité
Soit $\vec{X}$ un $\vec{v}$.a. $\phi_{\vec{X}}$ est continue, de module inférieur à 1 et $\phi_{\vec{X}}(\vec{0}) = 1$.

Fonction caractéristique et moments
Soit X une v.a. réelle. Si $|X|^m$ est intégrable pour un entier m, alors ϕ_X est de classe $\mathscr{C}^m$ et

$$\phi_X^{(m)}(u) = \mathrm{i}^m\mathbb{E}[\mathrm{e}^{\mathrm{i}uX}X^m].$$

En particulier, $\mathbb{E}[X] = -\mathrm{i}\phi'_X(0)$ et $\mathbb{E}[X^2] = -\phi''_X(0)$.

Propriété fondamentale
Si deux $\vec{v}$.a. ont mêmes fonctions caractéristiques, alors ils ont la même loi.

Fonction caractéristique et indépendance
X et Y sont indépendantes si et seulement si $\phi_{(X,Y)} = \phi_X\phi_Y$.

9 Types de convergence

Convergence p.s.
Une suite $(\vec{X}_n)_{n\in\mathbb{N}}$ de $\vec{v}$.a. converge presque sûrement vers $\vec{X}$ si

$$\mathbb{P}(\|\vec{X}_n - \vec{X}\| \to 0) = 1.$$

On note $\vec{X}_n \xrightarrow{\text{p.s.}} \vec{X}$.

Propriétés de la convergence p.s.
Soit $(\vec{X}_n)$ et $(\vec{Y}_n)$ telles que $\vec{X}_n \xrightarrow{\text{p.s.}} \vec{X}$ et $\vec{Y}_n \xrightarrow{\text{p.s.}} \vec{Y}$:
(i) pour toute fonction continue f, $f(\vec{X}_n) \xrightarrow{\text{p.s.}} f(\vec{X})$;
(ii) $(\vec{X}_n, \vec{Y}_n) \xrightarrow{\text{p.s.}} (\vec{X}, \vec{Y})$;
(iii) $\vec{X}_n + \vec{Y}_n \xrightarrow{\text{p.s.}} \vec{X} + \vec{Y}$, $\vec{X}_n \cdot \vec{Y}_n \xrightarrow{\text{p.s.}} \vec{X} \cdot \vec{Y}$ etc.

Convergence monotone
Soit (X_n) une suite de v.a. positives, croissante et convergeant p.s. vers X. On a $\mathbb{E}[X_n] \to \mathbb{E}[X]$.

Lemme de Fatou

Soit (X_n) une suite de v.a. positives. On a $\mathbb{E}[\liminf X_n] \leqslant \liminf \mathbb{E}[X_n]$.

Convergence dominée

Soit (X_n) une suite de v.a. quelconques, convergeant p.s. vers X. On suppose que $|X_n| \leqslant Z$ pour tout n, avec Z v.a. intégrable.
On a alors que X_n et X sont intégrables et que $\mathbb{E}[|X_n - X|] \to 0$.

Convergence en moyenne

La suite (X_n) converge en moyenne vers X si X_n et X sont intégrables et que $\mathbb{E}[|X_n - X|] \to 0$.

On note $X_n \xrightarrow{L^1} X$.

Convergence dominée

Si X_n est dominée par une v.a. intégrable, alors

$$[X_n \xrightarrow{\text{p.s.}} X] \implies [X_n \xrightarrow{L^1} X].$$

Convergence en probabilité

(X_n) converge en probabilité vers X si

$$(\forall \epsilon > 0) \quad \mathbb{P}(|X_n - X| \geqslant \epsilon) \to 0.$$

On note $X_n \xrightarrow{\mathbb{P}} X$.

Inégalité de Markov et convergence

L'inégalité de Markov permet de prouver que si X_n et X sont intégrables alors

$$[X_n \xrightarrow{L^1} X] \implies [X_n \xrightarrow{\mathbb{P}} X].$$

Cas bornée presque sûrement

S'il existe une constante $a > 0$ telle que $\mathbb{P}(|X_n| \leqslant a) = 1$, alors

$$[X_n \xrightarrow{\mathbb{P}} X] \implies [X_n \xrightarrow{L^1} X].$$

Convergences p.s. et en probabilité

On a

$$[X_n \xrightarrow{\text{p.s.}} X] \implies [X_n \xrightarrow{\mathbb{P}} X].$$

Réciproque partielle

On a

$$[X_n \xrightarrow{\mathbb{P}} X] \implies [(\exists \phi) \quad X_{\phi(n)} \xrightarrow{\text{p.s.}} X].$$

Convergence en loi

La suite (X_n) converge en loi vers X si pour toute fonction continue bornée f,

$$\mathbb{E}[f(X_n)] \to \mathbb{E}[f(X)].$$

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

La convergence en loi est très faible

On a

$$[X_n \xrightarrow{\text{p.s.}} X] \implies [X_n \xrightarrow{\mathcal{L}} X].$$

Difficile de remonter

On a

$$(\forall c \in \mathbb{R}) \quad [X_n \xrightarrow{\mathcal{L}} c] \implies [X_n \xrightarrow{\mathbb{P}} c].$$

Unicité des limites ?

(i) si $X_n \xrightarrow{\mathbb{P}} X$ et $X_n \xrightarrow{\mathbb{P}} Y$, alors $X = Y$ p.s. ;

(ii) si $X_n \xrightarrow{\mathcal{L}} X$ et $X_n \xrightarrow{\mathcal{L}} Y$, alors X et Y ont même loi.

Lien avec la fonction de répartition

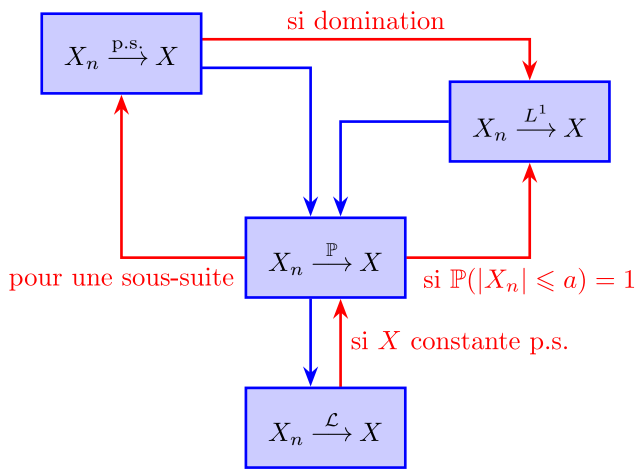
Soit (X_n) et X des v.a. de fonctions de répartition F_n et F. On a

$$[X_n \xrightarrow{\mathcal{L}} X] \iff [F_n(t) \longrightarrow F(t) \text{ partout où } F \text{ continue}].$$

Lien avec la densité

Si $X_n \xrightarrow{\mathcal{L}} X$ avec X à densité, alors pour tous $a < b$,

$$\mathbb{P}(X_n \in]a, b[) \longrightarrow \mathbb{P}(X \in]a, b[).$$



Lien avec la fonction caractéristique
Soit $(\vec{X}_n)$ des $\vec{\text{v}}$.a. à valeurs dans $\mathbb{R}^d$:

(i) si $\vec{X}_n \xrightarrow{\mathcal{L}} \vec{X}$, alors $\phi_{\vec{X}_n} \longrightarrow \phi_{\vec{X}}$ simplement ;

(ii) si $\phi_{\vec{X}_n}$ converge simplement vers ϕ continue en $\vec{0}$, alors ϕ est la fonction caractéristique d'un $\vec{\text{v}}$.a. $\vec{X}$ et $\vec{X}_n \xrightarrow{\mathcal{L}} \vec{X}$.

10 Lois classiques

Loi	Densité	$\mathbb{E}$	$\mathbb{V}$
Uniforme $\mathscr{U}[a,b]$	$\frac{\mathbf{1}_{[a,b]}(x)}{b-a}$	$\frac{b-a}{2}$	$\frac{(b-a)^2}{12}$
Exponentielle $\mathscr{E}(\lambda)$	$\lambda e^{-\lambda x}\mathbf{1}_{\mathbb{R}_+}(x)$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$
Normale $\mathcal{N}(\mu,\sigma^2)$	$\frac{1}{\sqrt{2\pi}}e^{-(x-\mu)^2/2\sigma^2}$	μ	σ^2
Gamma $\Gamma(\alpha,\theta)$	$\frac{\theta^\alpha x^{\alpha-1}}{\Gamma(\alpha)}e^{-\theta x}\mathbf{1}_{\mathbb{R}_+}(x)$	$\frac{\alpha}{\theta}$	$\frac{\alpha}{\theta^2}$
Chi-deux $\chi^2(n)$	$\frac{x^{n/2-1}}{2^{n/2}\Gamma(n/2)}e^{-x/2}\mathbf{1}_{\mathbb{R}_+}(x)$	n	$2n$
Cauchy	$\frac{1}{\pi}\frac{1}{1+x^2}$	$\varnothing$	$\varnothing$
Poisson $\mathscr{P}(\lambda)$	$\mathbb{P}(X=n) = e^{-\lambda}\frac{\lambda^n}{n!}$	λ	λ

11 Lois des grands nombres et théorème centrale limite

Loi faible des grands nombres
Soit (X_n) une suite de v.a. i.i.d. de carré intégrable. Si m est leur espérance, alors

$$M_n = \frac{X_1 + \cdots + X_n}{n} \begin{cases} \xrightarrow{\mathbb{P}} m \\ \xrightarrow{L^1} m \end{cases}.$$

Loi forte des grands nombres
Soit (X_n) une suite de v.a. i.i.d. intégrables. Si m est leur espérance, alors

$$M_n = \frac{X_1 + \cdots + X_n}{n} \begin{cases} \xrightarrow{\text{p.s.}} m \\ \xrightarrow{L^1} m \end{cases}.$$

Histogramme
Soit (X_n) une suite de v.a. indépendantes et de même fonction de répartition F. Pour tous $a < b$, on a

$$\frac{1}{n}\operatorname{Card}\{1 \leqslant i \leqslant n \ : \ a < X_i \leqslant b\} \xrightarrow{\text{p.s.}} F(b) - F(a).$$

Méthode de Monte-Carlo
Soit $(\vec{X}_n)$ suite de $\vec{\text{v}}$.a. i.i.d. à valeurs dans $A \subseteq \mathbb{R}^d$. On suppose qu'ils admettent f comme densité. Soit $g \ : \ A \to \mathbb{R}$ telle que $g(\vec{X}_1)$ soit intégrable. On a

$$\frac{g(\vec{X}_1) + \cdots + g(\vec{X}_n)}{n} \xrightarrow{\text{p.s.}} \mathbb{E}[g(\vec{X}_1)] = \int_A g(\vec{x})f(\vec{x})\,\mathrm{d}\vec{x}.$$

Théorème central limite
Soit (X_n) une suite de v.a. i.i.d. de carré intégrable, d'espérance m et de variance σ^2. Si

$$M_n = \frac{X_1 + \cdots + X_n}{n},$$

alors

$$\sqrt{n}(M_n - m) \xrightarrow{\mathcal{L}} \mathcal{N}(0,\sigma^2).$$

Ce que cela signifie
En gros, pour n assez grand, on a

$$M_n \approx m + \mathcal{N}\big(0, \frac{\sigma^2}{n}\big) = \mathcal{N}\big(m, \frac{\sigma^2}{n}\big).$$

Méthode delta
Soit (X_n) une suite de v.a. i.i.d. de carré intégrable, d'espérance m et de variance σ^2. On note $\overline{X}_n = (X_1 + \cdots + X_n)/n$. En appliquant le théorème centrale limite, éventuellement après avoir écrit une formule de Taylor, on montre que :

(i) si g continue, alors $g(\overline{X}_n) \xrightarrow{\text{p.s.}} m$;

(ii) si g est $\mathscr{C}^1$, alors
$$\sqrt{n}(g(\overline{X}_n) - m) \xrightarrow{\mathcal{L}} \mathcal{N}(0, g'(m)^2\sigma^2)\ ;$$

(iii) si $g'(m) = 0$ et g est $\mathscr{C}^2$, alors
$$n(g(\overline{X}_n) - m) \xrightarrow{\mathcal{L}} \frac{1}{2}\sigma^2 g''(m)\chi_1^2$$

12 Vecteurs gaussiens

Vecteur gaussien
Un $\vec{\text{v}}$.a. $\vec{X} \ : \ \Omega \to \mathbb{R}^n$ est dit gaussien si sa fonction caractéristique est de la forme

$$\phi_{\vec{X}} \ : \ \vec{u} \mapsto e^{i\langle \vec{u}, \vec{m}\rangle - \frac{1}{2}\langle \vec{u}, C\vec{u}\rangle}$$

où $\vec{m} \in \mathbb{R}^n$ et C est une matrice symétrique positive. On note $\vec{X} \sim \mathcal{N}(\vec{m}, C)$.

Espérance et variance d'un vecteur gaussien
Avec les notations précédentes, on a $\mathbb{E}[\vec{X}] = \vec{m}$ et $C = (\mathrm{Cov}(X_i, X_j))$.

Caractérisation d'un vecteur gaussien
$\vec{X}$ est un vecteur gaussien si et seulement si pour tout $\vec{a} \in \mathbb{R}^n$, $\langle \vec{a}, \vec{X} \rangle = \vec{a}^\top \vec{X}$ suit une loi normale.
Autrement dit, il faut et il suffit que toute combinaison linéaire des coordonnées de $\vec{X}$ suive une loi normale.

Vecteur gaussien non dégénéré
Un vecteur gaussien est dit non dégénéré si C est inversible (c'est-à-dire si C est définie positive).

Densité d'un vecteur gaussien
Un vecteur gaussien est non dégénéré si et seulement si il est à densité. Dans ce cas, sa densité est

$$f_{\vec{X}} : \vec{x} \mapsto \frac{1}{(2\pi)^{N/2}\sqrt{\det C}} \times \exp\left[-\frac{1}{2}\big\langle \vec{u} - \vec{m}, C(\vec{u} - \vec{m})\big\rangle\right].$$

Lois des coordonnées
Soit $\vec{X}$ un $\vec{\mathrm{v}}$. gaussien non dégénéré suivant $\mathcal{N}(\vec{m}, C)$. Alors $X_i \sim \mathcal{N}(m_i, C_{ii})$.

Réciproque dans le cas standard
$\vec{X} \sim \mathcal{N}(\vec{0}, I)$ si et seulement si $X_i \sim \mathcal{N}(0,1)$.

Normalisation d'un vecteur gaussien
Soit $\vec{Z} \sim \mathcal{N}(\vec{m}, C)$ un vecteur gaussien non dégénéré. On a :
(i) si $\vec{a} \in \mathbb{R}^n$ et A inversible, alors

$$A\vec{X} + \vec{a} \sim \mathcal{N}(A\vec{m} + \vec{a}, ACA^\top) ;$$

(ii) C admet une racine carrée inversible et

$$C^{-1/2}(\vec{X} - \vec{m}) \sim \mathcal{N}(\vec{0}, I).$$

Lien avec la covariance
Soit $\vec{X}$ un $\vec{\mathrm{v}}$. gaussien non dégénéré. Les v.a. X_i et X_j sont indépendantes si et seulement si $\mathrm{Cov}(X_i, X_j) = 0$.

Somme de lois normales indépendantes
Si $X \sim \mathcal{N}(\mu, \sigma^2)$ et $Y \sim \mathcal{N}(\nu, \tau^2)$ sont indépendantes, alors $X+Y \sim \mathcal{N}(\mu+\nu, \sigma^2+\tau^2)$.

13 Estimateurs

Modèle statistique
Un modèle statistique est un triplet $(\mathcal{X}, \mathcal{A}, \mathcal{P})$ où $\mathcal{A}$ est une tribu sur $\mathcal{X}$ et $\mathcal{P} = \{\mathbb{P}_\theta : \theta \in \Theta\}$ est une famille de probabilités sur $(\mathcal{X}, \mathcal{A})$.

Sur une paire d'éléments
On peut définir un modèle statistique sur $\{0, 1\}$ en le paramétrant par la probabilité d'obtenir 1 : $\mathbb{P}_\theta(\{1\}) = 1 - \mathbb{P}_\theta(\{0\}) = \theta$. Ici on a donc $\Theta = [0, 1]$.

Échantillon
Un n-échantillon est un $\vec{\mathrm{v}}$.a. $\vec{X}$ de n v.a. i.i.d. appartenant au modèle statistique.

Estimateur
On appelle estimateur une fonction $\mathcal{X}^n \to \Theta$. On la note souvent $\hat{\theta}_n$. On confond souvent $\hat{\theta}_n$ avec $\hat{\theta}_n(\vec{X})$ où $\vec{X}$ est un n-échantillon fixé.

Biais
Un estimateur $\hat{\theta}_n$ est non biaisé si $\mathbb{E}_\theta[\hat{\theta}_n] = \theta$ pour tout $\theta \in \Theta$, où $\mathbb{E}_\theta$ est l'espérance relativement à la probabilité $\mathbb{P}_\theta$.
(c'est-à-dire $\mathbb{E}_\theta[\hat{\theta}_n(\vec{X})] = \theta$)

Écart quadratique moyen
Si $g : \Theta \to \mathbb{R}$, alors l'écart quadratique moyen d'un estimateur $\hat{g}_n$ de $g(\theta)$ est

$$\begin{aligned}\mathrm{RQM}_\theta(\hat{g}_n) &= \mathbb{E}_\theta[\hat{g}_n - g(\theta)^2] \\ &= \mathbb{V}_\theta[\hat{g}_n] + \big(\mathbb{E}_\theta(\hat{g}_n) - g(\theta)\big)^2\end{aligned}$$

où $\theta \in \Theta$ est fixé.

Estimateur convergent
(i) on dit qu'un estimateur $\hat{\theta}_n$ est convergent si, pour tout $\theta \in \Theta$ et tout $\epsilon > 0$,

$$\mathbb{P}_\theta(|\hat{\theta}_n - \theta| > \epsilon) \to 0 \quad \text{quand} \quad n \to +\infty,$$

autrement dit, si pour tout $\theta \in \Theta$, $\hat{\theta}_n(\vec{X})$ converge en probabilité vers θ.
(ii) on dit que $\hat{\theta}_n$ est fortement convergent si, pour tout $\theta \in \Theta$,

$$\mathbb{P}_\theta(\hat{\theta}_n \to \theta) = 1,$$

autrement dit, si pour tout $\theta \in \Theta$, $\hat{\theta}_n(\vec{X})$ converge p.s. vers θ.

Estimateur asymptotiquement normal
Soit $g : \Theta \to \mathbb{R}$. Un estimateur $\hat{g}_n$ de $g(\theta)$ est dit asymptotiquement normal s'il existe deux fonctions $m_n, \sigma_n : \Theta \to \mathbb{R}$ telles que

$$(\forall \theta \in \Theta) \quad \frac{\hat{g}_n - m_n(\theta)}{\sigma_n(\theta)} \xrightarrow{\mathcal{L}} \mathcal{N}(0,1) \quad \text{sous } \mathbb{P}_\theta.$$

14 Estimateurs empiriques

Dans la suite on se place dans un modèle statistique de la forme $(\mathbb{R}, \mathcal{P}(\mathbb{R}), \mathscr{P})$.

Estimateur de la moyenne
Soit $\vec{X}$ un n-échantillon, dont les lois sont intégrables. On note μ_θ la moyenne d'une v.a. suivant $\mathbb{P}_\theta$.
La moyenne empirique est l'estimateur défini par $\overline{X}_n : \vec{x} \mapsto (x_1 + \cdots + x_n)/n$.

Propriétés de l'estimateur de la moyenne
Avec les notations précédentes :

(i) $\overline{X}_n$ est non biaisé ;

(ii) $\overline{X}_n$ est fortement convergent ;

(iii) si de plus les lois sont de carré intégrables, qu'on note σ_θ^2 la variance d'une loi $\mathbb{P}_\theta$, alors

$$\mathrm{RQM}_\theta(\overline{X}_n) = \frac{\sigma_\theta^2}{n} \ ;$$

(iv) $\overline{X}_n$ est asymptotiquement normal pour $(\mu_\theta, \sigma_\theta/\sqrt{n})$, c'est-à-dire que

$$\sqrt{n}\frac{\overline{X}_n - \mu_\theta}{\sigma_\theta} \xrightarrow{\mathcal{L}} \mathcal{N}(0,1)$$

ou encore

$$\sqrt{n}(\overline{X}_n - \mu_\theta) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \sigma_\theta^2).$$

Estimateur de la variance
Avec les mêmes notations, on définit

$$\overline{V}_n : \vec{x} \mapsto \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{X}_n(\vec{x}))^2$$

c'est-à-dire (abusivement)

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

Propriétés de l'estimateur de la variance

(i) $\overline{V}_n$ est biaisé : $\mathbb{E}_\theta[\overline{V}_n] = \frac{n-1}{n}\sigma_\theta^2$;

(ii) il est fortement convergent ;

(iii) si les lois sont de quatrième moment fini, alors

$$\mathrm{RQM}_\theta(\overline{V}_n) = \frac{\mu_\theta^{(4)} - \sigma_\theta^4}{n} + O(\frac{1}{n}),$$

où $\mu_\theta^{(4)} = \mathbb{E}_\theta[(X_1 - \mathbb{E}_\theta[X_1])^4]$;

(iv) si les lois sont de quatrième moment fini, alors $\overline{V}_n$ est asymptotiquement normal et

$$\sqrt{n}(\overline{V}_n - \sigma_\theta^2) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \mu_\theta^{(4)} - \sigma_\theta^4).$$

Estimateur de la variance non biaisé
On peut débiaiser l'estimateur de la variance en posant

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

Il vérifie exactement propriétés (ii), (iii) et (iv) ci-dessus, et est non biaisé.