Corrigé TD n°3 ①

EX1 chaque v.a X_i suit une loi normale $N(m_i+a, \sigma^2)$
$1 \leq i \leq n$

1)

Posons $Y_i = X_i - m_i \Rightarrow Y_i \hookrightarrow N(a, \sigma^2)$

$y_i = x_i - m_i$

La fonction de vraissemblance est :

$$L(x_1, .., x_n, a) = \prod_{i=1}^{n} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x_i - m_i - a)^2}{2\sigma^2}}$$

$$L(y_1, .., y_n, a, \sigma^2) = \prod_{i=1}^{n} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(y_i - a)^2}{2\sigma^2}}$$

$$= \frac{1}{\sigma^n(\sqrt{2\pi})^n} e^{-\frac{\sum_{i=1}^{n}(x_i - m_i - a)^2}{2\sigma^2}}$$

$$\Rightarrow \ln L(x_1, .., x_n, a) = -\frac{n}{2}\ln(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - m_i - a)^2$$

d'où $\frac{\partial \ln L}{\partial a} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i - m_i - a)$ et $\frac{\partial^2 \ln L}{\partial a^2} = -\frac{n}{\sigma^2}$

d'où $\frac{\partial \ln L}{\partial a} = 0 \Rightarrow \hat{a} = \frac{1}{n}\sum_{i=1}^{n}(x_i - m_i)$

donc $\hat{a}_n = \overline{Y}_n = \frac{1}{n}\sum_{i=1}^{n} Y_i = \frac{1}{n}\sum_{i=1}^{n}(X_i - m_i)$ est un estimateur

de maximum de vraissemblance du paramètre a et il est

efficace car $V(\hat{a}_n) = \frac{\sigma^2}{n}$ car les v.a. Y_i sont indep.

et $I_n(a) = -E\left(\frac{\partial^2 \ln L}{\partial a^2}\right) = \frac{n}{\sigma^2}$

on a bien $V(\hat{a}_n) = \frac{1}{I_n(a)}$

construction d'un intervalle de confiance :

$\hat{a}_n \hookrightarrow N\left(a, \left(\frac{\sigma}{\sqrt{n}}\right)^2\right)$

on cherche un intervalle de confiance de la forme

$[\hat{a}_n - \varepsilon, \hat{a}_n + \varepsilon]$ il faut donc determiner ε tq

$\hat{a}_n - \varepsilon \leq a \leq \hat{a}_n + \varepsilon$

$P(-\varepsilon \leq \hat{a}_n - a \leq \varepsilon) = 1 - \alpha$ ou bien

$$P\left(-\frac{\sqrt{n}\,\varepsilon}{\sigma} \leq \sqrt{n}\,\frac{\hat{a}_n - a}{\sigma} \leq \frac{\sqrt{n}}{\sigma}\varepsilon\right) = 1 - \alpha \quad (*)$$

or $\sqrt{n}\,\dfrac{\hat{a}_n - a}{\sigma} \hookrightarrow N(0,1)$ soit ϕ la fonction de repartition de cette loi alors $(*)$ donne

$$2\phi\left(\frac{\sqrt{n}\,\varepsilon}{\sigma}\right) - 1 = 1 - \alpha \Rightarrow \frac{\sqrt{n}}{\sigma}\varepsilon = \phi^{-1}\left(1 - \frac{\alpha}{2}\right)$$

si $\alpha = 0{,}05$ alors alors $u_\alpha = \phi^{-1}\left(1 - \frac{\alpha}{2}\right) = 1{,}96$

d'où $\varepsilon = u_\alpha \dfrac{\sigma}{\sqrt{n}}$

$f(x) = \dfrac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$

l'intervalle de confiance pour a est donc

$$\left[\hat{a}_n - u_\alpha \frac{\sigma}{\sqrt{n}}, \hat{a}_n + u_\alpha \frac{\sigma}{\sqrt{n}}\right].$$

2°) cette fois les X_i suivant la loi $N(m_i b, \sigma^2)$

posons pour $1 \leq i \leq n$ $Z_i = \dfrac{X_i}{m_i}$ alors $Z_i \hookrightarrow N\left(b, \left(\dfrac{\sigma}{m_i}\right)^2\right)$

$(**)$

la fonction de vraissemblance est :

$$L(x_1, \dots, x_n, b) = \prod_{i=1}^{n} \frac{m_i}{\sigma} \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{(z_i - b)m_i^2}{2\sigma^2}} = \frac{\prod_{i=1}^{n} m_i}{(\sigma\sqrt{2\pi})^n} e^{-\sum_{i=1}^{n} \frac{\left(\frac{x_i}{m_i} - b\right) m_i^2}{2\sigma^2}}$$

$$\Rightarrow \ln L(x_1, \dots, b) = \ln\left[\prod_{i=1}^{n} m_i\right] - n\ln(\sigma\sqrt{2\pi}) - \sum_{i=1}^{n} \frac{(x_i - b m_i)^2}{2\sigma^2}$$

(2)

d'où $\dfrac{\partial L}{\partial b} = \dfrac{1}{\sigma^2} \sum_{i=1}^{n} m_i (x_i - b m_i)$

$\dfrac{\partial L}{\partial b} = 0 \Rightarrow \hat{b}_n = \dfrac{\sum_{i=1}^{n} m_i x_i}{\sum_{i=1}^{n} m_i^2}$

estimateur compliqué à étudier

et $\dfrac{\partial^2 L}{\partial b^2} = -\dfrac{1}{\sigma^2} \sum_{i=1}^{n} m_i^2$

l'information de fisher a pour expression

$I_n(b) = E\left[-\dfrac{\partial^2 \ln L}{\partial b^2}\right] = \dfrac{1}{\sigma^2} \sum_{i=1}^{n} m_i^2 = \dfrac{n\bar{m}}{\sigma^2}$ où on a

posé $\bar{m} = \dfrac{1}{n} \sum_{i=1}^{n} m_i^2$

d'après (**) on peut poser $\hat{b}_n = \dfrac{1}{n} \sum_{i=1}^{n} z_i = \dfrac{1}{n} \sum_{i=1}^{n} \dfrac{X_i}{m_i}$

c'est un estimateur sans biais convergent, de loi normale d'espérance b et de variance : $V(\hat{b}_n) = \dfrac{1}{n^2} \sum_{i=1}^{n} \dfrac{\sigma^2}{m_i^2}$

$V(\hat{b}_n) = \dfrac{\sigma^2}{n h}$ où h est la moyenne harmonique des m_i^2

c-a-d $h = \dfrac{n}{\sum_{i=1}^{n} \frac{1}{m_i^2}}$ — comme on sait que $h < \bar{m}$ moyenne arith. des m_i^2

on en déduit que $V(\hat{b}_n) = \dfrac{\sigma^2}{n h} > \dfrac{1}{I_n(b)}$

c-a-d l'estimateur $\hat{b}_n$ n'est pas efficace, sauf dans le cas particulier où tous les m_i sont égaux.

EX2 X v.a de densité $f(x) = \begin{cases} \frac{2x}{\theta^2} & \text{si } 0 \leq x \leq \theta \ ; \ \theta > 0. \\ 0 & \text{sinon} \end{cases}$

1.) on a $E(X) = \int_{-\infty}^{+\infty} x f(x) dx = \int_0^\theta \frac{2}{\theta^2} x^2 dx = \frac{2}{\theta^2} \left[\frac{x^3}{3}\right]_0^\theta = \frac{2}{3}\theta$

soit $T_n = \frac{3}{2} \bar{X}_n$ où $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$ alors $E(T_n) = \theta$

$\Rightarrow T_n$ est un estimateur sans biais de θ.

de plus $V(T_n) = \left(\frac{3}{2}\right)^2 V(\bar{X}_n) = \frac{9}{4} \cdot \frac{1}{n} V(X)$

or $V(X) = E(X^2) - (E(X))^2$ et $E(X^2) = \int_0^\theta x^2 f(x) dx = \int_0^\theta \frac{2}{\theta^2} x^3 dx$

$= \frac{2}{\theta^2} \left[\frac{x^4}{4}\right]_0^\theta = \frac{\theta^2}{2} \Rightarrow V(X) = \frac{\theta^2}{2} - \frac{4}{9}\theta^2 = \frac{\theta^2}{18}$

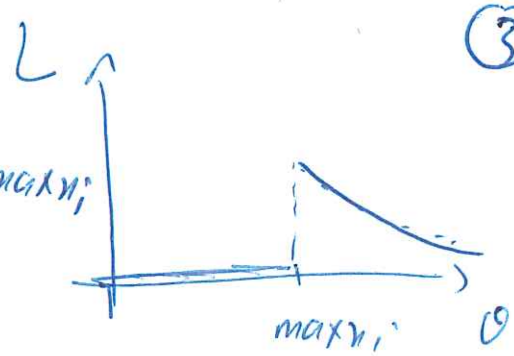
d'où $V(T_n) = \frac{9}{4} \cdot \frac{1}{n} \cdot \frac{\theta^2}{18} = \frac{\theta^2}{8n}$

$\Rightarrow T_n$ est un estimateur convergent de θ.

Rq $X(\Omega) = [0, \theta]$ dépend du paramètre à estimer c'est pourquoi on n'va pas étudier l'efficacité qui repose sur ~~la~~ comparaison $V(T_n)$ et $I_n(\theta)$ quand les hyp. de Cramer-RAO sont vérifiés ce qui n'est pas le cas ici car $X(\Omega)$ dépend de θ.

2.) la fonction de vraissemblance est :

$L(x_1, \dots, x_n, \theta) = \prod_{i=1}^n f(x_i, \theta) = \left(\frac{2}{\theta^2}\right)^n \prod_{i=1}^n x_i$ pour $0 \leq \min x_i \leq \max x_i \leq \theta$



La fonction L est nulle pour $\theta < \max x_i$, et décroissante pour $\theta > \max x_i$ ce qui montre que elle atteint son maximum en $\theta = \max x_i$ donc l'emv est :

$$\Pi_n = \max(X_1, \dots, X_n).$$

Cherchons la loi de probabilité de Π_n.

$$P(\Pi_n \leq x) = P(\max(X_1, \dots, X_n) \leq x) = P(X_1 \leq x \text{ et} \dots X_n \leq x)$$
$$= P(\cap(X_i \leq x)) = \prod_{i=1}^{n} P(X_i \leq x) = F_{X_i}^n(x, \theta) \text{ car}$$

les X_i sont indep. (F_{X_i} fonction de repartition de X_i)

la densité de Π_n est donc $g(x,\theta) = [F_X^n(x,\theta)]' = n F^{n-1}(x,\theta) f(x,\theta)$

or la fonction de repartition $F_X(x,\theta) = \begin{cases} 0 & \text{si } x < 0 \\ \frac{x^2}{\theta^2} & \text{si } 0 \leq x \leq \theta \\ 1 & \text{si } x > \theta \end{cases}$

d'où $g(x,\theta) = \begin{cases} 2n \dfrac{x^{2n-1}}{\theta^{2n}} & \text{si } 0 \leq x \leq \theta \\ 0 & \text{sinon} \end{cases}$

et $E(\Pi_n) = \dfrac{2n}{\theta^{2n}} \displaystyle\int_0^\theta x \cdot x^{2n-1} \, dx = \dfrac{2n}{\theta^{2n}} \left[\dfrac{x^{2n+1}}{2n+1} \right]_0^\theta = \dfrac{2n}{2n+1} \theta$

l'estimateur sans biais de θ est $\hat{\theta}_n = \dfrac{2n+1}{2n} \Pi_n$

de plus $E(\Pi_n^2) = \dfrac{2n}{\theta^{2n}} \displaystyle\int_0^\theta x^2 \cdot x^{2n-1} \, dx = \dfrac{2n}{2n+2} \theta^2$

d'où $V(\Pi_n) = \frac{2n}{2n+2}\theta^2 - \frac{(2n)^2}{(2n+1)^2}\theta^{2} = \frac{2n\theta^2}{(2n+2)(2n+1)^2}$

et $V(\hat{\theta}_n) = \left(\frac{2n+1}{2n}\right)^2 V(\Pi_n) = \frac{\theta^2}{2n(2n+2)}$

L'estimateur $\hat{\theta}_n$ est donc convergent et infiniment plus efficace que T_n ~~au~~ sens où $\frac{V(\hat{\theta}_n)}{V(T_n)} = \frac{2}{n+1} \xrightarrow[n\to+\infty]{} 0$

$V(\hat{\theta}_n) \underset{n\to+\infty}{=} o(V(\Pi_n))$

3.) on va construire un intervalle de confiance à partir de Π_n dont la loi est connue, de fonction de repartition :

$$G(x,\theta) = \begin{cases} 0 & \text{si } x<0 \\ \frac{x^{2n}}{\theta^{2n}} & \text{si } 0\leq x\leq\theta \\ 1 & \text{si } x>\theta \end{cases}$$

on cherche a et b tq $1-\alpha = P(\theta a \leq \Pi_n \leq \theta b)$ ou bien $\Leftrightarrow P\left(\frac{\Pi_n}{b} < \theta < \frac{\Pi_n}{a}\right)$

$1-\frac{\alpha}{2}-\frac{\alpha}{2} = G(\theta b,\theta) - G(\theta a,\theta)$

si on pose $G(\theta a,\theta) = \frac{\alpha}{2}$ et $G(\theta b,\theta) = 1-\frac{\alpha}{2}$

alors $a^{2n} = \frac{\alpha}{2}$ et $b^{2n} = 1-\frac{\alpha}{2}$ c.a.d $a = \left(\frac{\alpha}{2}\right)^{\frac{1}{2n}}$ et $b = \left(1-\frac{\alpha}{2}\right)^{\frac{1}{2n}}$

d'où $\left[\frac{\Pi_n}{b}, \frac{\Pi_n}{a}\right]$

Application

$n = 20$

$\Pi_n = 5$

$\alpha = 0{,}05 \Rightarrow \frac{\alpha}{2} = 0{,}025$; $1-\alpha = 0{,}975$

$a = (0{,}025)^{\frac{1}{40}}$; $b = (0{,}975)^{\frac{1}{40}}$

$\Rightarrow I = \left[5.(0{,}975)^{-\frac{1}{40}} \; ; \; 5.(0{,}025)^{-\frac{1}{40}}\right]$

EX3 $X_i \hookrightarrow N(\theta, (\sqrt{\theta})^2)$; $\theta > 0$

On a $E(X) = \theta$ et $V(X) = \theta$, on peut prendre comme estimateurs sans biais de θ, les moments empiriques :

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i \quad \text{et} \quad \bar{S}_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2$$

ces 2 estimateurs sont aussi convergents car :

$$V(\bar{X}_n) = \frac{\theta}{n} \quad \text{et} \quad V(\bar{S}_n^2) = \frac{2\theta^2}{n-1}$$ ← on l'admet.

le rapport $\frac{V(\bar{S}_n^2)}{V(\bar{X}_n)} = \frac{n}{n-1} \cdot 2\theta \xrightarrow[n \to +\infty]{} 2\theta$ et comme θ est inconnue on ne peut pas comparer ces 2 estimateurs.

Etudions leurs efficacité :

la fonction de vraissemblance est :

$$L(x_1, \dots, x_n, \theta) = \prod_{i=1}^{n} f(x_i, \theta) = \prod_{i=1}^{n}\left(\frac{1}{\sqrt{2\pi\theta}} e^{-\frac{1}{2\theta}(x_i - \theta)^2}\right)$$

$$\Rightarrow \ln L = -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln\theta - \frac{1}{2\theta}\sum_{i=1}^{n}(x_i - \theta)^2$$

d'où $$\frac{\partial \ln L}{\partial \theta} = -\frac{n}{2\theta} + \frac{1}{2\theta^2}\sum_{i=1}^{n}(x_i-\theta)^2 + \frac{1}{2\theta}\sum_{i=1}^{n} -2(x_i - \theta)$$

$$= -\frac{n}{2\theta} + \frac{1}{2\theta^2}\sum_{i=1}^{n} x_i^2 - \frac{n}{2}$$

fonction de densité $X \hookrightarrow N(m, \sigma^2)$ et $f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-m}{\sigma}\right)^2}$

et $$\frac{\partial^2 \ln L}{\partial \theta^2} = \frac{n}{2\theta^2} - \frac{1}{\theta^3}\sum_{i=1}^{n} x_i^2$$

d'où la quantité d'information de Fisher

$$I_n(\theta) = -E\left(\frac{\partial^2 \ln L}{\partial \theta^2}\right) = \frac{-n}{2\theta^2} + \frac{1}{\theta^3} n\, E(X^2)$$

or $V(X) = E(X^2) - (E(X))^2 \Rightarrow E(X^2) = \theta + \theta^2$

d'où $I_n(\theta) = \frac{-n}{2\theta^2} + \frac{n}{\theta^2} + \frac{n}{\theta} = \frac{n}{2\theta^2} + \frac{n}{\theta}$

on peut vérifier que $\frac{1}{V(\bar{X}_n)} < I_n(\theta)$ et $\frac{1}{V(\bar{S}_n^2)} < I_n(\theta)$

donc aucun de ces 2 estimateurs n'est efficace.

2°) Pour construire un intervalle de confiance on va utiliser que $\sqrt{n}\,\frac{\bar{X}_n - \theta}{\bar{S}_n}$ suit la loi T_{n-1} loi de student à n-1 degrés de liberté. $P(\bar{X}_n - \varepsilon < \theta < \bar{X}_n + \varepsilon) = 1-\alpha$

on cherche ε tq $P(-\varepsilon \le \bar{X}_n - \theta \le \varepsilon)$ ou bien

$$P\left(-\frac{\sqrt{n}\,\varepsilon}{\bar{S}_n} \le \sqrt{n}\,\frac{\bar{X}_n - \theta}{\bar{S}_n} \le \frac{\sqrt{n}}{\bar{S}_n}\varepsilon\right) = 1-\alpha \quad (*)$$

si on pose F la fonction de répartition de T_{n-1} alors

$(*) \Rightarrow \frac{\sqrt{n}\,\varepsilon}{\bar{S}_n} = F^{-1}(1-\alpha/2)$; posons $t_{n-1,\alpha} = F^{-1}(1-\alpha/2)$

alors $I = \left[\bar{X}_n - \frac{\bar{S}_n}{\sqrt{n}} t_{n-1,\alpha}, \bar{X}_n + \frac{\bar{S}_n}{\sqrt{n}} t_{n-1,\alpha}\right]$

Application pour $\alpha = 0,05$; $\bar{X}_n = 2,01$; $\bar{S}_n = \sqrt{\frac{48,12}{24}}$

$\sum_{i=1}^{n}(x_i - \bar{x})^2 = 48,12$ $\quad 1-\frac{\alpha}{2} = 0,975$ $\quad = \frac{50,23}{25}$ $\quad = 1,42$

et $t_{24,0,05} = 2,064$

$\bar{S}_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$ $\quad$ table $\quad$ d'où $I = [1,42 , 2,59]$