

Théorème central limite

Leçons : 218, 261, 262, 263

Théorème 1

Si $(X_i)_i$ est une suite de variables aléatoires iid et $E(X_0^2) < \infty$ alors $\frac{S_n - nE(X_0)}{\sqrt{n}\sigma}$ converge en loi vers $\mathcal{N}(0, 1)$ où $\sigma^2 = \text{Var}(X_0)$ et $S_n = \sum_{i=0}^{n-1} X_i$.

Lemme 2

Si $(z_n)_n$ est une suite de nombres complexes tendant vers 0, alors $\lim_{n \rightarrow +\infty} \left(1 + \frac{z_n}{n}\right)^n = 1$.

Démonstration. On a

$$\left| \left(1 + \frac{z_n}{n}\right)^n - 1 \right| = \left| \sum_{k=1}^n \binom{n}{k} \frac{z_n^k}{n^k} \right| \leq \sum_{k=1}^n \binom{n}{k} \frac{|z_n|^k}{n^k} = \left(1 + \frac{|z_n|}{n}\right)^n - 1.$$

On est donc ramené au cas où $\forall n \in \mathbb{N}, z_n \in \mathbb{R}$. A partir d'un certain rang, $z_n > -1$ et $\ln\left(\left(1 + \frac{z_n}{n}\right)^n\right) = n \ln\left(1 + \frac{z_n}{n}\right) = n\left(\frac{z_n}{n} + o\left(\frac{z_n}{n}\right)\right) = z_n + o(z_n)$

Donc en passant à l'exponentielle, la limite de $\left(1 + \frac{z_n}{n}\right)^n$ est bien 1. □

Démonstration (du théorème). Quitte à remplacer X_i par $\frac{X_i - E(X_0)}{\sigma}$, on peut supposer que $E(X_0) = 0$ et $\sigma = 1$. On note $X = X_0$. La fonction caractéristique de $\frac{S_n}{\sqrt{n}}$ est

$$\varphi_{S_n/\sqrt{n}}(t) = \left(\varphi_{X/\sqrt{n}}(t)\right)^n = \left(\varphi_X\left(\frac{t}{\sqrt{n}}\right)\right)^n$$

par indépendance des X_i .

De plus on sait que $\varphi'_X(0) = iE(X) = 0$, $\varphi''_X(0) = -E(X^2) = -1$. Donc selon la formule de Taylor-Young à l'ordre 2 au voisinage de 0, on a $\varphi_X(t) = 1 - \frac{t^2}{2} + t^2\varepsilon(t)$ où $\lim_{t \rightarrow 0} \varepsilon(t) = 0$.
Donc

$$\varphi_{S_n/\sqrt{n}}(t) = \left(1 - \frac{t^2}{2n} + \frac{t^2}{n}\varepsilon\left(\frac{t}{\sqrt{n}}\right)\right)^n = \left(1 - \frac{t^2}{2n}\right)^n \left(1 + \frac{t^2}{n\left(1 - \frac{t^2}{2n}\right)}\varepsilon\left(\frac{t}{\sqrt{n}}\right)\right)^n$$

Le premier terme tend vers $e^{-t^2/2}$, et selon le lemme, comme

$$\lim_{n \rightarrow +\infty} \frac{t^2}{1 - \frac{t^2}{2n}} = 1,$$

le second terme tend vers 1 quand n tend vers l'infini. Donc pour tout $t \in \mathbb{R}$, $\varphi_{S_n/\sqrt{n}}(t) \rightarrow e^{-t^2/2}$, qui est la fonction caractéristique de $\mathcal{N}(0, 1)$. Selon le théorème de Paul-Lévy, on a le résultat voulu. □

Exemple. Dans le cadre d'un sondage pour le deuxième tour de l'élection présidentielle, on interroge n personnes pour savoir s'ils comptent voter pour le candidat A (n_A personnes) ou le candidat B (n_B personnes). Chaque personne est modélisée par une variable aléatoire $X_i : \Omega \rightarrow \{A, B\}$ suivant la loi de Bernoulli de paramètre p . On suppose que les X_i sont indépendantes. C'est ce paramètre p qui nous est inconnu et qu'on cherche à déterminer *a posteriori* à partir des données observables (les valeurs des X_i). Cela est rendu possible par le théorème central limite.

Un *intervalle de confiance* à α près pour p est un intervalle **aléatoire** I de \mathbb{R} tel que $\mathbb{P}(p \in I) = 1 - \alpha$.

Soit G une variable gaussienne de loi $\mathcal{N}(0, 1)$ de fonction de répartition ϕ . On note $\bar{X}_n = \frac{X_1 + \dots + X_n}{n}$ (dans la modélisation ci-dessus, \bar{X}_n correspond à $\frac{n_A}{n}$). On sait que la loi de cette variable aléatoire est $\mathcal{B}(n, p)$ donc selon le TCL, $Y_n = \sqrt{n} \frac{\bar{X}_n - p}{\sqrt{p(1-p)}}$ converge en loi vers G .

De plus, selon la loi faible des grands nombres, $\bar{X}_n \xrightarrow[n \rightarrow +\infty]{\mathbb{P}} p$ donc selon le lemme de Slutsky, $(Y_n, \bar{X}_n) \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} (G, p)$ de sorte qu'en composant par des fonctions continues, on obtient

$$Y_n \frac{\sqrt{p(1-p)}}{\sqrt{\bar{X}_n(1-\bar{X}_n)}} = \frac{\sqrt{n}(\bar{X}_n - p)}{\sqrt{\bar{X}_n(1-\bar{X}_n)}} \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} G \times \frac{\sqrt{p(1-p)}}{\sqrt{p(1-p)}} = G.$$

Donc en vertu de la caractérisation de la convergence en loi avec les fonctions de répartitions, pour tout $a \geq 0$,

$$\mathbb{P} \left(-a \leq \frac{\sqrt{n}(\bar{X}_n - p)}{\sqrt{\bar{X}_n(1-\bar{X}_n)}} \leq a \right) \xrightarrow[n \rightarrow +\infty]{} C_a = \phi(a) - \phi(-a),$$

c'est-à-dire, avec $I_{a,n} = \left[-a \frac{\sqrt{\bar{X}_n(1-\bar{X}_n)}}{\sqrt{n}} + \bar{X}_n, a \frac{\sqrt{\bar{X}_n(1-\bar{X}_n)}}{\sqrt{n}} + \bar{X}_n \right]$, $\mathbb{P}(p \in I_{a,n}) \xrightarrow[n \rightarrow +\infty]{} C_a$.

En prenant C_a suffisamment petit, à partir d'un certain rang, $\mathbb{P}(p \in I_{a,n}) \leq 1 - \alpha : I_{a,n}$ est un intervalle de confiance à α près.

Complétons ce théorème par un autre théorème limite, le théorème des événements rares de Poisson. Il faut faire attention car la preuve du Ouvrard présente une imprécision sur le logarithme complexe.

Théorème 3 (Théorème des événements rares)

Soit pour tout $n \in \mathbb{N}^*$ une famille finie $\{A_{n,j} \mid 1 \leq j \leq M_n\}$ d'événements indépendants.

On pose $p_{n,j} = \mathbb{P}(A_{n,j})$ et on note $S_n = \sum_{j=1}^{M_n} \mathbb{1}_{A_{n,j}}$. On suppose que la suite de terme général M_n tend en croissant vers $+\infty$ et que

$$\sup_{1 \leq j \leq M_n} p_{n,j} \xrightarrow[n \rightarrow +\infty]{} 0, \quad \sum_{j=1}^{M_n} p_{n,j} \xrightarrow[n \rightarrow +\infty]{} \lambda$$

où $\lambda > 0$. Alors la suite $(S_n)_{n \in \mathbb{N}^*}$ converge en loi vers la loi de Poisson $\mathcal{P}(\lambda)$ de paramètre λ .

Démonstration. Soit $t \in \mathbb{R}$. Par indépendance des $A_{n,j}$, la fonction caractéristique de S_n est

$$\varphi_{S_n}(t) = \prod_{j=1}^{M_n} (p_{n,j} e^{it} + (1 - p_{n,j})) = \prod_{j=1}^{M_n} [1 + p_{n,j}(e^{it} - 1)].$$

On pose $z = e^{it} - 1$. Comme $\sup_{1 \leq j \leq M_n} p_{n,j} \xrightarrow{n \rightarrow +\infty} 0$, il existe $N \geq 1$ tel que pour $n \geq N$, et $j \in \{1, \dots, M_n\}$, $|p_{n,j}z| < \frac{1}{2}$ de sorte que $1 + p_{n,j}z = \exp \{ \log [1 + p_{n,j}z] \}$ où \log désigne une détermination principale du logarithme complexe sur le plan fendu de Cauchy $\mathbb{C} \setminus \mathbb{R}_-$.

Or, par la formule de Taylor avec reste intégral, si $|w| < 1$,

$$\log(1 + w) = w - w^2 \int_0^1 (1 - u) \frac{1}{(1 + uw)^2} du$$

Si $n \geq N$, et $u \in [0, 1]$, $j \in \llbracket 1, M_n \rrbracket$, $|1 + up_{n,j}z| \geq 1 - up_{n,j}|z| \geq 1 - \frac{u}{2} \geq \frac{1}{2}$, d'où

$$\left| \sum_{j=1}^{M_n} p_{n,j}^2 \int_0^1 \frac{1 - u}{(1 + up_{n,j}z)^2} du \right| \leq 2 \left| \sum_{j=1}^{M_n} p_{n,j}^2 \right| \leq 2 \left[\sup_{1 \leq j \leq M_n} p_{n,j} \right] \left[\sum_{j=1}^{M_n} p_{n,j} \right] \xrightarrow{n \rightarrow +\infty} 0.$$

Donc par la formule de Taylor précédente et par hypothèse,

$$\sum_{j=1}^{M_n} \log(1 + p_{n,j}z) = z \sum_{j=1}^{M_n} p_{n,j} - z^2 \sum_{j=1}^{M_n} p_{n,j}^2 \int_0^1 \frac{1 - u}{(1 + up_{n,j}z)^2} du \xrightarrow{n \rightarrow +\infty} \lambda z.$$

Donc

$$\exp \left(\sum_{j=1}^{M_n} \log(1 + p_{n,j}z) \right) \xrightarrow{n \rightarrow +\infty} \exp(\lambda z) = \exp(\lambda(e^{it} - 1)),$$

fonction caractéristique d'une loi de Poisson $\mathcal{P}(\lambda)$.

Mais $\exp \left(\sum_{j=1}^{M_n} \log(1 + p_{n,j}z) \right) = \prod_{j=1}^{M_n} \exp(\log(1 + p_{n,j}z)) = \prod_{j=1}^{M_n} 1 + p_{n,j}z = \varphi_{S_n}(t)$, donc selon le théorème de Lévy, $(S_n)_{n \in \mathbb{N}^*}$ converge en loi vers $\mathcal{P}(\lambda)$. \square

Remarque. Pour compléter ce développement un peu court, on peut calculer la fonction caractéristique d'une loi normale (cf « Inversion de Fourier »). Dans la leçon 218, il est bienvenu de mentionner le théorème des événements rares de Poisson plutôt que la recherche d'un intervalle de confiance.

Références :

- Philippe BARBÉ et Michel LEDOUX (2007). *Probabilité*. EDP Sciences, pp. 136-138.
- Jean-Yves OUVRARD (2009). *Probabilités (Master Agrégation)*. T. 2. Cassini, p. 311 (événements rares).