LSTAT 2040 - TP 3 : Solutions

Information de Fisher et borne de Rao–Cramer

Exercice 1

On calcule directement que $\mathrm{Var}[\overline{X}_n] = p(1-p)/n$. Afin de la comparer avec la borne de Rao–Cramer, nous calculons l'information de Fisher de l'échantillon $I_n(p) = nI(p)$ où l'information $I(p)$ contenue dans une seule observation est obtenue via la formule

$$I(p) = -\mathrm{E}[\partial_p^2 \log f(X;p)],$$

où $f(x;p)$ désigne la fonction de masse d'une $\mathrm{Be}(p)$. On a

$$\begin{aligned}
-\mathrm{E}[\partial_p^2 \log f(X;p)] &= -\mathrm{E}[\partial_p^2 \log(p^X(1-p)^{1-X})] \\
&= -\mathrm{E}[\partial_p^2 \{X \log(p) + (1-X)\log(1-p)\}] \\
&= -\mathrm{E}[-\tfrac{X}{p^2} - \tfrac{1-X}{(1-p)^2}] \\
&= \frac{p}{p^2} + \frac{1-p}{(1-p)^2} = \frac{1}{p(1-p)}.
\end{aligned}$$

On déduit que $I_n(p) = \frac{n}{p(1-p)}$, et l'on a bien que $\mathrm{Var}[\overline{X}_n] = I_n(p)^{-1}$.

Exercice 2

Un estimateur efficace T pour $g(\theta)$ existe si et seulement si

$$\partial_\theta \log f_n(x;\theta) = a_n(\theta)[T(x) - g(\theta)],$$

où a_n est une fonction de θ telle que $a_n(\theta) \neq 0$ et $f_n(x;\theta) := \prod_{i=1}^n f(x_i;\theta)$ est la densité jointe de $X_1, \ldots, X_n$. Dans chaque cas, on calcule le membre de gauche et on essaie de le factoriser en une expression de la forme du membre de droite. Notons en particulier que le membre de gauche peut également s'exprimer comme $\sum_{i=1}^n \partial_\theta \log f(x_i;\theta)$.

(a)

$$\begin{aligned}
\sum_{i=1}^n \partial_\theta \log f(x_i;\theta) &= \sum_{i=1}^n \left\{ \frac{x_i}{\theta} - \frac{1-x_i}{1-\theta} \right\} \\
&= \frac{n}{\theta(1-\theta)} \{\overline{x}_n - \theta\}.
\end{aligned}$$

L'expression est bien de la forme requise avec $a_n(\theta) = \frac{n}{\theta(1-\theta)}$ et $T(x) = \overline{x}_n$.

(b)

$$\begin{aligned}
\sum_{i=1}^n \partial_\theta \log f(x_i;\theta) &= \sum_{i=1}^n \left\{ \frac{1}{\theta} + \log(x_i) \right\} \\
&= -n \left\{ \frac{\sum_{i=1}^n \log(1/x_i)}{n} - \frac{1}{\theta} \right\}.
\end{aligned}$$

L'expression est bien de la forme requise avec $a_n(\theta) = -n$ et $T(x) = \frac{\sum_{i=1}^n \log(1/x_i)}{n}$.

(c)

$$\begin{aligned}\sum_{i=1}^{n} \partial_\theta \log f(x_i;\theta) &= \sum_{i=1}^{n} \left\{ \frac{1}{\theta \log \theta} - \frac{1}{\theta - 1} + \frac{x_i}{\theta} \right\} \\ &= \frac{n}{\theta} \left\{ \overline{x}_n - \left(\frac{\theta}{\theta - 1} - \frac{1}{\log \theta} \right) \right\}.\end{aligned}$$

L'expression est bien de la forme requise avec $a_n(\theta) = \frac{n}{\theta}$ et $T(x) = \overline{x}_n$.

(d)

$$\begin{aligned}\sum_{i=1}^{n} \partial_\theta \log f(x_i;\theta) &= \sum_{i=1}^{n} \left\{ \frac{1}{\theta} - x_i \right\} \\ &= -n \left\{ \overline{x}_n - \frac{1}{\theta} \right\}.\end{aligned}$$

L'expression serait de la forme voulue si on avait $g(\theta) = 1/\theta$, mais il n'est pas possible d'obtenir une factorisation correcte avec le choix $g(\theta) = \theta$. Aucun estimateur efficace n'existe pour ce paramètre.

Exercice 3

(a) On calcule pour tout $y > 0$,

$$\begin{aligned}\Pr(X_i^3 \le y) &= \Pr(X_i \le y^{1/3}) \\ &= \int_0^{y^{1/3}} \frac{3x^2}{\theta} \exp(-x^3/\theta) dx \\ &= 1 - \exp(-y/\theta).\end{aligned}$$

Pour $y \le 0$, on a trivialement $\Pr(X_i^3 \le y) = 0$. On déduit que $X_i^3 \sim \operatorname{Exp}(1/\theta)$. Par conséquent, on a bien

$$\mathrm{E}[T_n] = \mathrm{E}[X_1] = \theta,$$

et on déduit que T_n est non-biaisé pour θ.

(b) L'estimateur T_n est efficace si sa variance atteint la borne de Rao–Cramer. On calcule

$$\operatorname{Var}[T_n] = \frac{\operatorname{Var}[X_1]}{n} = \frac{\theta^2}{n}.$$

La borne de Rao–Cramer est donnée par $I_n(\theta)^{-1}$ où $I_n(\theta) = nI(\theta)$ est l'information de Fisher contenue dans l'échantillon. On a

$$\begin{aligned}I(\theta) &= -\mathrm{E}[\partial_\theta^2 \log f(X;\theta)] \\ &= -E[\partial_\theta^2 \{\log(3X^2) - \log(\theta) - \tfrac{X^3}{\theta} + \log(\mathrm{I}(0 < x < \infty))\}] \\ &= -\mathrm{E}[\tfrac{1}{\theta^2} - 2\tfrac{X^3}{\theta^3}] \\ &= \frac{1}{\theta^2},\end{aligned}$$

dès lors $I_n(\theta) = n/\theta^2$ et on voir que l'estimateur T_n est efficace.

Exercice 4

(a) Par intégration, on trouve

$$\mathrm{E}[X_1] = \frac{2\theta^2 + \theta}{\theta + 1} \qquad \text{et} \qquad \operatorname{Var}[X_1] = \frac{2\theta^4 + 4\theta^3 + \theta^2}{(\theta + 1)^2}.$$

(b) On remarque que l'on peut réécrire

$$\eta = \frac{(3+2\theta)(2+\theta)}{\theta+1} = \mathrm{E}[X_1] + 6.$$

Dès lors, l'estimateur $\hat{\eta} = \overline{X}_n + 6$ est sans biais pour η car

$$\mathrm{E}[\hat{\eta}] = \mathrm{E}[\overline{X}_n] + 6 = \mathrm{E}[X_1] + 6 = \eta.$$

(c) On doit comparer

$$\mathrm{Var}[\hat{\eta}] = \frac{2\theta^4 + 4\theta^3 + \theta^2}{n(\theta+1)^2}$$

avec la borne de Rao–Cramer $I_n^{-1}(\eta) = (g'(\theta))^2/I_n(\theta)$ où g est telle que $\eta = g(\theta)$. On calcule

$$\begin{aligned} I(\theta) &= -\mathrm{E}[\partial_\theta^2 \log f(X;\theta)] \\ &= -\mathrm{E}[\tfrac{1}{\theta^2} + \tfrac{1}{(\theta+1)^2} - \tfrac{2X}{\theta^3}] \\ &= \frac{2\theta^2 + 4\theta + 1}{\theta^2(\theta+1)^2}. \end{aligned}$$

De plus, on a

$$g'(\theta)^2 = \left(\frac{2\theta^2 + 4\theta + 1}{(\theta+1)^2}\right)^2,$$

donc, on déduit,

$$I_n^{-1}(\eta) = \frac{2\theta^4 + 4\theta^3 + \theta^2}{n(\theta+1)^2} = \mathrm{Var}[\hat{\eta}],$$

i.e., l'estimateur $\hat{\eta}$ est efficace.

Exercice 5

(a) Soit $y > 0$. On a

$$\begin{aligned} \Pr(-\log(X_1) \leq y) = \Pr(X_1 \geq \exp(-y)) &= \int_{\exp(-y)}^{1} \frac{1}{\theta} x^{1/\theta - 1} dx \\ &= 1 - \exp(-y/\theta). \end{aligned}$$

Pour $y \leq 0$, on a trivialement $\Pr(-\log(X_1) \leq y) = 0$. On déduit que $-\log(X_1) \sim \mathrm{Exp}(1/\theta)$. Dès lors

$$\mathrm{E}[T_n] = \mathrm{E}[-\log(X_1)] = \theta,$$

et T_n est non-biaisé pour θ.

(b) On rappelle qu'une somme de variables aléatoires indépendante distribuée selon une même loi exponentielle suit une loi Gamma. En particulier, on a

$$nT_n = \sum_{i=1}^{n} -\log(X_i) \sim \mathrm{Gamma}(n, 1/\theta).$$

On déduit que la densité de nT_n en $t > 0$ est donnée par

$$f_{nT_n}(t) = \frac{t^{n-1}\exp(-t/\theta)}{\theta^n (n-1)!} \mathrm{I}(0 < t < \infty).$$

(c) La variance de T_n est donnée par

$$\mathrm{Var}[T_n] = \frac{\mathrm{Var}[nT_n]}{n^2} = \frac{\theta^2}{n}.$$

Calculons maintenant la borne de Rao–Cramer. On a

$$\begin{aligned} I(\theta) &= -\,\mathrm{E}[\partial_\theta^2 \log f(X;\theta)] \\ &= -\,\mathrm{E}[\tfrac{1}{\theta^2} + \tfrac{2\log(X)}{\theta^3}] \\ &= \frac{2\theta}{\theta^3} - \frac{1}{\theta^2} = \frac{1}{\theta^2}. \end{aligned}$$

Dès lors, on trouve

$$I_n^{-1}(\theta) = \frac{\theta^2}{n} = \mathrm{Var}[T_n],$$

et on déduit que l'estimateur T_n est efficace.

(d) Soit on effectue un calcul direct et on voit que les deux informations sont identiques et valent toutes les deux $I_n(\theta)$. Soit on calcule l'information $I_{nT_n}(\theta)$ à partir de la densité obtenue au point (b) et on observe que la fonction $g : t > 0 \mapsto g(t) := t/n$ est différentiable, strictement monotone et indépendante de θ, dès lors par l'exercice 3.1 du syllabus, on a directement que

$$I_{T_n}(\theta) = I_{g(nT_n)}(\theta) = I_{nT_n}(\theta).$$

Exercice 6

(a) On a

$$\mathrm{E}[T_n] = \mathrm{E}[\overline{X}_n^2] - 1/n = \mathrm{Var}[\overline{X}_n] + \mathrm{E}[\overline{X}_n]^2 - 1/n = \theta^2.$$

(b) On commence par calculer la variance de T_n en utilisant l'indice :

$$\begin{aligned} \mathrm{Var}[T_n] = \mathrm{Var}[\overline{X}_n^2] &= \mathrm{E}[\overline{X}_n^4] - (\mathrm{E}[\overline{X}_n^2])^2 \\ &= \theta^4 + 6\theta^2 \frac{1}{n} + \frac{3}{n^2} - (\theta^2 + \frac{1}{n})^2 \\ &= \frac{4\theta^2}{n} + \frac{2}{n^2}. \end{aligned}$$

La borne de Rao–Cramer est donnée par $I_n^{-1}(\theta^2) = \frac{4\theta^2}{nI(\theta)}$ où

$$I(\theta) = -\,\mathrm{E}[(\partial_\theta \log f(X;\theta))^2] = \mathrm{E}[(X-\theta)^2] = 1.$$

Dès lors,

$$\mathrm{Var}[T_n] = \frac{4\theta^2}{n} + \frac{2}{n^2} > \frac{4\theta^2}{n} = I_n^{-1}(\theta^2),$$

et l'estimateur T_n n'est pas efficace.