

Développement : Théorème central limite

Leçons : 224, 239, 260, 261, 262.

Théorème. Soit $(X_n)_{n \in \mathbb{N}^*}$ suite de variables aléatoires iid dans L^2 . Alors

$$\sqrt{\frac{n}{\text{Var}(X_1)}} \left(\frac{S_n}{n} - \mathbb{E}[X_1] \right) \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \mathcal{N}(0, 1)$$

Démonstration.

Étape 1 : (Lemme) Pour tout $z \in \mathbb{C}$, $\left| e^z - \left(1 + \frac{z}{n} \right)^n \right| \leq e^{|z|} - \left(1 + \frac{|z|}{n} \right)^n$.

Formule du binôme : $e^z - \left(1 + \frac{z}{n} \right)^n = \sum_{k=0}^{\infty} \frac{z^k}{k!} - \sum_{k=0}^n \binom{n}{k} \frac{z^k}{n^k} = \sum_{k=n+1}^{\infty} \frac{z^k}{k!} + \sum_{k=0}^n \left[\frac{1}{k!} - \frac{\binom{n}{k}}{n^k} \right] z^k$

$= \sum_{k=n+1}^{\infty} \frac{z^k}{k!} + \sum_{k=0}^n \left[1 - \prod_{j=1}^{k-1} \left(1 - \frac{j}{n} \right) \right] \frac{z^k}{k!}$, et comme le terme entre crochets est positif, on a

$\left| e^z - \left(1 + \frac{z}{n} \right)^n \right| \leq \sum_{k=0}^{\infty} \frac{|z|^k}{k!} + \sum_{k=0}^n \left[1 - \prod_{j=1}^{k-1} \left(1 - \frac{j}{n} \right) \right] \frac{|z|^k}{k!} = e^{|z|} - \left(1 + \frac{|z|}{n} \right)^n$ en faisant le raisonnement inverse.

En particulier, si $(z_n)_{n \in \mathbb{N}} \in \mathbb{C}^{\mathbb{N}} \xrightarrow[n \rightarrow +\infty]{} \alpha$, on a $\left| e^{z_n} - \left(1 + \frac{z_n}{n} \right)^n \right| \leq e^{|z_n|} - \left(1 + \frac{|z_n|}{n} \right)^n \xrightarrow[n \rightarrow +\infty]{} 0$, et $\left(1 + \frac{z_n}{n} \right)^n \xrightarrow[n \rightarrow +\infty]{} e^\alpha$.

Étape 2 : Développement limité de la fonction caractéristique.

Quitte à poser pour tout i , $X'_i = \frac{X_i - \mathbb{E}[X_i]}{\text{Var}(X_i)}$, on peut supposer les $(X_i)_i$ centrées réduites.

On pose, pour tout $n \geq 1$, $Z_n = \frac{S_n}{\sqrt{n}}$. Alors, par indépendance, on a $\varphi_{S_n} = \prod_{i=1}^n \varphi_{X_i}$, et comme les X_i sont iid, on a

$\forall t \in \mathbb{R}, \varphi_{S_n}(t) = \prod_{i=1}^n \varphi_{X_1}(t) = \varphi_{X_1}(t)^n$. Il s'ensuit que $\varphi_{Z_n}(t) = \mathbb{E} \left[e^{it \frac{S_n}{\sqrt{n}}} \right] = \varphi_{S_n} \left(\frac{t}{\sqrt{n}} \right) = \varphi_{X_1} \left(\frac{t}{\sqrt{n}} \right)^n$.

Alors, comme $\left| \frac{d^2}{dt^2} (e^{itX_1}) \right| = X_1^2$ qui est intégrable, le théorème de dérivation des intégrales à paramètre donne que φ_{X_1} est deux fois dérivable, et que $\varphi'_{X_1}(0) = \mathbb{E}[iX] = 0$, $\varphi''_{X_1}(0) = \mathbb{E}[-X^2] = -1$. On a donc, d'après la formule de Taylor-Young, $\varphi_X(t) = 1 - \frac{t^2}{2} + o(t^2)$. Par suite, lorsque $n \rightarrow +\infty$, $\varphi_{Z_n}(t) = \left(1 - \frac{t^2}{2n} + o\left(\frac{1}{n}\right) \right)^n = \left(1 - \frac{t^2}{2n} + \frac{\varepsilon_n}{n} \right)^n$, où $\varepsilon_n \xrightarrow[n \rightarrow +\infty]{} 0$.

Étape 3 : Conclusion.

On pose $z_n = \frac{t^2}{2} + \varepsilon_n \xrightarrow[n \rightarrow +\infty]{} \frac{t^2}{2}$. Alors, d'après le lemme, $\varphi_{Z_n}(t) = \left(1 + \frac{z_n}{n} \right)^n \xrightarrow[n \rightarrow +\infty]{} e^{-\frac{t^2}{2}}$. On reconnaît la fonction caractéristique de la loi $\mathcal{N}(0, 1)$, d'où le résultat. □

Compléments

Intervalle de confiance asymptotiques

Soit $(X_n)_{n \geq 1}$ suite de variables aléatoires iid de loi $\mathcal{B}(\theta)$ où $\theta \in \mathbb{R}$ est inconnu. On définit, pour $n \geq 1$, la moyenne empirique $\bar{X}_n = \frac{1}{n} \sum_{k=1}^n X_k$, on peut montrer que c'est un estimateur faiblement consistant de θ (il converge en probabilité vers θ), sans biais (son espérance vaut θ).

Le théorème central limite donne alors $\sqrt{\frac{n}{\theta(1-\theta)}} (\bar{X}_n - \theta) \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} Z$ où Z suit une loi $\mathcal{N}(0, 1)$. Soit $\alpha \in]0, 1[$. Si on note q_α le quantile d'ordre α de la loi $\mathcal{N}(0, 1)$, on a $\mathbb{P}(|Z| \geq q_\alpha) = \alpha$. Comme la convergence en loi implique la convergence des fonctions de répartition, on a $\mathbb{P} \left(\left| \sqrt{\frac{n}{\theta(1-\theta)}} (\bar{X}_n - \theta) \right| \leq q_\alpha \right) \xrightarrow[n \rightarrow +\infty]{} \mathbb{P}(|Z| \leq q_\alpha) = 1 - \alpha$,

c'est-à-dire $\mathbb{P}\left(\theta \in \left[\bar{X}_n - \sqrt{\frac{\theta(1-\theta)}{n}}q_\alpha, \bar{X}_n + \sqrt{\frac{\theta(1-\theta)}{n}}q_\alpha\right]\right) \xrightarrow{n \rightarrow +\infty} 1 - \alpha$. On obtient un intervalle de confiance asymptotique de niveau $1 - \alpha$. On remarque que celui-ci dépend de θ ; on peut majorer $\theta(1 - \theta)$ par $\frac{1}{4}$ pour obtenir $\lim_{n \rightarrow \infty} \mathbb{P}\left(\theta \in \left[\bar{X}_n - \frac{q_\alpha}{\sqrt{2n}}, \bar{X}_n + \frac{q_\alpha}{\sqrt{2n}}\right]\right) \geq 1 - \alpha$.

Un équivalent de suite

On pose pour tout $n \in \mathbb{N}$, $u_n = \sum_{k=0}^n \frac{n^k}{k!}$ et $v_n = e^{-n}u_n$. Soit $(X_n)_{n \in \mathbb{N}^*}$ suite de variables aléatoires iid de loi $\mathcal{P}(1)$,

on pose $Y_n = \sum_{k=1}^n X_k$. Alors Y_n suit une loi de Poisson de paramètre n . Ainsi $v_n = \mathbb{P}(Y_n \leq n)$.

D'après le théorème central limite, $\sqrt{n} \left(\frac{Y_n}{n} - 1\right) \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} Z$ où Z suit une loi $\mathcal{N}(0, 1)$. On en déduit que

$$v_n = \mathbb{P}(Y_n \leq n) = \mathbb{P}\left(\sqrt{n} \left(\frac{Y_n}{n} - 1\right) \leq 0\right) \xrightarrow{n \rightarrow +\infty} \mathbb{P}(Z \leq 0) = \frac{1}{2} \text{ Donc } u_n = \sum_{k=0}^n \frac{n^k}{k!} \underset{n \rightarrow +\infty}{\sim} \frac{e^n}{2}.$$

Références

M. Barbe, P. Ledoux, *PROBABILITÉ*, EDP Sciences, p. 131.

J.-Y. Ouvrard, *PROBABILITÉS TOME 2*, Cassini, p. 310.

Notes personnelles

Le lemme sert surtout à éviter de passer par le logarithme complexe. Selon la leçon, on peut présenter l'une ou l'autre des applications (trop court sinon).