

1 Convergence de variables aléatoires

Un peu de calcul sur les événements

Prop. Si $(A_n)_n$ est croissante, $\mathbf{P}(\bigcup_n A_n) = \lim_{n \rightarrow \infty} \mathbf{P}(A_n)$. Si $(A_n)_n$ est décroissante, $\mathbf{P}(\bigcap_n A_n) = \lim_{n \rightarrow \infty} \mathbf{P}(A_n)$.

Def. $\limsup_{n \rightarrow \infty} A_n = \bigcap_{n \in \mathbf{N}} \bigcup_{k \geq n} A_k$.

Donc $\omega \in \limsup_n A_n \iff \forall n, \exists k \geq n, \omega \in A_k$. Donc $\limsup_n A_n$ est réalisé ssi une infinité de A_n est réalisé.

Lem (de Borel-Cantelli). Si $\sum_n \mathbf{P}(A_n) < \infty$, alors $\mathbf{P}(\limsup_n A_n) = 0$.

Autrement dit, il y a une proba 1 pour que seulement un nombre fini de A_n soient réalisés.

Démonstration. Soit $B_n = \bigcup_{k \geq n} A_k$, $\mathbf{P}(\limsup A_n) = \mathbf{P}(\bigcap_n B_n) = \lim_n \mathbf{P}(B_n)$. Or $\mathbf{P}(B_n) = \mathbf{P}(\bigcup_{k \geq n} A_k) \leq \sum_{k \geq n} \mathbf{P}(A_k) \xrightarrow{n \rightarrow \infty} 0$ (par hypothèse). \square

Convergence p.s., en probabilité et dans L^p

Def. (i) On dit que $X_n \xrightarrow{p.s.} X$ (**converge presque sûrement**), si $\forall \omega \mathbf{P}$ -p.p., $X_n(\omega) \rightarrow X(\omega)$. Cela signifie qu'il existe $A \in \mathcal{F}$ tel que $\mathbf{P}(A) = 1$ et $\forall \omega \in A, \lim_n X_n(\omega) = X(\omega)$.

(ii) On dit que X_n converge en probabilité vers X si $\forall \epsilon > 0, \mathbf{P}(\|X_n - X\| > \epsilon) \xrightarrow{n \rightarrow \infty} 0$.

(iii) On dit que X_n converge vers X dans $L^p(\Omega, \mathbf{R}^d)$ si $X_n, X \in L^p$ et $\mathbf{E}(\|X_n - X\|^p) \xrightarrow{n \rightarrow \infty} 0$.

Prop. On note $X_n = \begin{pmatrix} X_n^1 \\ \vdots \\ X_n^d \end{pmatrix}$ où X_n^k est la k^e composante de X_n . Alors $X_n \rightarrow X$ p.s. (resp. en probabilité, dans L^p) ssi $\forall k \in \llbracket 1; d \rrbracket, X_n^k \rightarrow X^k$ p.s. (resp. en probabilité, dans L^p).

Démonstration. Soit $X_n \xrightarrow{p.s.} X$. On fixe $k \in \llbracket 1; d \rrbracket$. Soit $\epsilon > 0$. On sait que $|X_n^k - X^k|^2 < \|X_n - X\|^2$. Donc l'événement $|X_n^k - X^k| > \epsilon$ implique $\mathbf{P}(|X_n^k - X^k| > \epsilon) \leq \mathbf{P}(\|X_n - X\| > \epsilon) \rightarrow 0$. Donc $\forall k, X_n^k \xrightarrow{p.s.} X^k$.

Réciproquement, soit $\epsilon > 0$. On a $\|X_n - X\|^2 = \sum_k |X_n^k - X^k|^2 \leq d \cdot \max_k |X_n^k - X^k|^2$. Donc $\mathbf{P}(\|X_n - X\| > \epsilon) \leq \mathbf{P}(\sqrt{d} \max_k |X_n^k - X^k| > \epsilon) = \mathbf{P}(\exists k, |X_n^k - X^k| > \frac{\epsilon}{\sqrt{d}}) \leq \sum_{k=1}^d \mathbf{P}(|X_n^k - X^k| > \frac{\epsilon}{\sqrt{d}}) \rightarrow 0$. \square

Prop. La convergence p.s. et la convergence L^p impliquent toutes les deux la convergence en probabilité.

Démonstration. (i) Supposons $X_n \xrightarrow{p.s.} X$. Soit $\epsilon > 0$. On a $\mathbf{P}(\|X_n - X\| > \epsilon) = \mathbf{E}(\mathbf{1}_{\|X_n - X\| > \epsilon})$. Or $\|X_n - X\| \rightarrow 0$ p.p. donc $\mathbf{1}_{\|X_n - X\| > \epsilon} \rightarrow 0$ p.p.

$$\lim_n \mathbf{E}(\mathbf{1}_{\|X_n - X\| > \epsilon}) = \mathbf{E}(\lim_n \mathbf{1}_{\|X_n - X\| > \epsilon}) = \mathbf{E}(0) = 0.$$

(ii) $\mathbf{P}(\|X_n - X\| > \epsilon) \leq \frac{\mathbf{E}(\|X_n - X\|^p)}{\epsilon^p} \rightarrow 0$. \square

Prop. Si $\forall \epsilon > 0, \sum_n \mathbf{P}(\|X_n - X\| > \epsilon) < \infty$ alors $X_n \xrightarrow{p.s.} X$.

Démonstration.

$$\begin{aligned} \forall \epsilon > 0 \mathbf{P}(\limsup \{\|X_n - X\| > \epsilon\}) &= 0 \\ \implies \forall \epsilon > 0 \mathbf{P}(\forall n, \exists k \geq n, \|X_k - X\| > \epsilon) &= 0 \\ \forall q \in \mathbf{N}^* \mathbf{P}(\exists n, \forall k \geq n, \|X_k - X\| \leq 1/q) &= 1 \end{aligned}$$

Donc $\mathbf{P}(\bigcap_{q \in \mathbf{N}^*} A_q) = 1$, ce qui se lit

$$\mathbf{P}(\forall q \in \mathbf{N}^*, \exists n, \forall k \geq n, \|X_k - X\| \leq 1/q) = \mathbf{P}(\lim_n \|X_n - X\| = 0) = 1$$

d'où $X_n \xrightarrow{p.s.} X$. \square

Prop. $X_n \xrightarrow{p.s.} X$ ssi on peut extraire une sous-suite φ_n telle que $X_{\varphi_n} \xrightarrow{p.s.} X$.

Prop. $X_n \xrightarrow{p} X \implies$ on peut extraire $X_{\varphi_n} \xrightarrow{p.s.} X$.

Prop. $X_n \xrightarrow{\mathbf{P}} X$ ssi de toute sous-suite X_{φ_n} on peut extraire une autre sous-suite X_{ψ_n} telle que $X_{\psi_n} \xrightarrow{p.s.} X$.

Th (de continuité). X_n, X v.a. sur \mathbf{R}^d . Soit $h: \mathbf{R}^d \rightarrow \mathbf{R}^p$ mesurable et continue sur C tel que $\mathbf{P}(X \in C) = 1$, alors

(i) Si $X_n \xrightarrow{p.s.} X$ alors $h(X_n) \xrightarrow{p.s.} h(X)$

(ii) Si $X_n \xrightarrow{\mathbf{P}} X$ alors $h(X_n) \xrightarrow{\mathbf{P}} h(X)$.

Th (Loi forte des grands nombres). Soit (X_n) i.i.d. telle que $\mathbf{E}(\|X_1\|) < \infty$. Alors $\frac{1}{n} \sum_{i=1}^n X_i \xrightarrow{p.s.} \mathbf{E}(X_1)$.

Th (Loi faible des grands nombres). Soit (X_n) i.i.d. telle que $\mathbf{E}(\|X_1\|^2) < \infty$. On a $\frac{1}{n} \sum_{i=1}^n X_i \xrightarrow{\mathbf{P}} \mathbf{E}(X_1)$.

Convergence en loi

Rappels : une mesure de proba μ sur $(\mathbf{R}^d, \mathcal{B}(\mathbf{R}^d))$ est caractérisée par sa fonction de répartition F_μ .
 $F_\mu(x_1, \dots, x_d) = \mu(\prod_{i=1}^d]-\infty; x_i])$.

On a :

- F_μ croissante.
- $F_\mu(-\infty) = 0, F_\mu(+\infty) = 1$
- F_μ est continue à droite et $\mu(x_0) = F_\mu(x_0) - F_\mu(x_0^-)$

Soit $X: \Omega \rightarrow \mathbf{R}^d$ une v.a. On note $P_X = \mathbf{P} \circ X^{-1}$ la loi de X . P_X est une mesure de proba sur \mathbf{R}^d . On note F_X sa fonction de répartition. Pour $d = 1, F_X(x) = \mathbf{P}(X \leq x)$.

Def. Soit $(\mu_n)_n, \mu$ des mesures de proba sur \mathbf{R}^d . On dit que μ_n converge faiblement (ou étroitement) vers μ si $F_{\mu_n}(x) \rightarrow F_\mu(x)$ en tout x point de continuité de F_μ . On note $\mu_n \Rightarrow \mu$.

Def. $(X_n)_n, X$ v.a. sur \mathbf{R}^d . On dit que X_n converge en loi vers X (noté $X_n \xrightarrow{\mathcal{L}} X$) si $P_{X_n} \Rightarrow P_X$.

Exemple trivial : $X_n = \frac{1}{n}$ (v.a. constantes). Alors $X_n \xrightarrow{\mathcal{L}} 0$.

Prop. $\left. \begin{array}{l} cv \text{ ps} \\ ou \\ cv L^p \end{array} \right\} \Rightarrow cv \text{ proba} \Rightarrow cv \text{ loi}$

Th (de représentation de Skorohod). Soit $(\mu_n)_n, \mu$ des mesures de proba sur \mathbf{R}^d telles que $\mu_n \Rightarrow \mu$. Il existe un espace de proba et des v.a. $(Y_n), Y$ sur cet espace telles que :

- $Y \sim \mu, \forall n, Y_n \sim \mu_n$
- $\forall \omega, Y_n(\omega) \rightarrow Y(\omega)$

Démonstration. (pour $d = 1$) Soit F, F_n les fonctions de répartition de μ, μ_n / Cas simple : supposons que F et F_n sont continues et strictement croissantes. On choisit $\Omega = [0; 1], \mathbf{P}$ la mesure de Lebesgue sur Ω et $\mathcal{F} = \mathcal{B}([0; 1])$. On pose $Y(\omega) = F^{-1}(\omega), Y_n(\omega) = F_n^{-1}(\omega)$.

1. Montrons que $Y \sim \mu$:

$$\begin{aligned} \mathbf{P}(Y \leq t) &= \mathbf{P}(\{\omega \mid F^{-1}(\omega) \leq t\}) \\ &= \mathbf{P}(\{\omega \mid \omega \leq F(t)\}) &= \lambda_{[0;1]}(\{\omega \mid \omega \leq F(t)\}) \\ &= \lambda([0; F(t)]) \\ &= F(t) \end{aligned}$$

De même $Y_n \sim \mu_n$.

2. Exercice : montrer $F_n^{-1}(\omega) \rightarrow F^{-1}(\omega)$. Dans le cas où F_n, F ne sont pas bijectives, définir $F^{-1}(\omega) := \inf\{t \mid F(t) \geq \omega\}$.

□

Th (de continuité). Soit $X_n \xrightarrow{\mathcal{L}} X$ définie sur $(\Omega, \mathcal{F}, \mathbf{P})$. $h: \mathbf{R}^d \rightarrow \mathbf{R}^p$ continue sur C telle que $\mathbf{P}(X \in C) = 1$. Alors $h(X_n) \xrightarrow{\mathcal{L}} h(X)$.

Démonstration. Il existe un autre espace de proba $(\Omega', \mathcal{F}', \mathbf{P}')$ et d'autres v.a. Y_n, Y sur Ω' telles que :

- Y_n, Y ont même loi que X_n, X
- $\forall \omega \in \Omega', Y_n(\omega) \rightarrow Y(\omega)$

Comme $Y_n \xrightarrow{p.s.} Y$ on a $h(Y_n) \xrightarrow{p.s.} h(Y)$, donc $h(Y_n) \xrightarrow{\mathcal{L}} h(Y)$ signifie que $F_{h(Y_n)}(x) \rightarrow F_{h(Y)}(x)$. Or $F_{h(Y_n)}(x) = \mathbf{P}'(h(Y_n) \leq x) = \mathbf{P}(h(X_n) \leq x)$ puisque X_n est égal en loi à Y_n . Donc $F_{h(X_n)}(x) \rightarrow F_{h(X)}(x)$. □

Th (de Portmanteau). On a équivalence entre :

- $X_n \xrightarrow{\mathcal{L}} X$,
- $\forall f: \mathbf{R}^d \rightarrow \mathbf{R}$ continue bornée, $\mathbf{E}(f(X_n)) \rightarrow \mathbf{E}(f(X))$,
- $\forall A \subset \mathbf{R}^d$ tel que $\mathbf{P}(X \in \partial A) = 0$, on a $\mathbf{P}(X_n \in A) \rightarrow \mathbf{P}(X \in A)$ où $\partial A = \bar{A} \setminus \overset{\circ}{A}$

Démonstration. • (i \implies ii) On choisit Y_n et Y comme avant. $Y_n \rightarrow Y$ donc $f(Y_n) \rightarrow f(Y)$

□

Lem (d'Helly). Soit $(F_n)_n$ une suite de fonctions de répartition. Il existe une sous-suite φ_n et $F: \mathbf{R} \rightarrow [0; 1]$ croissante, continue à droite, telle que $F_{\varphi_n}(x) \rightarrow_n F(x)$ en tout point de continuité de F .

Démonstration. On indexe \mathbf{Q} sur \mathbf{N} de sorte que $\mathbf{Q} = \{x_1, \dots, x_n\}$. De $(F_n(x_1))_n$ on peut extraire une sous-suite $(F_{\varphi_n^1}(x_1))_n$ qui converge vers un certain $F(x_1) \in [0; 1]$.

De $(F_{\varphi_n^1}(x_2))_n$ on peut trouver une extraction le long de laquelle la suite converge vers un certain $F(x_2)$. De la sorte pour tout $k \in \mathbf{N}^*$ on construit $(\varphi_n^k)_n$ extrait de $(\varphi_n^{k-1})_n$ tel que $\forall i \leq k, F_{\varphi_n^k}(x_i) \rightarrow_n F(x_i)$.

Posons maintenant $\psi_n = \varphi_n^n$. On a $\lim_n \psi_n = +\infty$ et $\forall n, \forall i, n \geq i \implies \psi_i \in \{\varphi_{n'}^i \mid n' \in \mathbf{N}^*\}$. Donc $\lim_n F_{\psi_n}(x_i) = F(x_i)$.

On pose, pour tout x dans $\mathbf{R} \setminus \mathbf{Q}$, $F(x) := \inf\{F(t) \mid t \geq x, t \in \mathbf{Q}\}$.

On montre que F est croissante (exercice). On montre enfin que pour tout point de continuité x de F , $\lim_n F_{\psi_n}(x) = F(x)$. C'est vrai par construction pour x rationnel. Pour x non-rationnel on a

$$F_{\psi_n}(r^-) \leq F_{\psi_n}(x) \leq F_{\psi_n}(r^+)$$

donc

$$F(r^-) \leq \lim_n F_{\psi_n}(x) \leq \overline{\lim}_n F_{\psi_n}(x) \leq F(r^+)$$

et pour $\varepsilon \downarrow 0$,

$$F(x - \varepsilon) \leq \lim_n F_{\psi_n}(x) \leq \overline{\lim}_n F_{\psi_n}(x) \leq F(x + \varepsilon)$$

d'où,

$$F(x^-) \leq \lim_n F_{\psi_n}(x) \leq \overline{\lim}_n F_{\psi_n}(x) \leq F(x).$$

□

On ajoute une condition pour que la limite vérifie $\lim_{x \rightarrow -\infty} F(x) = 0$ et $\lim_{x \rightarrow +\infty} F(x) = 1$.

Def. $(\mu_n)_n$ est dite **tendue** si $\forall \varepsilon > 0, \exists \mathcal{K}$ compact, $\forall n, \mu_n(\mathcal{K}) \geq 1 - \varepsilon$.

Dans le cas $d = 1$ on peut prendre $\mathcal{K} = [-K; K]$.

Def. $(X_n)_n$ est tendue si $\forall \varepsilon > 0, \exists \mathcal{K}$ compact, $\forall n, \mathbf{P}(X_n \in \mathcal{K}) \geq 1 - \varepsilon$.

Th (de Prokhorov). Soit $(\mu_n)_n$ tendue. Il existe une mesure de probabilité μ sur \mathbf{R}^d et une suite $(\varphi_n)_n$ telle que $\mu_{\varphi_n} \implies \mu$.

Prop. Si toute sous-suite faiblement convergente de $(\mu_n)_n$ tendue converge vers μ^* , alors $\mu_n \implies \mu^*$.

Démonstration. Supposons par l'absurde $\mu_n \not\implies \mu^*$. Alors $\exists f \in \mathcal{C}_b, \int f d\mu_n \not\rightarrow \int f d\mu^*$. Donc $|\int f d\mu_n - \int f d\mu^*| \not\rightarrow 0$. Il existe $\varepsilon > 0$ et $(\varphi_n)_n$ tels que $\forall n, |\int f d\mu_{\varphi_n} - \int f d\mu^*| > \varepsilon$. D'après Prokhorov, puisque $(\mu_n)_n$ est tendue, on peut extraire de $(\varphi_n)_n$ une autre sous-suite $(\psi_n)_n$ telle que $\mu_{\psi_n} \implies \mu^*$. Comme $f \in \mathcal{C}_b, \int f d\mu_{\psi_n} \rightarrow \int f d\mu^*$, ce qui contredit le fait que $\forall n, |\int f d\mu_{\psi_n} - \int f d\mu^*| > \varepsilon$. □

Fonction caractéristique, TCL

La fonction caractéristique d'une mesure de proba μ sur \mathbf{R}^d est

$$\varphi_\mu: \begin{array}{ll} \mathbf{R}^d & \rightarrow \mathbf{C} \\ t & \mapsto \int e^{i\langle t|x \rangle} d\mu(x) \end{array}$$

...

Rappel: $\varphi_\mu = \varphi_\nu \implies \mu = \nu$.

Ex. $\varphi_{\mathcal{N}(0,1)}(t) = e^{-t^2/2}$.

Pour $Y = AX + b$ on a $\varphi_Y(t) = e^{i\langle t|b \rangle} \varphi_X(A^T t)$.

Prop. φ_μ est continue en zéro.

Th (de Lévy). Soit $(\mu_n)_n, \mu$ des mesures de probabilité sur \mathbf{R}^d . $\mu_n \implies \mu$ ssi $\forall t \in \mathbf{R}^d, \varphi_{\mu_n}(t) \rightarrow \varphi_\mu(t)$.

Démonstration. Première implication: Si $X_n \xrightarrow{\mathcal{L}} X$ signifie $\forall f \in \mathcal{C}_b(\mathbf{R}^d), \mathbf{E}(f(X_n)) \rightarrow \mathbf{E}(f(X))$. Soit $t \in \mathbf{R}^d$ fixé. $f: x \mapsto e^{i\langle t|x \rangle}$ est continue bornée. Donc

$$\mathbf{E}(f(X_n)) \rightarrow \mathbf{E}(f(X))$$

$$\mathbf{E}(e^{i\langle t|X_n \rangle}) \rightarrow \mathbf{E}(e^{i\langle t|X \rangle}).$$

Deuxième implication : Supposons

$$\forall t, \phi_{X_n}(t) \longrightarrow \phi_X(t) \quad (*)$$

On note $\mu_n = \mathbf{P} \circ X_n^{-1}$ et $\mu = \mathbf{P} \circ X^{-1}$ les lois de X_n et X . $(*)$ implique que (μ_n) est tendue.

Conclusion de la preuve : montrons que $\mu_n \longrightarrow \mu$. Choisissons (μ_{ϕ_n}) convergeant faiblement, disons $\mu_{\phi_n} \Longrightarrow \nu$. Il suffit de montrer $\mu = \nu$. Si $\mu_{\phi_n} \Longrightarrow \nu$, alors $\forall t, \phi_{\mu_{\phi_n}}(t) \longrightarrow \phi_\nu(t)$. Or $\forall t, \phi_{\mu_n}(t) \longrightarrow \phi_\mu(t)$ par hypothèse. Donc $\forall t, \phi_\mu(t) = \phi_\nu(t)$. Donc $\mu = \nu = \lim_n \mu_n$.

Reste à montrer que $\forall t, \phi_{\mu_n}(t) \longrightarrow \phi_\mu(t)$ implique que (μ_n) est tendue. Dans le cas $d = 1$, pour tout $\mathcal{A} \subset \mathbf{R}$ on a :

$$\begin{aligned} \frac{1}{a} \int_{-a}^a (1 - \phi_{\mu_n}(t)) dt &= \frac{1}{a} \int_{-a}^a \left(1 - \int e^{itx} d\mu_n(x) \right) dt \\ &= \frac{1}{a} \int_{-a}^a \int (1 - e^{itx}) d\mu_n(x) dt \\ &= \int \frac{1}{a} \int_{-a}^a (1 - e^{itx}) dt d\mu_n(x) \\ &= \int \frac{1}{a} \left[t - \frac{e^{itx}}{ix} \right]_{-a}^a d\mu_n(x) \\ &= \int \left(2 - \frac{e^{iax} - e^{-iax}}{iax} \right) d\mu_n(x) \\ &= 2 \int \underbrace{\left(1 - \frac{\sin(|ax|)}{|ax|} \right)}_{\geq 0} d\mu_n(x) \\ &\geq 2 \int_{\mathcal{A}} \left(1 - \frac{\sin(|ax|)}{|ax|} \right) d\mu_n(x) \\ &\geq \int_{\mathcal{A}} 2 \left(1 - \frac{1}{|ax|} \right) d\mu_n(x) \end{aligned}$$

On choisit $\mathcal{A} = \left\{ x \in \mathbf{R} \mid 2 \left(1 - \frac{1}{|ax|} \right) > 1 \right\}$.

Alors

$$\begin{aligned} 2 \left(1 - \frac{1}{|ax|} \right) &\iff 1 - \frac{1}{a|x|} > \frac{1}{2} \\ &\iff \frac{1}{a|x|} < \frac{1}{2} \end{aligned}$$

□

Th (Procédé de Cramer-Wold). Soit X_n, X des v.a. sur \mathbf{R}^d . On a $X_n \xrightarrow{\mathcal{L}} X \iff \forall t, \langle t | X_n \rangle \xrightarrow{\mathcal{L}} \langle t | X \rangle$.

Démonstration. La première implication est donnée par le théorème de continuité.

Implication réciproque : Soit $t \in \mathbf{R}^d$ tel que $\langle t | X_n \rangle \xrightarrow{\mathcal{L}} \langle t | X \rangle$. Par le théorème de Lévy on a : $\forall u \in \mathbf{R}, \phi_{\langle t | X_n \rangle}(u) \longrightarrow \phi_{\langle t | X \rangle}(u)$. Donc $\phi_{\langle t | X_n \rangle}(1) \longrightarrow \phi_{\langle t | X \rangle}(1)$. Donc $\mathbf{E}(e^{i\langle t | X_n \rangle}) \longrightarrow \mathbf{E}(e^{i\langle t | X \rangle})$, d'où $\phi_{X_n}(t) \longrightarrow \phi_X(t)$. □

Théorème centrale limite

Not. $\mathcal{N}(m, \sigma^2)$ désigne la loi de densité $\rho(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-m)^2}{2\sigma^2}}$, et si $\sigma^2 = 0$ c'est la loi δ_m .

Pour X un vecteur gaussien, sa fonction caractéristique vérifie $\phi_X(t) = e^{i\langle t | m \rangle} e^{-\frac{t^T \Sigma t}{2}}$ où ...

Th (central limite). ...