

Rappel : (X_n) va. i.i.d

$$\frac{S_n}{n} = \frac{1}{n} \sum_{i=1}^n X_i \xrightarrow[n \rightarrow \infty]{\text{P, p.s., } L^1} \mathbb{E}[X_1]. \quad (\text{LGN})$$

2. Théorème Limite central (TLC)

Le TLC est un raffinement de P_n LGN. (CLT en anglais)

2.1 Théorème Limite central

Théorème : Soit (X_n) une suite de va. i.i.d avec $\mathbb{E}[|X_1|^2] < \infty$. On pose $\sigma^2 = \text{var}(X_1)$. Alors

$$\sqrt{n} \left(\frac{S_n}{n} - \mathbb{E}[X_1] \right) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}(0, \sigma^2),$$

de façon équivalente

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

Remarque : • Le TLC mesure que la vitesse de convergence dans la P_n LGN est $O\left(\frac{1}{\sqrt{n}}\right)$.

- La loi gaussienne $\mathcal{N}(0, 1)$ apparaît comme une loi limite universelle

Montrons le lemme suivant :

Lemme : Pour tout $z \in \mathbb{R}$, $p \geq 1$

$$\left| e^{iz} - \sum_{k=0}^p \frac{(iz)^k}{k!} \right| \leq \min \left(\frac{|z|^{p+1}}{(p+1)!}, \frac{2|z|^p}{p!} \right).$$

Preuve : on applique la formule de Taylor avec reste intégral :

$$e^{iz} - \sum_{k=0}^p \frac{(iz)^k}{k!} = \frac{i^{p+1}}{p!} \int_0^z (z-s)^p e^{is} ds \quad \textcircled{A}$$

On a

$$\left| e^{iz} - \sum_{k=0}^p \frac{(iz)^k}{k!} \right| \leq \frac{1}{p!} \left| \int_0^z (z-s)^p ds \right| = \frac{|z|^{p+1}}{(p+1)!}$$

Par ailleurs, une intégration par parties donne

$$\begin{aligned} \left| e^{iz} - \sum_{k=0}^p \frac{(iz)^k}{k!} \right| &\leq \left| \frac{i^p}{(p+1)!} \int_0^z (z-s)^{p+1} e^{is} ds - \frac{(iz)^p}{p!} \right| \\ &\leq \frac{|z|^p}{p!} + \frac{|z|^p}{p!} = \frac{2|z|^p}{p!} \end{aligned}$$

□

Preuve du TLC

En considérant $\frac{X_i - E[X_i]}{\sigma}$, on peut supposer que $E[X_i] = 0$ et $\text{Var}(X_i) = 1$.

On calcule la fonction caractéristique, par indépendance ; pour $t \in \mathbb{R}$

$$\varphi_{\frac{S_n}{\sqrt{n}}}(t) = \varphi_{X_1}\left(\frac{t}{\sqrt{n}}\right)^n$$

Comme $E[X^2] < \infty$, φ_{X_1} est 2 fois dérivable

$$\begin{aligned} \text{et } \varphi_{X_1}(t) &= 1 + t \varphi'_{X_1}(0) + \frac{t^2}{2} \varphi''_{X_1}(0) + o(t^2) \\ &= 1 + 0 - \frac{t^2}{2} + o(t^2). \end{aligned}$$

D'après le lemme vu au cours précédent, comparant $|\prod_i a_i - \prod_i b_i|$ avec $a_i = \varphi_{X_1}\left(\frac{t}{\sqrt{n}}\right)$, $b_i = \left(1 - \frac{t^2}{2n}\right)$

on a

$$\rightarrow \left| \varphi_{\frac{S_n}{\sqrt{n}}}(t) - \left(1 - \frac{t^2}{2n}\right)^n \right| \leq n \left| \varphi_{X_1}\left(\frac{t}{\sqrt{n}}\right) - \left(1 - \frac{t^2}{2n}\right) \right|.$$

D'après le dernier lemme avec $p=2$

$$\left| \varphi_{X_1}\left(\frac{t}{\sqrt{n}}\right) - \left(1 - \frac{t^2}{2n}\right) \right| \leq E \left[\min \left(\frac{|t|^3 |X_1|^3}{6n^{3/2}}, \frac{t^2 |X_1|^2}{n} \right) \right].$$

Ainsi

$$\left| \varphi_{\frac{S_n}{\sqrt{n}}}(t) - \left(1 - \frac{t^2}{2n}\right)^n \right| \leq \mathbb{E} \left[\min \left(\frac{|t|^3 |X_1|^3}{6\sqrt{n}}, \underbrace{\mathbb{E}^2 |X_1|^2}_{\substack{\rightarrow 0 \\ n \text{ fixé}}} \right) \right]$$

intégrable
domine

$$\xrightarrow[n \rightarrow \infty]{\text{cr domine}} 0.$$

Par ailleurs, puisque $n \rightarrow \infty$, on a

$$\left(1 - \frac{t^2}{2n}\right)^n = \exp\left(n \log\left(1 - \frac{t^2}{2n}\right)\right) \xrightarrow[n \rightarrow \infty]{} e^{-\frac{t^2}{2}}.$$

Par conséquent

$$\left| \varphi_{\frac{S_n}{\sqrt{n}}}(t) - e^{-\frac{t^2}{2}} \right| \xrightarrow[n \rightarrow \infty]{} 0.$$

$$\text{i.e. } \left| \varphi_{\frac{S_n}{\sqrt{n}}}(t) - \varphi_{\text{N}(0,1)}(t) \right| \xrightarrow[n \rightarrow \infty]{} 0. \quad \square$$

Remarque: Le TLC s'écrit de manière équivalente:
 $\forall [a, b] \subset \mathbb{R}$

$$\mathbb{P} \left(\frac{\sqrt{n}}{\sigma} \left(\frac{S_n}{n} - \mathbb{E}[X_1] \right) \in [a, b] \right) \rightarrow \int_a^b e^{-\frac{x^2}{2}} \frac{dx}{\sqrt{2\pi}}.$$

Théorème : Soit (X_n) une suite de variables indépendantes avec $\text{var}(X_i) < \infty \quad \forall i \in \mathbb{N}^*$.

On pose $\sigma_n = \sqrt{\text{var}(X_1) + \dots + \text{var}(X_n)}$.

On suppose de plus la condition de Lindeberg :

$$\forall \varepsilon > 0, \quad \frac{1}{\sigma_n^2} \sum_{i=1}^n \mathbb{E} \left[|X_i - \mathbb{E}[X_i]|^2 \mathbb{1}_{|X_i - \mathbb{E}[X_i]| > \varepsilon \sigma_n} \right] \xrightarrow{n \rightarrow \infty} 0. \quad (*)$$

Alors
$$Z_n = \frac{\sum_{i=1}^n (X_i - \mathbb{E}[X_i])}{\sigma_n} \xrightarrow[n \rightarrow \infty]{} \mathcal{N}(0, 1).$$

Preuve : Comme plus haut par indépendance, $\forall t \in \mathbb{R}$

$$\phi_{Z_n}(t) = \prod_{k=1}^n \mathbb{E} \left[e^{i \frac{t}{\sigma_n} (X_k - \mathbb{E}[X_k])} \right]$$

$$\text{avec } \mathbb{E} \left[e^{i \frac{t}{\sigma_n} (X_k - \mathbb{E}[X_k])} \right] = 1 - \frac{t^2}{2\sigma_n^2} \text{var}(X_k) + o\left(\frac{t^2}{\sigma_n^2}\right).$$

D'après le lemme $(\prod a_i - \prod b_i) \dots$, on a alors

$$\left| \phi_{Z_n}(t) - \prod_{i=1}^n \left(1 - \frac{t^2}{2\sigma_n^2} \text{var}(X_i) \right) \right| \leq \dots \xrightarrow[n \rightarrow \infty]{} 0. \quad (*)$$

On remarque enfin que

$$\prod_{k=1}^n \left(1 - \frac{t^2}{2\sigma_n^2} \text{var}(X_k) \right) = \exp \left[\sum_{i=1}^n \log \left(1 - \frac{t^2}{2\sigma_n^2} \text{var}(X_i) \right) \right]$$

$$\xrightarrow[n \rightarrow \infty]{} \exp \left(-\frac{t^2}{2} \cdot 1 + o(1) \right).$$

□

2.2 Revoir sur les applications de la LGW.

Méthode de Monte-Carlo

On a vu que si X_i iid et $g: \mathbb{R} \rightarrow \mathbb{R}$ t.q.
 $g(X_i) \in \mathbb{L}^1$ alors

$$\frac{1}{n} \sum_{i=1}^n g(X_i) \xrightarrow[n \rightarrow \infty]{} \mathbb{E}[g(X_1)] = \int g(x) P_X(dx).$$

A l'aide du TLC, on peut préciser

$$\frac{\sqrt{n}}{\sqrt{\text{var}(g(X_1))}} \left(\frac{1}{n} \sum_{i=1}^n g(X_i) - \mathbb{E}[g(X_1)] \right) \xrightarrow{d} \mathcal{N}(0, 1).$$

Ainsi la vitesse de convergence de $\frac{1}{n} \sum_{i=1}^n g(X_i)$
vers $\mathbb{E}[g(X_1)]$ est de l'ordre de $O\left(\frac{1}{\sqrt{n}}\right)$.

Cette vitesse est indépendante de la régularité
de la fonction g , contrairement aux autres
méthodes d'intégration usuelles.

Estimation paramétrique

On reprend l'exemple du cours précédent $X_i \sim B(p)$
iid avec $X_i = 1$ si vote pour A
0 si vote pour B
 p inconnue.

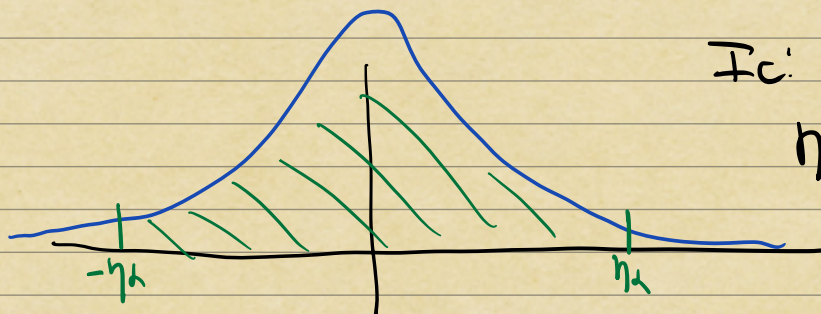
On a vu hier que p.m. LGV 1-Pique :

$$\frac{S_n}{n} = \frac{\sum_{i=1}^n X_i}{n} \xrightarrow{n \rightarrow \infty} E[X_1] = p.$$

Fixons un seuil de confiance $1-\alpha = 95\%$. par ex.

Ensuite, il existe η_α h.

$$IP(|\mathcal{N}(0,1)| \leq \eta_\alpha) \geq 1-\alpha$$



Ex: si $\alpha = 5\%$,

$$\eta_\alpha = 1.96.$$

D'après p.m. LGV de la TLC, on a

$$IP\left(\frac{\sqrt{n}}{\sigma} \left| \frac{S_n}{n} - p \right| \leq \eta_\alpha\right) \xrightarrow{n \rightarrow \infty} IP(|\mathcal{N}(0,1)| \leq \eta_\alpha) \geq 1-\alpha$$

||

$$IP\left(p \in \left[\frac{S_n}{n} - \frac{\eta_\alpha \cdot \sigma}{\sqrt{n}}, \frac{S_n}{n} + \frac{\eta_\alpha \cdot \sigma}{\sqrt{n}}\right]\right)$$

Comme $\sigma = p(1-p) \leq \frac{1}{4}$ et $1.96 \leq 2$, ainsi

$$IP\left(p \in \underbrace{\left[\frac{S_n}{n} - \frac{1}{2\sqrt{n}}, \frac{S_n}{n} + \frac{1}{2\sqrt{n}}\right]}_{\text{fourchette}}\right) \xrightarrow{n \rightarrow \infty} 1-\alpha = 95\%.$$

Autrement dit, avec une probabilité 95% de se
pas se tromper, on peut affirmer que

$$p \in \left[\frac{S_n}{n} - \frac{1}{2\sqrt{n}}, \frac{S_n}{n} + \frac{1}{2\sqrt{n}} \right].$$

La taille p_n fournie est $\frac{1}{\sqrt{n}}$.

Ainsi, si on interroge 10^4 personnes, on a une
précision de 10^{-2} .

2.3 Vitesse de convergence dans le TLC.

Théorème (Berry-Esseen)

Soit (X_i) une suite de v.a. i.i.d avec $E[|X_i|^3] < \infty$.

Alors

$$\sup_{x \in \mathbb{R}} \left| P\left(\frac{S_n}{\sigma} \left(\frac{S_n}{n} - E[X_i] \right) \leq x\right) - P(N(0,1) \leq x) \right| \leq \frac{C \cdot E[|X_i|^3]}{\text{var}(X_i) \sqrt{n}}.$$

où C est une constante universelle (≤ 3).

Preuve: voir DM.

Squellée: relier $P(X \leq x) \sim P_X(x)$

$$\text{via } P(X \leq x) = \int_{-\infty}^x P_X(du) = \int_{\mathbb{R}} \mathbb{1}_{(-\infty, x]}(u) P_X(du)$$

on approche $\mathbb{H}_{\infty, n}$ par une fonction régulière
on utilise ensuite la transformée de Fourier
inverse.