49. Théorème central limite et intervalle de confiance

[BL07, §V.5, p136–137] [RS12, §3.4, p25]

ÉNONCÉ

THÉORÈME. [THÉORÈME CENTRAL LIMITE]
Si les $(X_i)_{i\in\mathbb{N}^*}$ sont i.i.d. de carré intégrable, alors :

$$\frac{1}{\sqrt{n}}\sum_{i=1}^n (X_i - \mathbb{E}[X_1]) \xrightarrow[n\to+\infty]{\mathcal{L}} Z \sim \mathcal{N}(0, \mathrm{Var}(X_1))$$

APPLICATION. [INTERVALLE DE CONFIANCE ASYMPTOTIQUE]
Soient $(X_i)_{i\in\mathbb{N}^*}$ des réalisations i.i.d. de $\mathcal{B}(p)$ pour un $p\in[0,1]$ inconnu.
$\mathrm{IC}_\alpha(p) = \left[\hat{p}_n \pm \frac{q_{1-\alpha/2}}{2\sqrt{n}}\right]$ est un intervalle de confiance asymptotique de niveau α pour p, où $\hat{p}_n = \frac{1}{n}\sum_{k=1}^n X_i$ est la moyenne empirique et q_t est le quantile d'ordre t de $\mathcal{N}(0,1)$.

DÉVELOPPEMENT

Si les $(X_i)_{i\in\mathbb{N}^*}$ sont constantes p.s., le résultat est clair.
Sinon on peut supposer $\mathbb{E}[X_i]=0$ et $\mathrm{Var}(X_i)=1$ quitte à remplacer X_i par $\frac{X_i-\mathbb{E}[X_i]}{\sqrt{\mathrm{Var}(X_i)}}$.

Posons $S_n = \sum_{i=1}^n X_i$ et montrons que $\forall t\in\mathbb{R}, \phi_{\frac{S_n}{\sqrt{n}}}(t) \xrightarrow[n\to+\infty]{} \mathrm{e}^{-\frac{t^2}{2}}$.

Fixons $t\in\mathbb{R}$. On a :

$$\begin{aligned}\phi_{\frac{S_n}{\sqrt{n}}}(t) &= \mathbb{E}\left[\mathrm{e}^{it\frac{S_n}{\sqrt{n}}}\right] = \prod_{i=1}^n \mathbb{E}\left[\mathrm{e}^{it\frac{X_i}{\sqrt{n}}}\right] && \text{par indépendance}\\ &= \phi_{X_1}\left(\frac{t}{\sqrt{n}}\right)^n && \text{car les } (X_i)_{i\in\mathbb{N}^*} \text{ sont i.i.d.}\end{aligned}$$

X_1 étant de carré intégrable, on peut appliquer le théorème de dérivation sous le signe intégral deux fois, et on obtient que ϕ_{X_1} est deux fois dérivable sur $\mathbb{R}$, avec $\phi_{X_1}'(0) = i\mathbb{E}[X]=0$ et $\phi_{X_1}''(0) = -\mathbb{E}[X^2] = -1$. De plus, le théorème de continuité sous le signe intégral donne que ϕ_{X_1} est $\mathcal{C}^2$, et donc on peut appliquer la formule de TAYLOR-YOUNG à l'ordre 2 en 0 :

$$\forall u\in\mathbb{R}, \phi_{X_1}(u) = 1 - \frac{u^2}{2} + o(u^2)$$

On a donc $\phi_{\frac{S_n}{\sqrt{n}}}(t) = \left(1-\frac{t^2}{2n}+o(1/n)\right)^n = \left(1+\frac{z_n}{n}\right)^n$, où $z_n = -\frac{t^2}{2}(1+o(1))\in\mathbb{C}$.

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

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

$$\mathrm{e}^z - \left(1+\frac{z}{n}\right)^n = \sum_{k\ge0}\frac{z^k}{k!} - \sum_{k=0}^n \binom{n}{k}\frac{z^k}{n^k} = \sum_{k=0}^n \frac{z^k}{k!}\alpha_n^k$$

$$\text{d'où } \left|\mathrm{e}^z - \left(1+\frac{z}{n}\right)^n\right| \le \sum_{k=0}^n \frac{|z|^k}{k!}\alpha_n^k = \mathrm{e}^{|z|} - \left(1+\frac{|z|}{n}\right)^n$$

Revenons à notre suite de complexes $(z_n)_{n\in\mathbb{N}}$. On a $\mathrm{e}^{|z_n|} \xrightarrow[n\to+\infty]{} \mathrm{e}^{\frac{t^2}{2}}$ et

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

Ainsi, par le Lemme, $\phi_{\frac{S_n}{\sqrt{n}}}(t) \sim_{n\to+\infty} \mathrm{e}^{z_n} \xrightarrow[n\to+\infty]{} \mathrm{e}^{-\frac{t^2}{2}}$.
Le théorème de LÉVY permet alors de conclure que $S_n/\sqrt{n} \xrightarrow[n\to+\infty]{\mathcal{L}} \mathcal{N}(0,1)$.

Passons à l'application. Les $(X_i)_{i\in\mathbb{N}^*}$ satisfont les hypothèses du théorème central limite et on a $\mathbb{E}[X_1]=p$, $\mathrm{Var}(X_1)=p(1-p)$. Ainsi :

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

La fonction de répartition de $\mathcal{N}(0,1)$ est continue donc le théorème de PORTMANTEAU donne :

$$\forall a<b\in\mathbb{R}, \mathbb{P}\left(a\le \frac{\sqrt{n}}{\sqrt{p(1-p)}}(\hat{p}_n-p)\le b\right) \xrightarrow[n\to+\infty]{} \mathbb{P}(a\le\mathcal{N}(0,1)\le b)$$

Prenant $a = q_{\alpha/2}$ et $b = q_{1-\alpha/2}$, on a $a=-b$ et donc :

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

Ce que l'on réécrit $\mathbb{P}\left(p\in\left[\hat{p}_n \pm q_{1-\alpha/2}\frac{\sqrt{p(1-p)}}{\sqrt{n}}\right]\right) \xrightarrow[n\to+\infty]{} 1-\alpha$.

En remarquant que $p(1-p)\le\frac{1}{4}$, on en déduit l'intervalle de confiance asymptotique annoncé.
On aurait aussi pu utiliser le lemme de SLUTSKY, qui permet de remplacer p par $\hat{p}_n$ dans les bornes de l'intervalle.