- Estimation statistique -

Ex1:

Rappel: (cours échantillonnage pdf):

↦ La loi des grandes nombres (LGN):

$$(\overline{X}_n) \xrightarrow{P_n} E(X) = m.$$

on peut estimer m en utilisant comme estimateur $\overline{X}$.

↦ $E(\overline{X}) = m \Rightarrow$ sans biais.

↦

$$(\overline{X}) \xrightarrow{P_n} m$$

$$V(\overline{X}) = \frac{V(X)}{n} = \frac{\sigma^2}{n} = \frac{(250)^2}{5} = 12500$$

cette estimateur et les observations amènent à l'estimation $\hat{m} = \frac{1}{n} \sum x_i =$

$$\frac{(627 + 878 + 925 + 1105 + 1213)}{5} = 950.$$

2/: $X \sim N(m, 250) \Rightarrow T = \frac{X - 950}{250} \sim N(0,1)$

↦

$$T = \frac{X - 950}{250} \sim N(0,1)$$

$P(-t \leq T \leq t) = 1 - \alpha = 0{,}95 \Rightarrow t = 1{,}96$. (d'après tableau. P. 43).

~~ESP~~/X

$$\Rightarrow -1{,}96 \leq T \leq 1{,}96$$

$$-(1{,}96 \times 250) \times 950 \leq X \leq (1{,}96 \times 250) \times 950$$

$$\Rightarrow 460 \leq X \leq 1440.$$

$\Rightarrow X \in [460, 1440] \Rightarrow$ Les touristes chinois. bondant la France

dans (deux .ex) = 1 dans = 667 ⇒ non. L'article du journal "Le monde n'est pas précis .

· Exercice 2:

1/ S_n^2 est un estimateur biaisé pour σ^2:

$$S_n^2 = \frac{1}{n} \sum_{i=0}^{n} (X_i - \bar{X})^2 = \frac{1}{n} \sum_{i=0}^{n} (X_i^2 - \bar{X}^2)$$

sa converge en proba vers σ^2 .

(LGN et slutsky) ⟹ Démonstration Ds le Cours échantionnage .

2/: $S_n'^2$ est un estimateur sans biais pour σ^2

3/ en utulisant le 1er estimateur S_n^2 ⟹

$$\sigma^2 = \frac{1}{5} \left((627 - 950)^2 + (878 - 950)^2 + (925 - 950)^2 + (1105 - 950)^2 + (1213 - 950)^2\right) = 40666 \Rightarrow \sigma = 202$$

estimation pour σ^2 en utulisant le second estimation $S_n'^2$

$$\sigma^2 = \frac{1}{n-1} = \frac{5}{4} \cdot 40666 = 50833 \Rightarrow \sigma = 225.$$

4/: donc S_n' est le meilleur estimateur de σ. ⟹ $S_n' = \sqrt{\frac{1}{1-n} \cdot \sum_{i=0}^{n} (x_i - \bar{x})^2}$

Ex 3 : (1)

- **Le modèle d'inférence :**

⇒ Population : Ado, agés de 14/15 ans, et scolarisés en classe de 3ième de la région "R".

. Caractère étudié : surpoids/ou non. on note p la prévalence de surpoids p est un paramètre de valeur inconnue. $0<p<1$.

- échantillon prélevé : taille 145 supposé tiré au hasard avec un taux de sondage <10%

- La variable du modèle : Z définie par $\begin{cases} Z=1 \text{ si surpoids;} \\ Z=0 \text{ sinon;} \end{cases}$

on a évidemment $P(Z=1)=p$.

$$F=\frac{\sum_{i=1}^{n} Z_i}{n}$$

2/ : On peut proposer pour comme estimateur pour p, la fréquence empirique. Cet estimateur est sans biais et convergent.

. Cet estimateur et les observations rapportés amènent l'estimation $\hat{p}=\frac{27}{145}=18,6\%$.

3/ : à revoir dans l'exercice 19.

Ex 4 :

- T_n est un estimateur sans biais : $E(T_n)=\theta$.
- $\lim\limits_{n\to+\infty} V(T_n)=0$.

on veut montrer que $\lim\limits_{n\to+\infty} E((T_n-\theta)^2)=0$

. $V(T_n)=V(T_n-\theta)=E((T_n-\theta)^2)-E((T_n-\theta))^2$.

. or : $E(T_n-\theta)=E(T_n)-E(\theta)=\theta-\theta=0$.

$$\lim_{n\to+\infty} V(T_n)=\lim_{n\to+\infty} E((T_n-\theta)^2)=0.$$

2/ : $\lim\limits_{n\to+\infty} V(T_n)=0$

$\lim\limits_{n\to+\infty} b_n=0 \quad (b_n=E(T_n)-\theta)$.

(1)

• on veut montrer que la $\lim_{n\to+\infty} E((T_n-\theta)^2)=0$.

$$E((T_n-\theta)^2) = E(T_n^2 - 2T_n\theta + \theta^2) = E(T_n^2) - 2E(T_n\theta) + \theta^2.$$

$$E((T_n-\theta)^2) = E(T_n^2) - 2\theta E(T_n) + \theta^2.$$

$$= E(T_n^2) - 2E^2(T_n) + E^2(T_n) -$$

$$2\theta E(T_n) + \theta^2.$$

$$= E((T_n - E(T_n))^2) + (E(T_n) - \theta)^2.$$

$$= V(T_n) + b_n^2$$

$$\lim_{n\to+\infty} E((T_n-\theta)^2) = \lim_{n\to+\infty} (V(T_n) + b_n^2) = \lim_{n\to+\infty} V(T_n) +$$

$$\lim_{n\to+\infty} b_n^2 = 0$$

Donc T_n converge en moyenne quadratique vers θ.

• Exercice 5:

1/: a ?.

• $0 < 1-2a < 1$, $\Rightarrow$ $0 < a < \frac{1}{2}$.

$$E(x) = \sum_{i=1}^{n} x_i \cdot p_i.$$

$$= -a + 0\times(1-2a) + 1\times a = 0.$$

$$V(x) = \sum x_i^2 p_i - [E(x)]^2 = 2a$$

$$P(X<0) = P(X=-1) = a$$

2/: on pose $\Pi = \frac{1}{n}\sum_{i=1}^{n} x_i^2$.

on cherche α / $T = \alpha\,\Pi$ ...

$+/: E(T)^{(2)} = \alpha \cdot E(\Pi) = \frac{\alpha}{m} \sum_{i=1}^{m} E(x_i^2) = \frac{\alpha}{m}, m \cdot 2a = 2\alpha a.$
$= a.$

$$\Rightarrow \boxed{\alpha = \frac{1}{2}}.$$

$$\boxed{T = \frac{1}{2} \Pi}$$

~~2/: +/ Posons $Z_i = 1$ si $X_i < 0$.~~

~~$Z_i = 0$ sinon.~~

Montrons que T est C.V

Q: • $V(T) = \frac{1}{4} V(\Pi) = \frac{1}{4m^2} \sum_{i=1}^{m} V(x_i^2)$.

$$V(x_i^2) = E(x_i^4) - [E(x_i^2)]^2.$$

$$= (a+a) - (2a)^2 = (2a - 4a^2)$$

$$\Rightarrow V(T) = \frac{1}{4m^2} m(2a - 4a^2) = \frac{2}{2m}(1-2a) \xrightarrow[m \to +\infty]{} 0$$

$\Rightarrow$ T cv. en moyenne quadratique vers a, par conséquent T cv en Pr. vers a.

3/: $Z_i = 1$ si $X_i < 0$
$Z_i = 0$ sinon

on Pose $Y = \sum_{i=1}^{m} Z_i$.

$$. E(Y) = m \cdot a, \quad V(Y) = m \cdot a(1-a).$$

donc Y suit une $\sim \beta(m, a)$.

./ $W = \frac{Y}{m}$, estimateur sans biais?

Comme $E(W) = a$, alors W est un estimateur sans biais.

la convergence de $w = \frac{Y}{m}$.

En utilisant la L.G.N on a : $w \xrightarrow{P_n} a$. (P26).

4/ :

Pour l'estimateur $T = \frac{1}{2} \Pi$.

Nos observations amènent. $\hat{a} = \frac{1}{2} \left(\frac{37}{150} + \frac{45}{150} \right)$

$\Rightarrow \hat{a} = 0{,}27$.

Pour l'estimateur $w = \frac{Y}{m}$

les observations amènent $\hat{a} =$

$V(w) = \frac{a(1-a)}{m}$; $V(T) = \frac{a}{2m}(1-2a)$

$V(w) - V(T) = \frac{2a(1-a)}{2m} - \frac{a}{2m}(1-2a)$

$= \frac{2a - 2a^2 - a + 2a^2}{2m} = \frac{a}{2m} > 0$.

$\Rightarrow V(w) > V(T)$,

donc on préfère T comme estimateur pour "a".

Ex6 :

1/ :

$\Pi = \frac{F_1 + F_2}{2}$; $S = aF_1 + a(1-a)F_2$

$E(\Pi) = \frac{1}{2}\left(E(F_1) + E(F_2)\right) = \frac{1}{2}(P + P) = P$.

Donc Π est un estimateur sans biais pour p :

$E(S) = a E(F_1) + (1-a) E(F_2) = p \cdot a - (1-a) p$

$= p \Rightarrow$ Donc S est un estimateur sans biais

2/ :

$a = ?$; S soit de variance

$$S = aF_1 + (1-a)F_2$$

$$\Rightarrow V(S) = a^2 V(F_1) + (1-a)^2$$

$$= a^2 \cdot \frac{p(1-p)}{n_1} + \frac{(1-a)^2 \cdot p \cdot (1-p)}{n'_2}$$

$$\Rightarrow \quad = p(1-p)\left(\frac{a^2}{n_1} + \frac{(1-a)^2}{n_2}\right)$$

on prend $f(a) = \frac{a^2}{n_1} + \frac{(1-a)^2}{n_1}$

f est dérivable sur $]0,1[$ et $\forall a \in]0,1[$, on a

$$f'(a) = \frac{2a}{n_1} - \frac{(2(1-a))}{n_2} = 2\left(\frac{a}{n_1} - \frac{(1-a)}{n_2}\right)$$

- donc : $f'(a) = 0 \Rightarrow \frac{a}{n_1} - \frac{(1-a)}{n_2} = 0$.

$$\Rightarrow a n_2 = n_1(1-a)$$

$$\Rightarrow a(n_1 + n_2) = n_1$$

$$\Rightarrow a = \frac{n_1}{n_1+n_2}.$$

- Le tableau de variation :

a	0		$\frac{n_1}{n_1+n_2}$		1
$f'(a)$					
$f(a)$	$\frac{1}{n_2}$	↘	$\frac{1}{n_1+n_2}$	↗	$\frac{1}{n_1}$

$$f\left(\frac{n_1}{n_1+n_2}\right) = \frac{\frac{n_1^2}{(n_1+n_2)^2}}{n_1} + \frac{\left(\frac{n_2}{n_1+n_2}\right)^2}{n_2} = \frac{n_1}{(n_1+n_2)^2} + \frac{n_2}{(n_1+n_2)^2}$$

$$= \frac{1}{\cdots} \cdot (3)$$

montrer que l'éstimateur S' est

$$V(S) = p\cdot(1-p)\left(\frac{a^2}{m_1}+\frac{(1-a)^2}{m_2}\right)$$

$$V(\Pi) = \frac{1}{4}\left(\frac{p\cdot(1-p)}{m_1}+\frac{p(1-p)}{m_2}\right) = \frac{p(1-p)}{4}\left(\frac{1}{m_1}+\frac{1}{m_2}\right)$$

$$\frac{a^2}{m_1}+\frac{(1-a)^2}{m_2} \leq \frac{1}{4}\left(\frac{1}{m_1}+\frac{1}{m_2}\right)$$

$$\frac{1}{m_1+m_2} \leq \left(\frac{1}{m_1}+\frac{1}{m_2}\right)\cdot\frac{1}{4},$$

$$\frac{4}{m_1+m_2} \leq \frac{1}{m_1}+\frac{1}{m_2}.$$

$$\frac{4\,m_1\cdot m_2}{m_1+m_2} \leq m_1+m_2.$$

$$4\,m_1\cdot m_2 \leq (m_1+m_2)^2.$$

$$0 \leq (m_1-m_2)^2$$

raisonment ds ce sens

$\Rightarrow$ S est le meilleur estimateur.

a):

$$\Pi = \frac{F_1+F_2}{2} \Rightarrow S = aF_1+(1-a)F_2.$$

$$\Pi = S \Leftrightarrow m_1 = m_2 \;?$$

"$\Rightarrow$":

$$\Pi = S \Rightarrow \frac{f_1}{2}+\frac{f_2}{2} = af_1+(1-a)F_2.$$

$$\Rightarrow a = \frac{1}{2} \Rightarrow \frac{1}{2} = \frac{m_1}{m_1+m_2} \Rightarrow 2m_1 = m_1+m_2.$$

$\Rightarrow \boxed{m_1 = m_2} \Rightarrow$ 1er sens est vérifiée.

"$\Leftarrow$": $m_1 = m_2 \Rightarrow \boxed{a = \frac{1}{2}} \Rightarrow S = \Pi \Rightarrow$ Donc Deuxième sens est vérifiée

Exercice 9 :

$$P(X=k) = \lambda^k \cdot \frac{e^{-\lambda}}{k!}$$

1/ un estimateur w par la méthode MV pour λ :

$$L(x_1, x_2, \ldots, x_n, \lambda) = \prod_{i=1}^{n} P_\lambda (X = x_i) .$$

$$= \lambda^{\sum_{i=1}^{n} x_i} \cdot \frac{e^{-n\lambda}}{\prod_{i=1}^{n} x_i !} = e^{-n\lambda} \frac{\lambda^{\sum_{i=1}^{n} x_i}}{\prod_{i=1}^{n} x_i !} = \frac{e^{-n\lambda} \cdot \lambda^{n\bar{x}}}{\prod_{i=1}^{n} x_i !}$$

$\Rightarrow$ on cherche à maximiser la fonction qui a $\lambda \longrightarrow L(x_1, x_2, \ldots, x_n, \lambda)$.

$\Rightarrow$: ou ce qui est ~~tout à~~ equivalent

$$\ln (L(x_1, x_2, \ldots, x_n, \lambda)) = \ln \left(\frac{e^{-n\lambda} \lambda^{\sum_{i=1}^{n} x_i}}{\prod_{i=1}^{n} x_i} \right)$$

$$= (-n\lambda + n\bar{x} \ln(\lambda)) - \ln \left(\prod_{i=1}^{n} x_i !\right)$$

• condition nécessaire on a :

$$\frac{\partial \ln(L)}{\partial \lambda} = 0 \Rightarrow -n + \frac{n\bar{x}}{\lambda} = 0 . \Rightarrow \frac{n\bar{x}}{\lambda} = n \Rightarrow \lambda = \bar{x} .$$

• condition suffisante :

$$\frac{\partial^2 \ln(L)}{\partial \lambda^2} = -\frac{n\bar{x}}{\lambda^2} < 0 . \Rightarrow \text{elle est vérifiée pour } \forall \lambda .$$

$\Rightarrow$ Donc l'estimateur du maximum (MV) est la statistique

$w = \bar{X}$.

2/ L'estimateur du MV est toujours asymptotiquement sans biais et converge en probabilité, de plus $W = \overline{X}$.

$$\boxed{E(\overline{X}) = E(X) = \lambda} \Rightarrow \text{donc } W \text{ est sans biais.}$$

W est un estimateur "naturel" pour $E(X)$ (LGN) $\overline{X} \xrightarrow{PR} E(X) = \lambda$).

3/ Une estimation pour λ fournie par l'estimateur W et les observations est $\hat{\lambda} = \frac{16 x_1 + 15 x_2 + 5 x_3 + 4 x_4}{52}$

$$\Rightarrow \hat{\lambda} \simeq 1{,}5$$

$$P(X=0) = \frac{\hat{\lambda}^0 e^{-\hat{\lambda}}}{0!} = e^{-\hat{\lambda}} = e^{-1{,}5} = 0{,}22$$

$$\boxed{P(X=0) = 0{,}22.}$$

La fréquence observée pour l'évènement $(X=0)$ est $\frac{12}{52} = 0{,}23$.

⚠ ⇒ on remarque que notre estimateur est précis.

· Ex 9 :

$$f(x) = \begin{cases} f \text{ est une densité de probabilité ssi :} \\ \cdot f \text{ est positive} \\ \cdot f \subseteq p.p. \end{cases}$$

$$\int_{\mathbb{R}} f(x) \cdot dx = 1.$$

$$\int_{\mathbb{R}} f(x) \cdot dx = \int_{-1}^{0} a \cdot dx + \int_{0}^{1} (1-a) \cdot dx = [ax]_{-1}^{0} + [(1-a)x]_{0}^{1}$$

$$= a + (1-a) = 1.$$

$\vdash$ f est une densité de probabilité.

$$E(x) = \int_{\mathbb{R}} x \cdot f(x) \cdot dx = \int_{-1}^{0} ax \cdot dx + \int_{0}^{1} (1-a) \cdot x \cdot dx.$$

$$= \left[\frac{ax^2}{2}\right]_{-1}^{0} + \left[\frac{(1-a)}{2} x^2\right]_{0}^{1} = -\frac{a}{2} + \frac{(1-a)}{2} = \frac{1}{2} - a.$$

$\cdot\ E(x) = \frac{1}{2} - a.$

$\cdot\ V(x) = E(x^2) - [E(x)]^2.$

$$\cdot\ E(x^2) = \int_{\mathbb{R}} x^2 \cdot f(x)\, dx = \int_{-1}^{0} x^2 \cdot a\, dx + \int_{0}^{1} x^2 (1-a) \cdot dx.$$

$$= \left[\frac{x^3 a}{3}\right]_{-9}^{0} + \left[\frac{(1-a)}{3} x^3\right]_{0}^{1} =$$

$$\frac{a}{3} + \frac{(1-a)}{3} = \frac{1}{3}.$$

$$\Rightarrow V(x) = \frac{1}{3} - \left(\frac{1}{2} - a\right)^2.$$

$P(X \leq 0) = F(0)$

si $x \in [-1, 0]$

$$F(x) = \int_{-1}^{x} a \cdot dt = [at]_{-1}^{x} = ax + a = a(x+1)$$

$\Rightarrow P(x<0) = f(0).$