Corrigé

EX 1 X_n une v.a.r. de densité $f_n(x) = n e^{-nx}$ pour $x > 0$

on a $E(X_n^2) = \int_0^{+\infty} x^2 \underbrace{n e^{-nx}}_{(-e^{-nx})'} dx = \left[-x^2 e^{-nx}\right]_0^{+\infty} + 2\int_0^{+\infty} x e^{-nx} dx$

$$= 2\left[-x \frac{e^{-nx}}{n}\right]_0^{+\infty} + \frac{2}{n}\int_0^{+\infty} e^{-nx} dx = \frac{2}{n}\left[-\frac{e^{-nx}}{n}\right]_0^{+\infty} = \frac{2}{n^2}$$

$\Rightarrow E(|X_n - 0|^2) = \frac{2}{n^2} \xrightarrow[n\to+\infty]{} 0 \Rightarrow X_n \xrightarrow[m.q.]{} 0$

$X_n \xrightarrow[m.q]{} X$ ssi $\lim_{n\to+\infty} E(|X_n - X|^2) = 0$

EX 2 $(X_1, \dots, X_n)$ un echantillon d'une v.a.r. X de densité

$$f(x,\theta) = \begin{cases} \frac{1}{\theta} x^{\frac{1}{\theta}-1} & \text{si } 0 \le x \le 1 \\ 0 & \text{sinon} \end{cases}$$

θ parametre inconnu st^t positif.

a) soit $Y = -\ln X$; soit G la fonction de repartition de Y alors $G(y) = P(Y \le y) = P(-\ln X \le y)$

- si $y \le 0$ alors $G(y) = 0$
- si $y > 0$ alors $G(y) = P(\ln X \ge -y) = P(X \ge e^{-y})$

$$= 1 - P(X \le e^{-y}) = 1 - F(e^{-y}) \text{ où } F \text{ st}$$

la fonction de repartition de la v.a.r X.

$\Rightarrow$ la densité de la v.a.r. Y st $g(y) = G'(y) = e^{-y} f(e^{-y})$

$$= e^{-y} \frac{1}{\theta} e^{-y(\frac{1}{\theta}-1)} = \frac{1}{\theta} e^{-\frac{y}{\theta}}$$

d'où $g(y) = \begin{cases} \frac{1}{\theta} e^{-\frac{y}{\theta}} & \text{si } y \geq 0 \\ 0 & \text{sinon} \end{cases}$

on reconnait la densité de la loi exponentielle de paramètre $\frac{1}{\theta}$.

b) . $E(X) = \frac{1}{\theta} \int_0^1 x . x^{\frac{1}{\theta}-1} dx = \frac{1}{\theta} \int_0^1 x^{\frac{1}{\theta}} dx = \frac{1}{\theta} . \frac{1}{1+\frac{1}{\theta}} \left[x^{\frac{1}{\theta}+1} \right]_0^1$

$= \frac{1}{1+\theta}$

. $V(X) = E(X^2) - (E(X))^2$.

or $E(X^2) = \frac{1}{\theta} \int_0^1 x^2 x^{\frac{1}{\theta}-1} dx = \frac{1}{\theta} \int_0^1 x^{\frac{1}{\theta}+1} dx = \frac{1}{\theta} \frac{1}{\frac{1}{\theta}+2} \left[x^{\frac{1}{\theta}+2} \right]_0^1$

$= \frac{1}{2\theta+1}$

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

c) L'expression de la vraissemblance est :

$L(x_1,..,x_n,\theta) = \prod_{i=1}^{n} f(x_i,\theta) = \prod_{i=1}^{n} \frac{1}{\theta} x_i^{\frac{1}{\theta}-1} = \frac{1}{\theta^n} \left(\prod_{i=1}^{n} x_i \right)^{\frac{1}{\theta}-1}$

$\Rightarrow \ln(L(x,\theta)) = -n \ln\theta + (\frac{1}{\theta}-1) \sum_{i=1}^{n} \ln x_i$

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

$\frac{\partial \ln L}{\partial \theta} = 0 \Leftrightarrow -\frac{1}{n} \sum_{i=1}^{n} \ln x_i = \theta$

. $\frac{\partial^2 \ln L}{\partial \theta^2} = \frac{n}{\theta^2} + \frac{2}{\theta^3} \sum_{i=1}^{n} \ln x_i \Rightarrow \frac{\partial^2 \ln L}{\partial \theta^2} \Big|_{\theta = -\frac{1}{n}\sum_{i=1}^{n} \ln x_i} = \frac{n}{\theta^2} - \frac{2n\theta}{\theta^3}$

$= -\frac{n}{\theta^2} < 0$

donc l'estimateur du maximum de vraissemblance est : ②

$$T_n = -\frac{1}{n}\sum_{i=1}^{n} \ln X_i$$

. $E(T_n) = E(-\frac{1}{n}\sum_{i=1}^{n}\ln X_i) = E(-\ln X) = E(Y) = \theta$ donc

T_n est un estimateur sans biais.

. $V(T_n) = V(-\frac{1}{n}\sum_{i=1}^{n}\ln X_i) = \frac{1}{n^2}V(\sum_{i=1}^{n} Y_i) = \frac{V(Y)}{n} = \frac{\theta^2}{n} \underset{n\to+\infty}{\longrightarrow} 0$

$\Rightarrow T_n$ est un estimateur convergent. Pour savoir s'il est efficace on calcule la quantité d'information de Fisher :

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

$$= -\frac{n}{\theta^2} + \frac{2}{\theta^3}\cdot n\theta = \frac{n}{\theta^2}$$

ainsi on a $V(T_n) = \frac{\theta^2}{n} = \frac{1}{I_n(\theta)}$ donc T_n est efficace.

<u>EX3</u> l'hypothèse alternative étant une hypothèse multiple on détermine d'abord la région critique du test entre les 2 hypothèses simples

$$\begin{cases} H_0 : \sigma = \sigma_0 = 5 \\ H_1 : \sigma = \sigma_1 \quad \text{avec } \sigma_1 > 5. \end{cases}$$

L'expression de la vraissemblance est :

$$L(x_1, .., x_n, \sigma) = \prod_{i=1}^{n} \frac{1}{\sigma\sqrt{2\pi}} \cancel{e^{-\frac{(x_i-100)^2}{2\sigma^2}}} e^{-\frac{1}{2\sigma^2}(x_i-100)^2}$$

$$= \frac{1}{(\sigma\sqrt{2\pi})^n} e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-100)^2}$$

$$\Rightarrow \frac{L_0}{L_1} = \frac{L(x,\sigma_0)}{L(x,\sigma_1)} = \left(\frac{\sigma_1}{\sigma_0}\right)^n e^{-\frac{1}{2}\left(\frac{1}{\sigma_0^2} - \frac{1}{\sigma_1^2}\right)\sum_{i=1}^{n}(x_i-100)^2}$$

le Th. de Neyman et Pearson qui permet d'obtenir le test de puissance maximum, définit la region critique par

$\frac{L_0}{L_1} \leq K \Rightarrow$ en prenant les log $-\frac{1}{2}\left(\frac{1}{\sigma_0^2}-\frac{1}{\sigma_1^2}\right)\sum_{i=1}^{m}(x_i-100)^2 \leq k_2$

puisque $\sigma_1^2 > \sigma_0^2 \Rightarrow \sum_{i=1}^{m}(x_i-100)^2 \geq C$

$\Updownarrow$

$\frac{1}{\sigma_1^2} < \frac{1}{\sigma_0^2}$

La valeur du seuil c est determinée par :

$$\alpha = P(W \mid H_0) = P\left(\sum_{i=1}^{m}(X_i-100)^2 \geq c \mid \sigma=\sigma_0=5\right)$$

sous l'Hypothèse H_0 les v.a. X_i-100 sont normals, centrées, d'écart type $\sigma=\sigma_0=5$ et independants $\Rightarrow \sum_{i=1}^{m}\left(\frac{X_i-100}{5}\right)^2$ suit une loi du χ^2_m d'où $P\left(\sum_{i=1}^{m}\frac{(X_i-100)^2}{25} \geq \frac{c}{25}\right)=\alpha \Leftrightarrow P\left(\chi^2_m \geq \frac{c}{25}\right)=\alpha$

$\Rightarrow P\left(\chi^2_m \leq \frac{c}{25}\right) = 1-\alpha$. posons $\chi^2_{m,1-\alpha}$ le fractile d'ordre $1-\alpha$ de la loi χ^2_m alors $\frac{c}{25}=\chi^2_{m,1-\alpha} \Rightarrow c = 25\chi^2_{m,1-\alpha}$

$$\Rightarrow W = \left\{x \in \mathbb{R}^m \,/\, \sum_{i=1}^{m}(x_i-100)^2 \geq 25\,\chi^2_{m,1-\alpha}\right\}.$$

La region critique est independante de la valeur choisie pour σ_1 donc elle est aussi celle du test U.P.P. pour l'hypothèse alternative $\sigma > 5$.

La fonction puissance est définie par :

$$\eta = 1-\beta = P(W/H_1) = P\left(\sum_{i=1}^{m}(X_i-100)^2 \geq c \mid \sigma > 5\right)$$

$$= P\left(\sum_{i=1}^{m}(X_i-100)^2/\sigma^2 \geq \frac{c}{\sigma^2}\right) = 1-F_m\left(\frac{c}{\sigma^2}\right) \text{ où}$$

F_m est la fonction de repartition de la loi χ^2_m.

Application numérique $m=10$; $\alpha=0{,}05$ alors

$$c \simeq 25 \cdot 18{,}3 = 457{,}5$$

Rq si m était inconnue alors on a toujours la zone de rejet $W = \{x \in \mathbb{R}^n / \sum_{i=1}^{n}(x_i - m)^2 \geq k\}$

on remplace alors m par $\bar{X}_n \Rightarrow W = \{x \in \mathbb{R}^n / \sum_{i=1}^{n}(x_i - \bar{x}_n)^2 \geq k\}$

pour déterminer k, on utilise le fait que

$\frac{(n-1)\bar{S}_n^2}{\sigma^2}$ suit la loi du χ^2_{n-1} ; $\bar{S}_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2$

d'où $P(W | H_0) = \alpha \Leftrightarrow P\left(\sum_{i=1}^{n}(X_i - \bar{X}_n)^2 \geq k \,|\, \sigma = \sigma_0\right)$

d'où $P\left((n-1)\cancel{\sum_{i=1}^{n}(X_i - \bar{X}_n)^2}\bar{S}_n^2 \geq k \,|\, \sigma = \sigma_0 = \sigma\right) = \alpha$

d'où $P\left(\frac{(n-1)\cancel{\sum_{i=1}^{n}(X_i - \bar{X}_n)^2}\bar{S}_n^2}{\sigma_0^2} \geq \frac{k}{\sigma_0^2}\right) = \alpha$

c.à.d $P\left(\chi^2_{n-1} \geq \frac{k}{\sigma_0}\right) = \alpha \Rightarrow P\left(\chi^2_{n-1} \leq \frac{k}{\sigma_0^2}\right) = 1 - \alpha$

si on pose $\chi^2_{n-1, 1-\alpha}$ le fractile d'ordre $1-\alpha$ de la loi du χ^2 à $(n-1)$ degrés de liberté alors $k = \sigma_0^2 \chi^2_{n-1, 1-\alpha}$

$\Rightarrow W = \{x \in \mathbb{R}^n / \sum_{i=1}^{n}(x_i - \bar{x}_n)^2 \geq \sigma_0^2 \chi^2_{n-1, 1-\alpha}\}$