Chapitre 3 : Théorème centrale limite.

• Convergence en loi

<u>Définition</u>: $Y_1, Y_2 \dots$ à valeurs ds $\mathbb{R}$ cv en loi vers la v.a. Y à valeurs ds $\mathbb{R}$ si :

$$\lim_{n \to \infty} E[h(Y_n)] = E[h(Y)] \quad h \text{ continue, bornée.}$$

On note $Y_n \xrightarrow[n \to \infty]{\mathcal{L}} Y \Leftrightarrow \lim_{n \to \infty} F_{Y_n}(t) = F_Y(t)$

<u>Propriété</u>: Si Y_n cv en loi vers Y et si $f : \mathbb{R} \to \mathbb{R}$ continue
alors $f(Y_n) \xrightarrow[n \to \infty]{\mathcal{L}} f(Y)$

<u>ex</u> : $X_n \sim Unif(\{1, 2, \dots, n\})$ $\quad \frac{X_n}{n} \xrightarrow[n \to \infty]{\mathcal{L}oi} ?$

Pour h continue bornée

$$E\left(h\left(\frac{X_n}{n}\right)\right) = \sum_{k=1}^{n} h\left(\frac{k}{n}\right) P(X_n = k)$$

$$= \frac{1}{n} \sum_{k=1}^{n} h\left(\frac{k}{n}\right) \xrightarrow[n \to \infty]{\text{loi}} \int_0^1 h(x)dx$$

$$\quad = E[h(U)]$$

où $U \sim Unif(]0,1[)$

D'où $\frac{X_n}{n} \xrightarrow[n \to \infty]{\mathcal{L}oi} U \sim Unif(]0,1[)$

• Convergence en loi des moyennes empiriques

Prenons des v.a.r. $X_1, X_2, \dots$ iid dont la loi admet un moment d'ordre 2.

On sq $\underline{Var(X_1) > 0}$

La loi de $S_n = \sum X_i$ dépend de la loi des X_i.

$E(S_n) = n\,E(X_1)$

$V(S_n) = n\,Var(X_1)$

La loi de $S_n^* = \frac{S_n - n\,E(X_1)}{\sqrt{n\,V(X_1)}}$ dépend aussi de la loi des X_i mais

$$E(S_n^*) = \frac{1}{\sqrt{n\,V(X_1)}}$$

$$V(S_n^*) = \frac{Var(S_n)}{n\,Var(X_1)} = 1$$

<u>idée</u> : si n devient très grand, la loi S_n^* dépend de moins en moins de la loi des X_i, elle tend vers la $\mathcal{N}(0,1)$

Remarque : $S_n^* = \frac{\bar{X}_n - E(X_1)}{\sqrt{V(X_1)}} \sqrt{n} = \frac{\bar{X}_n - E(\bar{X}_n)}{\sqrt{V(\bar{X}_n)}} \sqrt{n}$

Théorème central limite :

Si les X_i sont des v.a.r.i.i.d. qui ont un moment d'ordre 2 et si $V(X_1) > 0$

alors $S_n^* = \dfrac{\sum X_i - nE(X_1)}{\sqrt{nV(X_1)}} = \sqrt{n}\,\dfrac{\overline{X_n} - E(X_1)}{\sqrt{V(X_1)}} \xrightarrow[n \to \infty]{Loi} Z \sim \mathcal{N}(0,1)$

Autrement dit pour n très grand on a :

$$P(a < S_n^* < b) = F_{S_n^*}(b) - F_{S_n^*}(a) \xrightarrow[n \to \infty]{Loi} \int_a^b \frac{e^{-x^2/2}}{\sqrt{2\pi}}\,dx$$

$$\phi(b) - \phi(a)$$

pour n grand : $P(a < S_n^* < b) \simeq \int_a^b \frac{e^{-x^2/2}}{\sqrt{2\pi}}\,dx$

L'application du TCL à une suite i.i.d. de $Ber(p)$ donne :

↳ Théorème de De Moivre-Laplace :

Si, $\forall p \in]0,1[$ fixé et $\forall n \in \mathbb{N}^*$ on a $S_n \sim Bin(n,p)$

alors $\boxed{\dfrac{S_n - np}{\sqrt{np(1-p)}} \xrightarrow[n \to \infty]{Loi} Z \sim \mathcal{N}(0,1)}$

Théor. cent. Lim. => Sous $V(X_1) > 0$ la vitesse de convergence ds la loi forte des grands nombre est en $1/\sqrt{n}$.

Interprétation : pour n très grand, loi loi de $\overline{X_n}$ est proche $\mathcal{N}(E(\overline{X_n}), V(\overline{X_n}))$

exemple : LES PAQUETS DE SUCRE.

Poids : 1 kg → poids réelle : v.a. $E(X_1) = 1071$

Palette : 1400 kg $\sigma = 6 \Rightarrow V(X_1) = 36$

Emballage : 71 g.

un dispositif de levage de 1500 kg suffit-il ?

Les X_i sont bornés (min = 0 kg, max = 5 kg)

↳ On utilise le corollaire de Tchebychev.

$$P(\sum X_i > 1500\,000) = P\left(\frac{1}{1400}\sum X_i - E(X_1) > 1500\,000 \times \frac{1}{1400} - 1071\right)$$

$$\leq P\left(\left|\frac{1}{1400}\sum X_i - 1071\right| \geq 0,4286\right)$$

$$\leq \frac{Var(X_1)}{1400 \times (0,4286)^2} = 0,1400$$

décourageant.

↳ On utilise le TCL

$$P(\sum X_i > 1\,500\,000) = P\left(\frac{\sum X_i - 1400 \times 1071}{\sqrt{1400 \times 6^2}} > \underbrace{\frac{1500\,000 - 1400 \times 1071}{\sqrt{1400 \times 6^2}}}_{\simeq 2,67}\right)$$

Comme $n = 1400$ grand, la fdR de $\dfrac{\sum X_i - 1400 \times n}{\sqrt{pV(X_1)}}$ est proche de ϕ

$$p \simeq P(Z > 2,67) = 1 - \underbrace{\phi(2,67)}_{= 0,9962} \qquad Z \sim \mathcal{N}(0,1)$$

D'où $p \simeq 0,38\%$.

Exemple de construction d'un IC à partir du TCL :

↳ LES SUCRES → le fabriquant trafique-t-il la machine ?

$\sigma = 6g$ → on veut estimer m

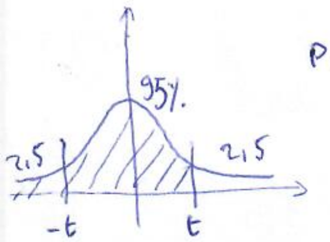
$X_1, \dots, X_{1400}$ iid

$E(X_1) = m + 71$

$V(X_1) = 36$

$n = 1400$ → TCL ⇒ $P\left(-t \leq \frac{\sum X_i - 1400(m+71)}{\sqrt{n\,Var(X_1)}} \leq t\right) \approx P(|Z| \leq t)$ $Z \sim \mathcal{N}(0,1)$

IC à 95% → On cherche t tq $P(|Z| \leq t) \simeq 0,95$



$P(Z \leq t) = 0,975$

↳ on trouve $t = 1,96$

dc $P\left(\left|\frac{\sum X_i - 1400(m+71)}{\sqrt{n\,Var(X_1)}}\right| \leq 1,96\right)$

⇔ $P\left(\left|\overline{X}_{1400} \pm 1,96\sqrt{\frac{36}{1400}}\right| \leq m\right) \simeq 0,95$

D'où $IC = \left[\overline{X}_{1400} \pm 1,96\sqrt{\frac{36}{n}}\right]$

Tchebychev vs TCL

Tchebychev ↓ Donne une majoration exacte que $\overline{X}_n$ s'écarte de son espérance.
Les IC sont en général plus large mais valables $\forall n \in \mathbb{N}^*$

TCL ↳ donne une valeur approchée que $\overline{X}_n$ s'écarte de l'espérance $E(X_1)$.
Les IC sont souvent plus étroits mais leur niveau de confiance est approximatif et l'approximat° n'est bonne que si n est grand.

· Théorème de Berry-Essen :

Si les X_i sont ind. et ont la m̂ loi qui a un moment d'ordre 2.
et si $\rho^3 = E(|X_1 - E(X_1)|^2) > 0$

alors $\forall n \in \mathbb{N}^*, t \in \mathbb{R}$

$$\left|P\left(\sqrt{n}\,\frac{\overline{X}_n - E(X_1)}{\sqrt{Var(X_1)}} \leq t\right) - P(Z \leq t)\right| \leq \frac{0,8}{\sqrt{n}} \times \frac{\rho^3}{\sqrt{V(X_1)}^3}$$

$Z \sim \mathcal{N}(0,1)$

$\frac{p^2 + (1-p)^2}{\sqrt{p(1-p)}}$

↳ contrôle de l'erreur dans le th. de De Moivre-Laplace.

Exemple: machine de remplissage de sucre tq

$$E(|X_i - E(X_1)|^3) \simeq 20\, g^3$$

Alors d'après le TCL:

$$0{,}95 \simeq \underbrace{P\left(-1{,}96 \leq \sqrt{1400}\ \frac{\overline{X_{1400}} - (m+71)}{\sqrt{36}} \leq 1{,}96\right)}_{\hookrightarrow\ F_{S^*}(1{,}96) - F_{S^*_{1400}}(-1{,}96)} \simeq P(|Z| \leq 1{,}96)$$

D'après Barry Essien,

$$\forall t \in \mathbb{R},\ |F_{S^*_{1400}}(t) - F_Z(t)| \leq \frac{0{,}8}{\sqrt{1400}} \cdot \frac{20}{6^3}$$

$$\left| (F_{S^*}(1{,}96) - F_{S^*}(-1{,}96)) - (F_Z(1{,}96) - F_Z(-1{,}96)) \right|$$

$$\leq \left| (F_{S^*}(1{,}96) - F_{S^*}(-1{,}96)) - F(S^*(1{,}96) - F_{S^*}(-1{,}96)) \right|$$

$$\leq \frac{0{,}8}{\sqrt{1400}} \cdot \frac{20}{6^3} + \frac{0{,}8}{\sqrt{1400}} \cdot \frac{20}{6^3} \simeq 2 \times 0{,}0198$$

↳ IC à 95% ± 0,4% ⩾ 94,6%

- Théorème centrale limite à variance inconnue

TCL avec autonormalisation.

Si $X_1, X_2, X_3 \ldots$ ind. et de $\hat{m}$ loi qui a $\sigma > 0$

alors $\sqrt{n}\ \dfrac{\overline{X_n} - E(X_1)}{\sqrt{V_n}} \xrightarrow[n\to\infty]{Loi} Z \sim \mathcal{N}(0,1)$

avec $\overline{X_n} = \frac{1}{n} \sum X_i$

$V_n = \left(\frac{1}{n} \sum X_i^2\right) - (\overline{X_n})^2$