Théorème central limite et intervalle de confiance

Lemme 1. *Pour tout $z \in \mathbb{C}$, on a* $\left|e^z - \left(1+\frac{z}{n}\right)^n\right| \leqslant e^{|z|} - \left(1+\frac{|z|}{n}\right)^n$.

Démonstration.

En posant $\alpha_n^k = 1 - \frac{n!}{(n-k)!n^k}\mathbb{1}_{k\leqslant n} \in [0,1]$, on a :

$$\left|e^z - \left(1+\frac{z}{n}\right)^n\right| = \left|\sum_{k\geqslant 0}\frac{z^k}{k!} - \sum_{k=0}^{n}\frac{z^k}{n^k}\right| = \left|\sum_{k=0}^{n}\frac{z^k}{k!}\alpha_n^k\right| \leqslant \sum_{k=0}^{n}\frac{|z|^k}{k!}\alpha_n^k = e^{|z|} - \left(1+\frac{|z|}{n}\right)^n$$

□

Théorème 2 (Théorème central limite)**.** *Soit $(X_n)_{n\in\mathbb{N}^\star}$ une suite de variables aléatoires indépendantes, identiquement distribuées et de carré intégrable. Alors :*

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{X_i - \mathbb{E}\left[X_i\right]}{\sqrt{\mathrm{Var}(X_i)}} \quad \xrightarrow[n\to+\infty]{\mathcal{L}} \quad \mathcal{N}(0,1)$$

Démonstration.

Si les $(X_n)_{n\in\mathbb{N}^\star}$ sont constantes presque sûrement, le résultat est clair. On va donc les supposer non constantes presque sûrement. De plus, quitte à remplacer X_i par $\frac{X_i - \mathbb{E}[X_i]}{\mathrm{Var}(X_i)}$, on peut supposer que $\mathbb{E}\left[X_i\right] = 0$ et $\mathrm{Var}(X_i) = 1$ pour tout $i \in \mathbb{N}^\star$. Posons alors $S_n = \sum_{i=1}^n X_i$.

Fixons $t \in \mathbb{R}$. Par indépendance et équidistribution, on a alors :

$$\varphi_{\frac{S_n}{\sqrt{n}}}(t) = \mathbb{E}\left[e^{it\frac{S_n}{\sqrt{n}}}\right] = \prod_{i=1}^{n}\mathbb{E}\left[e^{it\frac{X_i}{\sqrt{n}}}\right] = \varphi_{X_1}\left(\frac{t}{\sqrt{n}}\right)^n$$

X_1 étant de carré intégrable, φ_{X_1} est deux fois dérivable sur $\mathbb{R}$, avec :

$$\varphi'_{X_1}(0) = i\mathbb{E}\left[X_1\right] = 0 \quad \text{et} \quad \varphi''_{X_1}(0) = -\mathbb{E}\left[X_1^2\right] = -\mathrm{Var}(X_1) = -1$$

On obtient alors le développement de Taylor-Young à l'ordre 2 en 0 de φ_{X_1} :

$$\varphi_{X_1} = 1 - \frac{u^2}{2} + \mathrm{o}\left(u^2\right)$$

On a donc :

$$\varphi_{\frac{S_n}{\sqrt{n}}}(t) = \left(1 - \frac{t^2}{2n} + \mathrm{o}\left(\frac{1}{n}\right)\right)^n = \left(1+\frac{z_n}{n}\right)^n \quad \text{où} \quad z_n = -\frac{t^2}{2}\Big(1+\mathrm{o}(1)\Big) \in \mathbb{C}$$

On a de plus :

$$e^{|z_n|} - \left(1+\frac{|z_n|}{n}\right)^n = e^{\frac{t^2}{2}|1+\mathrm{o}(1)|} - e^{|z_n|(1+\mathrm{o}(1))} \xrightarrow[n\to+\infty]{} e^{\frac{t^2}{2}} - e^{\frac{t^2}{2}} = 0$$

Ainsi, par le Lemme 1, on a :

$$\varphi_{\frac{S_n}{\sqrt{n}}}(t) \underset{n\to+\infty}{\sim} e^{z_n} \xrightarrow[n\to+\infty]{} e^{-\frac{t^2}{2}}$$

Le théorème de Lévy permet alors de conclure que $\frac{S_n}{\sqrt{n}} \xrightarrow[n\to+\infty]{\mathcal{L}} \mathcal{N}(0,1)$. □

Application 3. *Soit $(X_n)_{n\in\mathbb{N}^\star}$ une suite de variables aléatoires indépendantes, identiquement distribuées et de loi $\mathcal{B}(p)$ pour $p \in [0,1]$ inconnu. Le théorème central limite donne un intervalle de confiance asymptotique de niveau α pour p en fonction de la moyenne empirique $\widehat{p_n} = \frac{1}{n}\sum_{k=1}^n X_i$. Il s'agit de :*

$$IC_\alpha = \left[\widehat{p_n} \pm \frac{q_{1-\frac{\alpha}{2}}}{2\sqrt{n}}\right]$$

où q_t est le quantile d'ordre t de $\mathcal{N}(0,1)$.

Démonstration.

Les $(X_n)_{n\in\mathbb{N}^\star}$ satisfont les hypothèses du théorème central limite, avec $\mathbb{E}[X_1] = p$ et $\mathrm{Var}(X_1) = p(1-p)$. Ainsi :

$$\frac{\sqrt{n}}{\sqrt{p(1-p)}}(\widehat{p} - p) = \frac{1}{\sqrt{n}}\sum_{i=1}^n \frac{X_i - p}{\sqrt{p(1-p)}} \quad \xrightarrow[n\to+\infty]{\mathcal{L}} \quad \mathcal{N}(0,1)$$

La fonction de répartition de $\mathcal{N}(0,1)$ étant continue, on a, pour tous $a, b \in \mathbb{R}$:

$$\mathbb{P}\left(a \leqslant \frac{\sqrt{n}}{\sqrt{p(1-p)}}(\widehat{p} - p) \leqslant b\right) \xrightarrow[n\to+\infty]{} \mathbb{P}\left(a \leqslant \mathcal{N}(0,1) \leqslant b\right)$$

En prenant $a = q_{\frac{\alpha}{2}}$ et $b = q_{1-\frac{\alpha}{2}}$, on a $a = -b$, et donc :

$$\mathbb{P}\left(\left|\frac{\sqrt{n}}{\sqrt{p(1-p)}}(\widehat{p} - p)\right| \leqslant q_{1-\frac{\alpha}{2}}\right) \xrightarrow[n\to+\infty]{} \mathbb{P}\left(q_{\frac{\alpha}{2}} \leqslant \mathcal{N}(0,1) \leqslant q_{1-\frac{\alpha}{2}}\right) = 1 - \alpha$$

Ce que l'on réécrit comme :

$$\mathbb{P}\left(p \in \left[\widehat{p_n} \pm q_{1-\frac{\alpha}{2}}\frac{\sqrt{p(1-p)}}{\sqrt{n}}\right]\right) \xrightarrow[n\to+\infty]{} 1 - \alpha$$

On en déduit le résultat annoncé en remarquant que $p(1-p) \leqslant \frac{1}{4}$. ☐

Références

[BL07] Philippe Barbe and Michel Ledoux. *Probabilité.* EDP Sciences, 2007