# Exo 1 TD2 (fin)

$$\hat{\sigma}_n^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2$$

4) Normalité asymptotique de $\hat{\sigma}_n^2$.

↳ TCL

$$\sqrt{n}\left(\hat{\sigma}_n^2 - \sigma^2\right) \xrightarrow[n \to +\infty]{\mathcal{L}} \mathcal{N}(0, \ast)$$

↖ variance

$$\hat{\sigma}_n^2 = \underbrace{\frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)^2}_{\text{Terme 1}} - \underbrace{(\mu - \bar{X}_n)^2}_{\text{Terme 2}}$$

(par la question 3))

Terme 1 ↳ TCL

Terme 2 ↳ on veut prouver que ce terme tend vers 0.

Méthode 1 : on prouve que le terme 1 tend en loi vers $\mathcal{N}(-)$
on prouve que le terme 2 tend en loi vers 0

Méthode 2 : on prouve que le terme 1 tend en loi vers $\mathcal{N}(-)$
on prouve que le terme 2 tend en proba vers 0
+ on conclue par Slutsky.

Terme 1 $\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)^2\right)$

On va appliquer le TCL.

- $(X_i-\mu)^2$ i.i.d.
- $\mathbb{E}\left[(X_i-\mu)^2\right] = Var(X_i) = \sigma^2$
- $Var\left((X_i-\mu)^2\right) = \mathbb{E}\left[(X_i-\mu)^4\right] - \mathbb{E}\left[(X_i-\mu)^2\right]^2$

$$= \tau^4 - \sigma^4$$

On a donc par le TCL

$$\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)^2 - \sigma^2\right) \xrightarrow[n\to+\infty]{\mathcal{L}} \mathcal{N}(0, \tau^4-\sigma^4)$$

Méthode 1

Terme 2 → on va aussi étudier sa cv en loi

$$\sqrt{n}(\mu-\bar{X}_n)^2 = \underbrace{\sqrt{n}\,(\bar{X}_n-\mu)}_{a)}\ \underbrace{(\bar{X}_n-\mu)}_{b)}$$

- cv en loi
- cv en proba vers une constante

- a) On va appliquer le TCL
  - $(\bar{X}_n-\mu) = \left(\frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)\right)$
  - $(X_i-\mu)$ i.i.d. (car $X_i$ sont i.i.d.)
  - $\mathbb{E}[(X_i-\mu)] = 0$ car $\mathbb{E}[X_i]=\mu$
  - $Var(X_i-\mu) = Var(X_i) = \sigma^2 < +\infty$

$$\boxed{\begin{array}{l} Var(X+\underset{\text{réel}}{a}) = Var(X) \\ Var(aX) = a^2\,Var(X) \end{array}}$$

Par le TCL

$$\underbrace{\sqrt{n}\,((\bar{X}_n-\mu)-0)}_{=\sqrt{n}(\bar{X}_n-\mu)} \xrightarrow[n\to+\infty]{\mathcal{L}} \mathcal{N}(0,\sigma^2)$$

• b) $\bar{X}_n - \mu$ cv en proba ?

$\bar{X}_n \xrightarrow[n \to +\infty]{\mathbb{P}, p.s.} \mu$ par LFGN $\left( \begin{array}{l} X_i \text{ i.i.d} \\ \mathbb{E}[X_1] = \mu \end{array} \right)$

$\Rightarrow \bar{X}_n - \mu \xrightarrow[n \to +\infty]{\mathbb{P}} 0$

Par Slutsky

• a) $\xrightarrow[n \to +\infty]{\mathcal{L}} \mathcal{N}(0, \sigma^2)$

b) $\xrightarrow{\mathbb{P}} 0$

$\left.\right\}$ a) $\times$ b) $\xrightarrow[n \to +\infty]{\mathcal{L}} \underbrace{0 \, \mathcal{N}(0, \sigma^2)}_{= 0}$

$\Rightarrow \sqrt{n} \left( \hat{\sigma}_n^2 - \sigma^2 \right) \xrightarrow[n \to +\infty]{\mathcal{L}} \mathcal{N}(0, \tau^4 - \sigma^4)$

Autre méthode ? : Cas en pratique

Terme 2 $\longrightarrow \sqrt{n}\,(\bar{X}_n - \mu)^2$

Par l'inégalité de Markov

$$\mathbb{P}\Big(\underbrace{\sqrt{n}\,(\bar{X}_n - \mu)^2}_{(n^{1/4}(\bar{X}_n - \mu))^2} > \varepsilon\Big) \leq \frac{\mathbb{E}\left[\sqrt{n}\,(\bar{X}_n - \mu)^2\right]}{\varepsilon}$$

$$= \sqrt{n}\,\frac{\mathbb{E}\left[(\bar{X}_n - \mu)^2\right]}{\varepsilon}$$

$$= \frac{\sqrt{n}\,\mathrm{Var}(\bar{X}_n)}{\varepsilon}$$

$$= \frac{\sqrt{n}\,\sigma^2}{n\,\varepsilon} = \frac{\sigma^2}{\sqrt{n}\,\varepsilon}$$

$$\xrightarrow[n \to +\infty]{} 0$$

$$\left.\begin{array}{l} \text{Term } 1 \xrightarrow[n \to +\infty]{\mathcal{L}} \mathcal{N}(0, \tau^4 - \sigma^4) \\ \text{Term } 2 \xrightarrow[n \to +\infty]{\mathbb{P}} 0 \end{array}\right\}$$

Par Slutsky

$$\sqrt{n}\,(\hat{\sigma}_n - \sigma^2) \xrightarrow[n \to +\infty]{\mathcal{L}} \mathcal{N}(0, \tau^4 - \sigma^4)$$

J'ai utilisé Slutsky :

$$\begin{cases} X_n \xrightarrow{\mathcal{L}} X \\ Y_n \xrightarrow{\mathbb{P}} c \text{ constante} \end{cases}$$

$$\Rightarrow X_n + Y_n \xrightarrow{\mathcal{L}} X + c$$

( remarque : un estimateur est une v.a. ) $\rightarrow$ on veut sa loi, sa convergence asymp., son espérance, sa variance ...

5) $$\mathbb{E}\left[\hat{\sigma}_m^2\right] = \mathbb{E}\left[\frac{1}{m}\sum_{i=1}^{m}(X_i-\mu)^2 - (\overline{X}_m-\mu)^2\right]$$

question 3)

$\mathbb{E}$ linéaire

$$= \frac{1}{m}\sum_{i=1}^{m}\mathbb{E}\left[(X_i-\mu)^2\right] - \mathbb{E}\left[(\overline{X}_m-\mu)^2\right]$$

$X_i$ sont i.d.

$$= \frac{1}{m}\, m\, \mathbb{E}\left[(X_1-\mu)^2\right] - \mathbb{E}\left[(\overline{X}_m-\mu)^2\right]$$

$$= \mathrm{Var}(X_1) - \mathrm{Var}(\overline{X}_m)$$

$$= \sigma^2 - \frac{\sigma^2}{m}$$

$$= \frac{m-1}{m}\sigma^2 .$$

$\hat{\sigma}_m^2$ n'est pas sans biais.

6) · Posons $S_n^2 = \dfrac{n}{n-1}\hat{\sigma}_n^2$ (il est sans biais)

$$\mathbb{E}[S_n^2] = \mathbb{E}\left[\frac{n}{n-1}\hat{\sigma}_n^2\right] = \frac{n}{n-1}\mathbb{E}[\hat{\sigma}_n^2] = \sigma^2.$$

· Normalité asymptotique :

$$\sqrt{n}\,(S_n^2 - \sigma^2) \xrightarrow[n \to +\infty]{\mathcal{L}} \mathcal{N}(0, \ldots)$$

on connaît $\sqrt{n}\,(\hat{\sigma}_n^2 - \sigma^2)$

$$\sqrt{n}\,(S_n^2 - \sigma^2) = \sqrt{n}\left(\frac{n}{n-1}\hat{\sigma}_n^2 - \sigma^2\right)$$

$$= \sqrt{n}\left(\frac{n-1}{n-1}\underbrace{(\hat{\sigma}_n^2 - \sigma^2)}_{\text{Terme 1}} + \underbrace{\frac{1}{n-1}\hat{\sigma}_n^2}_{\text{Terme 2}}\right)$$

On a $\sqrt{n}\,(\hat{\sigma}_n^2 - \sigma^2) \xrightarrow[n \to +\infty]{\mathcal{L}} \mathcal{N}(0, \tau^4 - \sigma^4)$

$$\frac{\sqrt{n}}{n-1}\hat{\sigma}_n^2 \quad ?$$

CV en proba ?

- $\hat{\sigma}_n^2 \xrightarrow[n\to+\infty]{\mathbb{P}} \frac{n-1}{n}\sigma^2$
- $\frac{\sqrt{n}}{n-1} \xrightarrow[n\to+\infty]{} 0$

$$\Rightarrow \frac{\sqrt{n}}{n-1}\hat{\sigma}_n^2 \xrightarrow[n\to+\infty]{\mathbb{P}} \underbrace{\sigma^2 \times 0}_{=0}$$

~~LGN~~

$(X_i - \bar{X}_n)^2$ i.i.d

$\mathbb{E}[(X_i - \bar{X}_n)^2] = \sigma^2 = \mathbb{E}[\hat{\sigma}_n^2]$

$\neq \sigma^2$

$= \frac{n-1}{n}\sigma^2$

$$\hat{\sigma}_n^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2$$

LFGN $\frac{1}{n}\sum X_i \xrightarrow{p.s.} \mu$

$\mathbb{E}[X_1] = \mu$

$X_i$ i.i.d

$\mathbb{E}[X_1] = \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right]$

linéarité et $X_i$ i.i.d.

mieux expliqué slide suivant →

$$\frac{\sqrt{n}}{n-1} \hat{\sigma}_n^2 \quad \text{(Terme 2)}$$

① ce que l'on a fait en cours

LGN $\quad \hat{\sigma}_n^2 = \frac{1}{n} \sum_{i=1}^{n} (X_i - \bar{X}_n)^2 \quad$ • $(X_i - \bar{X}_n)^2$ i.i.d.

• $\mathbb{E}[(X_1 - \bar{X}_n)^2] = \mathbb{E}\left[\frac{1}{n} \sum_{i=1}^{n} (X_i - \bar{X}_n)^2\right]$

$\uparrow$

(Remarque de votre camarade
on utilise : $\mathbb{E}$ linéaire + $(X_i - \bar{X}_n)^2$ ont la même loi)

$$= \mathbb{E}[\hat{\sigma}_n^2] \quad \text{(question 5)}$$

$$= \frac{n-1}{n} \sigma^2$$

Si on applique la LGN, on a

$$\hat{\sigma}_n^2 \xrightarrow[n \to +\infty]{\mathbb{P}} \frac{n-1}{n} \sigma^2$$

même si $\frac{n-1}{n} \xrightarrow[n \to +\infty]{} 1$

ce n'est pas propre !

(2) $\frac{\sqrt{n}}{n-1}\hat{\sigma}_n^2 = \frac{n}{\sqrt{n}\,(n-1)}\hat{\sigma}_n^2$

. CV en proba de $\frac{n}{n-1}\hat{\sigma}_n^2$ ?

$\frac{n}{n-1}\hat{\sigma}_n^2 = \frac{n}{n-1}\frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2$

$= \frac{1}{n}\sum_{i=1}^{n}\frac{n}{n-1}(X_i - \overline{X}_n)^2$

. $\frac{n}{n-1}(X_i - \overline{X}_n)^2$ i.i.d

. $\mathbb{E}\left[\frac{n}{n-1}(X_i - \overline{X}_n)^2\right]$

$= \frac{n}{n-1}\mathbb{E}\left[(X_i - \overline{X}_n)^2\right] = \sigma^2$

Par la LGN, $\hat{\sigma}_n^2 \xrightarrow[n\to+\infty]{\mathbb{P}} \sigma^2$

. On a $\boxed{\frac{1}{\sqrt{n}} \xrightarrow[n\to+\infty]{} 0}$

. On en conclue $\frac{\sqrt{n}}{n-1}\hat{\sigma}_n^2 \xrightarrow[n\to+\infty]{} 0$

Par Slutsky,

$$\underbrace{\sqrt{n}\left(S_n^2-\sigma^2\right)}_{\downarrow} \xrightarrow[n\to+\infty]{\mathcal{L}} \mathcal{N}\left(0,\tau^4-\sigma^4\right)$$

$$= \underbrace{\text{Terme 1}}_{\xrightarrow{\mathcal{L}} \mathcal{N}(0,\tau^4-\sigma^4)} + \underbrace{\text{Terme 2}}_{\xrightarrow{\mathbb{P}} 0}$$

## Exo 2

$(X, Y)$ $\mu_X, \mu_Y$ ; $\sigma_X^2, \sigma_Y^2$.

Estimer la covariance $C = \mathbb{E}\left[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])\right]$

1) $C_n = \frac{1}{n} \sum_{i=1}^{n} (X_i - \overline{X}_n)(Y_i - \overline{Y}_n)$

Estimateur naturel aurait été $\frac{1}{n} \sum_{i=1}^{n} (X_i - \mu_X)(Y_i - \mu_Y)$

mais $\mu_X, \mu_Y$ inconnus.

On les remplace par leurs estimateurs.

2) $$\sum_{j=1}^{n} (x_j - \bar{X}_m)(y_j - \bar{Y}_m)$$

$$= \sum_{j=1}^{n} \big((x_j - \mu_x) + (\mu_x - \bar{X}_m)\big)\big((Y_j - \mu_y) + (\mu_y - \bar{Y}_m)\big)$$

$$= \boxed{\sum_{j=1}^{n} (x_j - \mu_x)(y_j - \mu_y)} + (\mu_y - \bar{Y}_m)\underbrace{\sum_{j=1}^{n} (x_j - \mu_x)}_{\propto \bar{X}_m} + (\mu_x - \bar{X}_m)\underbrace{\sum_{j=1}^{n} (y_j - \mu_y)}_{\propto \bar{Y}_m} + m(\mu_x - \underset{\bar{X}_m}{\bar{X}_m})(\mu_y - \underset{\bar{Y}_m}{\bar{Y}_m})$$

$$= \quad + m(\mu_y - \bar{Y}_m)(\bar{X}_m - \mu_x) + m(\mu_x - \bar{X}_m)(\bar{Y}_m - \mu_y) + m(\mu_x - \bar{X}_m)(\mu_y - \bar{Y}_m)$$

On a utilisé :

$$\sum (x_j - \mu_x) = m\left(\frac{1}{m}\sum_{j=1}^{n} x_j - \mu_x\right) = m(\bar{X}_m - \mu_x)$$

$$= \sum x_j - m\mu_x$$

$$= \quad -2m(\mu_y - \bar{Y}_m)(\mu_x - \bar{X}_m) + m(\mu_y - \bar{Y}_m)(\mu_x - \bar{X}_m) = \text{ce que l'on voulait.}$$

3) $n^2(\bar{X}_n - \mu_X)(\bar{Y}_n - \mu_Y) =$

$$= \sum_{j=1}^{n} (X_j - \mu_X)(Y_j - \mu_Y) + \sum_{i \neq j} (X_i - \mu_X)(Y_j - \mu_Y) \quad ?$$

$$\cancel{n^2}\left(\frac{1}{\cancel{n}} \sum_{i=1}^{n} X_i \overset{n\,\checkmark}{-} \mu_X\right)\left(\frac{1}{\cancel{n}} \sum_{i=1}^{n} Y_i \overset{n\,\checkmark}{-} \mu_Y\right)$$

détaile en (*)

$X_1 - \mu_X$ $\quad$ $X_2 - \mu_X$

$(a+b)(c+d)$

$Y_1 - \mu_Y$ $\quad$ $Y_2 - \mu_Y$

$$= \left(\sum_{i=1}^{n} (X_i - \mu_X)\right)\left(\sum_{i=1}^{n} (Y_i - \mu_Y)\right)$$

$$= \sum_{i=1}^{n} \Big[ \underbrace{(X_i - \mu_X)(Y_i - \mu_Y)}_{ac + bd} + \underbrace{\sum_{i \neq j} (X_i - \mu_X)(Y_j - \mu_Y)}_{bc + ad} \Big]$$

« termes croisés en indice »

(*) $n^2(\bar{X}_n - \mu_X)(\bar{Y}_n - \mu_Y)$

$$= n\left(\frac{1}{n}\sum_{i=1}^{n} X_i - \mu_X\right) n\left(\frac{1}{n}\sum_{i=1}^{n} Y_i - \mu_Y\right) = \left(\sum_{i=1}^{n} X_i - n\mu_X\right)\left(\sum_{i=1}^{n} Y_i - n\mu_Y\right)$$

4) $\mathbb{E}[C_m] \propto f(C)$ } question 2)

$$= \mathbb{E}\left[\frac{1}{m}\sum_{i=1}^{m}(X_i-\mu_x)(Y_i-\mu_Y) - (\bar{X}_m-\mu_x)(\bar{Y}_m-\mu_Y)\right]$$

$$= \frac{1}{m}\sum_{i=1}^{m}\overbrace{\mathbb{E}\left[(X_i-\mu_x)(Y_i-\mu_Y)\right]}^{=C} - \mathbb{E}\left[(\bar{X}_m-\mu_x)(\bar{Y}_m-\mu_Y)\right]$$

$$\neq \mathbb{E}\left[(\bar{X}_m-\mu_X)(\bar{Y}_m-\mu_Y)\right]$$

PIEGE

$$= \mathbb{E}\left[(X_i-\mu_x)\right]\mathbb{E}\left[(Y_i-\mu_Y)\right] \;?$$ on l'a si $X_i \perp\!\!\!\perp Y_i$

NON !

- on ne nous le dit pas dans l'énoncé
- on veut estimer la covariance.

$\Rightarrow (X_i, Y_i) \perp\!\!\!\perp (X_j, Y_j)$ $\underline{\underline{i \neq j}}$

$(X_1, Y_1) \ldots (X_m, Y_m)$ indépendants $\not\Rightarrow$ $X_1 \perp\!\!\!\perp Y_1, X_2 \perp\!\!\!\perp Y_2 \ldots$

$X_m \perp\!\!\!\perp Y_m$

$$\mathbb{E}[C_n] = \frac{1}{n}\sum_{i=1}^{n} C \quad - \frac{1}{n^2}\mathbb{E}\left[n^2(\overline{X}_n - \mu_X)(\overline{Y}_n - \mu_Y)\right]$$

Ça ressemble à la Q3)

$$\mathbb{E}\left[n^2(\overline{X}_n - \mu_X)(\overline{Y}_n - \mu_Y)\right] = \mathbb{E}\left[\sum_{i=1}^{n} (X_i - \mu_X)(Y_i - \mu_Y) + \sum_{j \neq i} (X_j - \mu_X)(Y_i - \mu_Y)\right]$$

$$= \sum_{i=1}^{n} \overbrace{\mathbb{E}\left[(X_i - \mu_X)(Y_i - \mu_Y)\right]}^{= C} + \underbrace{\sum_{j \neq i} \mathbb{E}\left[(X_j - \mu_X)(Y_i - \mu_Y)\right]}$$

$$= \sum_{j \neq i} \mathrm{Cov}(X_j, Y_i)$$

$$= 0$$

$$= n\, C$$

car $(X_j, Y_j) \perp\!\!\!\perp (X_i, Y_i)$ $\boxed{i \neq j}$

$\Leftarrow$ $X_j \perp\!\!\!\perp X_i$, $Y_j \perp\!\!\!\perp Y_i$, $Y_j \perp\!\!\!\perp X_i$, $X_j \perp\!\!\!\perp Y_i$

$$\mathbb{E}[C_n] = C - \frac{1}{n} C = \frac{n-1}{n} C.$$

5) Estimateur sans biais de $C$ ?

$$\hat{C_n} = \frac{n}{n-1} C_n.$$

6 et 7)