14,24 vite

THÉORÈME CENTRAL LIMITE

ZQ p540 - 555 - 563. + Garet

Prérequis: . formule du binôme de Newton, série exp, Taylor avec reste intégral, Taylor Young
. thm d'holomorphie sous le signe intégral, principe de prolongement analytique, critère de Riemann
. thm de Levy, $X \in \mathcal{L}^n \Rightarrow \varphi \in \mathcal{C}^n$, formule de transfert.
. une suite cv est bornée, continuité séquentielle

Thm: Soit $(X_n)_{n\geq 1}$ une suite de v.a. iid dans $\mathcal{L}^2$. On note $S_n = \sum_{i=1}^n X_i$, $m = \mathbb{E}[X_1]$, $\sigma^2 = Var(X_1) > 0$

Alors $\dfrac{S_n - nm}{\sigma\sqrt{n}} \underset{n\to+\infty}{\Longrightarrow} \mathcal{N}(0,1)$.

Dém: Quitte à considérer $Y_n = \dfrac{X_n - m}{\sigma}$, on peut supposer $m=0$ et $\sigma = 1$. On cherche donc à montrer que $\dfrac{S_n}{\sqrt{n}} \Longrightarrow \mathcal{N}(0,1)$. Pour cela on va utiliser le thm de Levy.

Etape 1: Fonction caractéristique de $\mathcal{N}(0,1)$ Garet

Si $X \sim \mathcal{N}(0,1)$, pour $t\in\mathbb{R}$ $\varphi_X(t) = \dfrac{1}{\sqrt{2\pi}} \displaystyle\int_{\mathbb{R}} e^{itx - x^2/2} dx$ par formule de transfert

Pour $(z,x) \in \mathbb{C}\times\mathbb{R}$, on pose $f(z,x) = e^{zx} e^{-x^2/2}$. Soit $R>0$.

. $\forall z \in D(0,R)$, $x \mapsto f(z,x)$ est intégrable sur $\mathbb{R}$:

$\forall x \in \mathbb{R}$ $|e^{-\frac{x^2}{2}} e^{zx}| = e^{-\frac{x^2}{2}} e^{|x| Re z|} \leq e^{-x^2/2} e^{|x| R} \underset{x\to\pm\infty}{=} o\left(\dfrac{1}{x^2}\right)$ intégrable par critère de Riemann.

. $\forall x \in \mathbb{R}$ $z \mapsto f(z,x)$ est holomorphe sur $D(0,R)$

. $\forall z \in D(0,R)$ $\forall x \in \mathbb{R}$ $|e^{-x^2/2} e^{zx}| \leq \underbrace{e^{-\frac{x^2}{2}} e^{|x| R}}_{\text{indep de } z} \in \mathcal{L}^1$ d'après le 1er point.

Donc d'après le théorème d'holomorphie sous le signe intégral, $G : z \mapsto \dfrac{1}{\sqrt{2\pi}} \displaystyle\int_{\mathbb{R}} f(z,x) dx$ est holomorphe sur $D(0,R)$ pour tout $R>0$, donc sur $\mathbb{C}$.

Or si $u\in\mathbb{R}$, $G(u) = e^{u^2/2} \underbrace{\dfrac{1}{\sqrt{2\pi}} \displaystyle\int_{\mathbb{R}} e^{-\frac{(x-u)^2}{2}} dx}_{\text{densité de } \mathcal{N}(u,1)} = e^{u^2/2}$. Ainsi G et $z \mapsto e^{z^2/2}$ sont holomorphes et coïncident sur $\mathbb{R}$ (qui a un pt d'accumulation) donc sur $\mathbb{C}$ connexe par ppe de prolongement analytique.

En particulier, en $z = it$, $\varphi_X(t) = G(it) = e^{-t^2/2}$.

Etape 2: Developpement de Taylor de la fonction caractéristique de X_1

On va donc montrer que $\varphi_{\frac{S_n}{\sqrt{n}}}(t) \underset{n\to+\infty}{\longrightarrow} e^{-t^2/2}$ pour tout $t\in\mathbb{R}$.

Or $\varphi_{\frac{S_n}{\sqrt{n}}}(t) = \varphi_{S_n}\left(\dfrac{t}{\sqrt{n}}\right) = \left(\varphi_{X_1}\left(\dfrac{t}{\sqrt{n}}\right)\right)^n$ car les X_i sont iid.

Comme $X_1 \in \mathcal{L}^2$, $\varphi_{X_1} \in \mathcal{C}^2$ et on connait ses dérivées en fonctions des moments de X_1:

. $\varphi_{X_1}'(0) = i\mathbb{E}[X_1] = im = 0$

. $\varphi_{X_1}''(0) = i^2 \mathbb{E}[X_1^2] = -(Var(X_1) - \underbrace{\mathbb{E}[X_1]^2}_{=m=0}) = -\sigma^2 = -1$

donc par un dvlpt de Taylor de φ_{X_1} à l'ordre 2 en l'origine :

$$\varphi_{X_1}\left(\frac{t}{\sqrt{n}}\right) = \varphi_{X_1}(0) + \frac{t}{\sqrt{n}}\varphi'_{X_1}(0) + \frac{t^2}{2n}\varphi''_{X_1}(0) + \frac{\varepsilon_n}{n} \quad \text{où } \varepsilon_n \underset{n\to+\infty}{\longrightarrow} 0$$

$$= 1 - \frac{t^2}{2n} + \frac{\varepsilon_n}{n}.$$

Donc $\varphi_{\frac{S_n}{\sqrt{n}}}(t) = \left(1 + \frac{-\frac{t^2}{2} + \varepsilon_n}{n}\right)^n$ On veut passer à la limite qd $n \to +\infty$ mais une fct° caract est à valeurs complexes donc c'est complexe ds la parenthèse (caché dans le ε_n)

Etape 3: Si $(z_n) \in \mathbb{C}^{\mathbb{N}}$ cv vers $z \in \mathbb{C}$ alors $\left(1 + \frac{z_n}{n}\right)^n \underset{n\to+\infty}{\longrightarrow} e^z$

D'après la formule du binôme de Newton, on a pour $n \in \mathbb{N}$

$$e^{z_n} - \left(1 + \frac{z_n}{n}\right)^n = \sum_{k=0}^{+\infty} \frac{(z_n)^k}{k!} - \sum_{k=0}^{n} \binom{n}{k} \frac{z_n^k}{n^k} = \sum_{k=0}^{+\infty} a_k^{(n)} z_n^k \quad \text{où } a_k^{(n)} = \begin{cases} \frac{1}{k!}\left(1 - \frac{n(n-1)\cdots(n-k+1)}{n^k}\right) & k \leq n \\ \frac{1}{k!} & \text{si } k > n \end{cases}$$

Comme $1 - \underbrace{\underbrace{\frac{n}{n}}_{\leq 1} \underbrace{\frac{n-1}{n}}_{\leq 1} \cdots \underbrace{\frac{n-k+1}{n}}_{\leq 1}}_{\leq 1} \geq 0$, les $a_k^{(n)}$ sont positifs. donc

$$\left|e^{z_n} - \left(1 + \frac{z_n}{n}\right)^n\right| = \left|\sum_{k=0}^{+\infty} a_k^{(n)} z_n^k\right| \leq \sum_{k=0}^{+\infty} a_k^{(n)} |z_n|^k = e^{|z_n|} - \left(1 + \frac{|z_n|}{n}\right)^n$$

$$\leq e^{|z_n|} - e^{n \ln\left(1 + \frac{|z_n|}{n}\right)}$$

Or par Taylor reste intégral appliqué à $\ln$ entre 1 et $1+x$, $x \geq 0$,

$$\ln(1+x) = x - \frac{x^2}{2} + \underbrace{\int_1^{x+1} \frac{(x+1-t)^3}{6} \frac{2}{t^3}\, dt}_{\geq 0} \geq x - \frac{x^2}{2}$$

d'où $\left|e^{z_n} - \left(1 + \frac{z_n}{n}\right)^n\right| \leq e^{|z_n|} - e^{n\left(\frac{|z_n|}{n} - \frac{|z_n|^2}{2n^2}\right)} = e^{|z_n|}\left(1 - e^{-\frac{|z_n|^2}{2n}}\right)$

Or par Taylor reste intégral appliqué à exp (**) entre 0 et $x \geq 0$,

$$e^{-x} = 1 - x + \underbrace{\int_0^x \frac{(x-t)^2}{2} e^{-t}\, dt}_{\geq 0} \geq 1 - x \quad \text{d'où } 1 - e^{-x} \leq x \leftarrow \text{convexité de } e^{-\cdot}$$

d'où $\left|e^{z_n} - \left(1 + \frac{z_n}{n}\right)^n\right| \leq e^{|z_n|} \frac{|z_n|^2}{2n}$

Donc $\left|e^z - \left(1 + \frac{z_n}{n}\right)^n\right| \leq \left|e^z - e^{z_n}\right| + \left|e^{z_n} - \left(1 + \frac{z_n}{n}\right)^n\right| \leq \underbrace{\left|e^z - e^{z_n}\right|}_{\to 0} + \underbrace{\frac{1}{2n}}_{\to 0} \underbrace{|z_n|^2 e^{|z_n|}}_{\text{bornée car } (z_n) \text{ cv}} \underset{n\to+\infty}{\longrightarrow} 0$

car $z_n \to z$ et exp $\mathcal{C}^0$

Etape 4: Conclusion.

Donc $\varphi_{\frac{S_n}{\sqrt{n}}}(t) = \left(1 - \frac{t^2}{2n} + \frac{\varepsilon_n}{n}\right)^n \underset{n\to+\infty}{\longrightarrow} e^{-\frac{t^2}{2}}$ avec $z_n = -\frac{t^2}{2} + \varepsilon_n \underset{n\to+\infty}{\longrightarrow} -\frac{t^2}{2}$

pr tout $t \in \mathbb{R}$.

Ce qui conclut.

Application : Calcul d'intervalle de confiance asymptotique

On lance n fois une pièce de monnaie. On symbolise les résultats des lancers par une suite $(X_i)_{i\geq 1}$ de v.a. iid de loi $\mathcal{B}(p)$. On cherche à estimer le paramètre p.
(On prendra $n = 1000$ pour les applications numériques).

1ère méthode : intervalle de confiance par excès avec l'inégalité de Tchebychev

$$\forall \varepsilon > 0 \quad \mathbb{P}(|\overline{X_n} - p| \geq \varepsilon) \leq \frac{Var(\overline{X_n})}{\varepsilon^2} \leq \frac{1}{4n\varepsilon^2}$$

par étude de la fct° $x \mapsto x(1-x)$ + (X_i) iid

Donc si on veut un intervalle de confiance à 95 %, il faut prendre $\frac{1}{4n\varepsilon^2} \leq 0,05$

donc $\varepsilon \simeq 0,08$

Ainsi, $\mathbb{P}(p \in [\overline{X_n} \pm 0,08]) \geq 0,95$

2e méthode : intervalle de confiance asymptotique avec le TCL.

D'après le TCL, $\frac{\sqrt{n}(\overline{X_n} - p)}{\sqrt{p(1-p)}} \Rightarrow \mathcal{N}(0,1)$.

Or d'après la LGN faible, $\sqrt{\overline{X_n}(1-\overline{X_n})} \xrightarrow[n\to+\infty]{\mathbb{P}} \sqrt{p(1-p)}$

donc d'après le thm de Slutsky, $\frac{\sqrt{n}(\overline{X_n} - p)}{\sqrt{\overline{X_n}(1-\overline{X_n})}} \underset{n\to+\infty}{\Longrightarrow} \mathcal{N}(0,1)$

On note q_α le quantile d'ordre $1 - \frac{\alpha}{2}$ de la loi $\mathcal{N}(0,1)$: $\mathbb{P}(N \leq q_\alpha) = 1 - \frac{\alpha}{2}$ si $N \sim \mathcal{N}(0,1)$.

d'où $\mathbb{P}\left(-q \leq \frac{\sqrt{n}(\overline{X_n} - p)}{\sqrt{\overline{X_n}(1-\overline{X_n})}} \leq q\right) \xrightarrow[n\to+\infty]{} \mathbb{P}(N \leq q) - \mathbb{P}(N \leq -q) = 2\mathbb{P}(N \leq q) - 1 = 1 - \alpha$

Donc pour n assez grand,

$$\mathbb{P}\left(p \in \left[\overline{X_n} \pm \frac{q_\alpha}{\sqrt{n}} \sqrt{\overline{X_n}(1-\overline{X_n})}\right]\right) \simeq 1 - \alpha.$$

En pratique, on prend $\alpha = 0,05$ et donc on a $q_\alpha = 1,96$.

Ce qui donne : $\frac{q_\alpha}{\sqrt{n}} \sqrt{\overline{X_n}(1-\overline{X_n})} \simeq \frac{1,96}{32} \frac{1}{\sqrt{4}} \simeq 0,03$.

0,03 c'est mieux que 0,08 !

Le TCL est un thm clé en proba. Il souligne le rôle central des variables gaussiennes qui peuvent être vues comme le comportement global d'une multitude de petits phénomènes. Par ex les chocs de molécules d'eau sur une molécule de pollen ou les effets des conditions atmosphériques sur le plan de vol d'un avion peuvent ê modélisés par des gaussiennes.
En pratique, quand $n \geq 30$, le TCL fournit une bonne approximation de la situat° et on peut estimer son erreur grâce à l'inégalité de Berry-Esseen.