STATISTIQUE MATHÉMATIQUE - CORRIGÉ DE L'EXAMEN .

ENSAI - Mastères spécialisés - 2021/2022 -

École nationale de la statistique et de l'analyse de l'information

Durée de l'épreuve : 2 heures. Tous documents autorisés, calculatrices interdites. Le barème est indicatif et sans engagement (il sera tenu compte de la longueur du sujet). La qualité de la rédaction entre pour une part importante dans la notation.

I. 13 points

1°. Soit $x = (x_1, ..., x_n)$ une réalisation de l'échantillon. En posant $s = \sum_{i=1}^{n} x_i$, on a :

$$L(x, \lambda) = \prod_{i=1}^{n} \left(e^{-\lambda} \frac{\lambda^{x_i}}{x_i!} \right) = e^{-n\lambda} \frac{\lambda^s}{\prod(x_i!)} \tag{1}$$

$\boxed{\text{1 point}}$

Le modèle est régulier (homogène, dominé par la mesure de comptage sur $\mathbb{N}^n$, la vraisemblance est de classe C^2 en λ). $\boxed{\text{0.5 point}}$

2°. Si c désigne une constante ne dépendant pas de λ, on a :

$$\frac{\partial L(x, \lambda)}{\partial \lambda} = c \times (-n\lambda^s e^{-n\lambda} + e^{-n\lambda} s\lambda^{s-1}) \tag{2}$$

qui est nul si, et seulement si, $\lambda = s/n$. Par ailleurs il s'agit bien d'un maximum (c'est important de le vérifier) car

$$\frac{\partial^2 L(s, \lambda)}{\partial \lambda^2} < 0 \tag{3}$$

Ainsi, $\widehat{\lambda} = \overline{X} = S/n$ est un estimateur du maximum de vraisemblance de λ $\boxed{\text{1 point}}$

Posons $\psi(x) = xe^{-x}$. ψ est clairement une fonction de classe C^∞, en particulier elle est de classe C^2 et $\beta = \psi(\lambda)$. D'après le théorème de reparamétrisation $\boxed{\text{0.5 point}}$, un estimateur du maximum de vraisemblance est donc donné par

$$\widehat{\beta} = \overline{X} \exp(-\overline{X}) \tag{4}$$

$\boxed{\text{1 point}}$

Les X_i sont des v.a.i.i.d. intégrables car $\mathbb{E}[|X_i|] = \mathbb{E}[X_i] = \lambda$, on peut donc appliquer la loi forte des grandes nombres $\boxed{\text{1 point}}$ et l'on a alors

$$\widehat{\lambda} \xrightarrow[n\to+\infty]{\mathbb{P}-\text{p.s.}} \lambda \tag{5}$$

et le théorème de l'application continue assure alors que

$$\widehat{\beta} \xrightarrow[n\to+\infty]{\mathbb{P}-\text{p.s.}} \beta \tag{6}$$

$\widehat{\beta}$ est donc fortement consistant. *a fortiori*, il est également consistant. $\boxed{\text{0.5 point}}$

Les X_i sont également de carré intégrable et l'on peut donc appliquer le théorème de la limite centrale $\boxed{\text{1 point}}$:

$$\sqrt{n}\,(\widehat{\lambda} - \lambda) \rightsquigarrow \mathcal{N}(0, \lambda) \tag{7}$$

$\widehat{\lambda}$ est donc $\sqrt{n}$-consistant et asymptotiquement normal. Par ailleurs, ψ est de classe C^1 et $\psi'(x) = (1-x)e^{-x}$. En appliquant la méthode delta, on en déduit que $\widehat{\beta}$ est $\sqrt{n}$-consistant et asymptotiquement normal et que

$$\sqrt{n}\,(\widehat{\beta} - \beta) \rightsquigarrow \mathcal{N}\left(0, \lambda \times \psi'(\lambda)^2\right) \tag{8}$$

$\boxed{\text{1 point}}$

3°. Un calcul facile (plusieurs fois fait en cours) donne

$$\mathbb{I}_X(\lambda) = \frac{n}{\lambda} \tag{9}$$

Le modèle (initial) est par ailleurs régulier et la variance de l'échantillon vaut λ/n. La borne de Cramer Rao est donc atteinte et l'on en déduit que l'estimateur est efficace. Mais la question porte sur $\mathbb{I}_X(\beta)$. En utilisant le théorème de reparamétrisation, il vient :

$$\mathbb{I}_X(\beta) = \frac{\mathbb{I}_X(\lambda)}{\psi'(\lambda)^2} = n\frac{e^{2\lambda}}{\lambda(1-\lambda)^2}. \tag{10}$$

$\widehat{\beta}$ n'est pas efficace car il est biaisé. En effet, un calcul fait en cours donne :

$$\mathbb{E}[\widehat{\beta}] = \frac{1}{n} \sum_{k\geq 1} ke^{-k/n} e^{-n\lambda} \frac{(n\lambda)^k}{k!} \tag{11}$$

$$= \lambda \exp\left(-\frac{1}{n} - n\lambda(1 - e^{-1/n})\right) \neq \beta. \tag{12}$$

$\boxed{\text{1.5 points}}$

4°. Les $(Y_i)_i$ sont clairement des v.a. de Bernoulli de paramètre

$$\mathbb{P}[Y_i = 1] = \mathbb{P}[X_i = 1] = \beta \tag{13}$$

En posant $\overline{Y} = (\sum Y_i)/n$ nous obtenons un estimateur plug-in de β, non biaisé. $\boxed{\text{1 point}}$

Le modèle d'échantillonnage associé aux Y_i est un modèle exponentiel de statistique naturelle $\overline{Y}$. L'espace des paramètres est

$$\Lambda(\beta) = \{\ln \frac{\beta}{1-\beta};\ \beta \in [0,1]\} \tag{14}$$

qui est égal à $\mathbb{R}$ et est donc un ouvert de $\mathbb{R}$. Par suite, $\overline{Y}$ est une statistique exhaustive complète pour β. $\boxed{\text{1 point}}$

D'après le théorème de Lehmann-Scheffé, comme $\overline{Y}$ est également un estimateur sans biais de β, on en déduit que $\overline{Y} = \mathbb{E}[\overline{Y}|\overline{Y}]$ est un estimateur VUMSB. $\boxed{\text{1 point}}$

5°. D'après la formule de Bayes, la vraisemblance *a posteriori* est proportionnelle à

$$\lambda^{a-1} b^a e^{-b\lambda} e^{-n\lambda} \frac{\lambda^s}{\prod_{i=1}^n x_i} \tag{15}$$
$$\propto \lambda^{a+s-1} e^{-(n+b)\lambda} \tag{16}$$

avec $s = \sum_{i=1}^n x_i$.

Il s'agit donc d'une loi $\gamma(a+S, n+b)$ 1 point dont la moyenne vaut

$$\mathbb{E}[H|X] = \lambda^\star = \frac{a+S}{n+b} \tag{17}$$

$\lambda^\star$ est donc biaisé, mais asymptotiquement sans biais.
0.5 point

II. 14 points

1°. Soit $x = (x_1, ..., x_n)$ une réalisation de l'échantillon. La vraisemblance est

$$L(x,\theta) = (2\theta^2)^{-n} \prod_{i=1}^n x_i \times \mathbb{1}_{[0,2\theta]^n}(x) \tag{18}$$

1 point

2°. Le modèle n'est pas homogène car le support dépend de θ : il n'est pas régulier. Si $x_{(n)} \geq 2\theta$, alors $L(x,\theta) = 0$ et sinon la vraisemblance est une fonction décroissante de θ. 1 point On en déduit donc que (attention au facteur 1/2)

$$\widehat{\theta}_1 = \frac{1}{2} \sup_{i=1}^n X_i = \frac{1}{2} X_{(n)} \tag{19}$$

0.5 point

3°. Commençons par déterminer la fonction de répartition F ; pour $x \in [0, 2\theta]$,

$$F(x) = \int_0^x \frac{t}{2\theta^2} dt = \left(\frac{x}{2\theta}\right)^2 \tag{20}$$

Par ailleurs $F(x) = 0$ si $x \leq 0$ et $F(x) = 1$ si $x \geq 2\theta$.

Soit $G_n(x)$ la fonction de répartition de $\widehat{\theta}_1$. Pour $x \in [0,\theta]$, on a :

$$G_n(x) = \mathbb{P}[X_{(n)} \leq 2x] = F(2x)^n = \left(\frac{x}{\theta}\right)^{2n} \tag{21}$$

et $G_n(x)$ est nulle avant 0 et vaut 1 après θ (attention aux indicatrices!). 1 point

En dérivant, on obtient alors la densité $g_n(x)$ de $\widehat{\theta}_1$ (qui est portée par l'intervalle $[0,\theta]$) :

$$g_n(x) = \frac{2n}{\theta} \left(\frac{x}{\theta}\right)^{2n-1} \mathbb{1}_{[0,\theta]}(x) \tag{22}$$

0.5 point

On peut alors calculer espérance, variance et erreur quadratique :

$$\mathbb{E}[\widehat{\theta}_1] = \frac{2n}{\theta^{2n}} \int_0^\theta x^{2n} dx = \frac{2n}{2n+1}\theta \tag{23}$$

On en déduit alors que l'estimateur est biaisé, mais asymptotiquement sans biais. 1 point

$$\mathbb{E}[\widehat{\theta}_1^2] = \frac{2n}{\theta^{2n}} \int_0^\theta x^{2n+1} dx = \frac{n}{n+1}\theta^2 \tag{24}$$

et par suite,

$$\mathbb{V}[\widehat{\theta}_1] = \frac{n}{(n+1)(2n+1)^2}\theta^2 \tag{25}$$

1 point

On en déduit alors (sans même avoir besoin de calculer l'erreur quadratique) que l'estimateur converge en norme L^2 (car il est asymptotiquement sans biais et que sa variance tend vers 0) et également qu'il est consistant car la convergence en moyenne quadratique implique la convergence en probabilité. 0.5 point

4°. Soit $0 \leq x \leq \theta$ (attention à l'ordre donné par l'énoncé et au signe de x),

$$\mathbb{P}\left[n\left(\theta - \widehat{\theta}_1\right) \leq x\right] = \mathbb{P}\left[\widehat{\theta}_1 \geq \theta - \frac{x}{n}\right] \tag{26}$$
$$= 1 - \left(1 - \frac{x}{n\theta}\right)^{2n} \tag{27}$$
$$= 1 - \exp\left(2n \ln\left(1 - \frac{x}{n\theta}\right)\right) \tag{28}$$
$$= 1 - \exp\left(-\frac{2x}{\theta} + o(1)\right) \tag{29}$$

Ainsi, pour $x \geq 0$,

$$\lim_{n\to+\infty} \mathbb{P}\left[n\left(\theta - \widehat{\theta}_1\right) \leq x\right] = 1 - \exp\left(-\frac{2x}{\theta}\right) \tag{30}$$

qui est la fonction de répartition d'une loi exponentielle de paramètre $2/\theta$. On en déduit la convergence en loi de $n\left(\theta - \widehat{\theta}_1\right)$ vers une loi exponentielle $\mathscr{E}(2/\theta)$.
2 points

5°.

$$\mathbb{E}[X] = \int_0^{2\theta} \frac{x^2}{2\theta^2} f(x) dx = \frac{4}{3}\theta \tag{31}$$

En posant $\widehat{\theta}_2 = 3\overline{X}/4$ on obtient donc un estimateur des moments θ. Il est clairement sans biais et la loi forte des grands nombres montre que $\widehat{\theta}_2$ converge presque sûrement vers θ. On a donc consistance forte de l'estimateur en appliquant le théorème de l'application continue à $\phi(x) = 3x/4$. 1 point

Ensuite, le théorème de la limite centrale (licite car les v.a.i.i.d. de l'échantillon sont de carré intégrable) donne

$$\sqrt{n}\,(\overline{X} - 4\theta/3) \rightsquigarrow \mathcal{N}\left(0, \sigma^2\right) \tag{32}$$

où $\sigma^2 = \mathbb{V}(X) = 2\theta^2/9$

1 point

Puis en appliquant la méthode delta à $\phi(x)$, il vient :

$$\sqrt{n}\left(\widehat{\theta}_2 - \theta\right) \rightsquigarrow \mathcal{N}(0, 1/8) \tag{33}$$

$\boxed{\text{1 point}}$

6°. Par définition de la médiane (théorique) M,

$$\int_0^M \frac{x}{2\theta^2} dx = \frac{1}{2} \tag{34}$$

et l'on a donc $M = \theta\sqrt{2}$ $\boxed{\text{1 point}}$

Si $n = 2m + 1$, la médiane empirique est la statistique d'ordre $X_{(m)}$; l'estimateur plug-in associé à la médiane est donc

$$\widehat{\theta}_3 = \frac{X_{(m)}}{\sqrt{2}} \tag{35}$$

$\boxed{\text{0.5 point}}$