Université Paris Dauphine — Année 2002-2003
U.F.R. Mathématiques de la Décision
MD2 - Statistique

Correction Examen 11/06/2003

Exercice 1

1 Soit $f_n(x_1,\ldots,x_n;\theta)$ la densité de probabilité du vecteur aléatoire $(X_1,\ldots,X_n)$:

$$\begin{aligned} f_n(x_1,\ldots,x_n;\theta) &= \prod_{i=1}^n f_\theta(x_i) \\ &= \theta^n \left(\prod_{i=1}^n x_i\right)^{\theta-1} 1_{\max(x_1,\ldots,x_n)<1} 1_{\min(x_1,\ldots,x_n)>0} \end{aligned}$$

Soit $V_n(\theta;x_1,\ldots,x_n)$ la fonction de vraisemblance associée à l'échantillon $(x_1,\ldots,x_n)$.
Pour une réalisation $(x_1,\ldots,x_n)$ du vecteur aléatoire $(X_1,\ldots,X_n)$, il s'agit d'une fonction du paramètre $\theta > 0$:

$$\begin{aligned} V(\theta;x_1,\ldots,x_n) &= \theta^n \left(\prod_{i=1}^n x_i\right)^{\theta-1} \\ &= \theta^n \exp\left((\theta-1)\sum_{i=1}^n \log(x_i)\right) \end{aligned}$$

La vraisemblance associée à $(x_1,\ldots,x_n)$ ne dépend donc de l'échantillon qu'au travers de $\sum_{i=1}^n \log(x_i)$.

2 Soit $Y_i = -\log(X_i)$ et $f_\theta(y)$ la densité de probabilité de Y_i. D'après le théorème du changement de variable :

$$\begin{aligned} f_\theta(y) &= \exp(-y)\theta(\exp(-y))^{\theta-1} 1_{]0,+\infty[}(y) \\ &= \theta\exp(-\theta y) 1_{]0,+\infty[}(y) \end{aligned}$$

Ainsi, Y_i suit une loi exponentielle de paramètre θ.

3 Soit $LV(\theta; x_1, \ldots, x_n)$ la fonction de log-vraisemblance associée à l'échantillon $(x_1, \ldots, x_n)$:

$$LV(\theta; x_1, \ldots, x_n) \quad = \quad n \log(\theta) + (\theta - 1) \sum_{i=1}^{n} x_i$$

Soit θ^* l'estimation du maximum de vraisemblance de θ. D'après la condition nécessaire du premier ordre θ^* est tel que :

$$\frac{n}{\theta^*} + \sum_{i=1}^{n} \log(x_i) = 0.$$

$$\iff \theta^* = -\frac{n}{\sum_{i=1}^{n} \log(x_i)}.$$

Aussi, comme $-\frac{n}{\theta^2} < 0$, $LV()$ est une fonction strictement concave et θ^* est l'estimation du maximum de vraisemblance de θ.
Ainsi, l'estimateur du maximum de vraisemblance de θ est

$$\hat{\theta}(X_1, \ldots, X_n) = -\frac{n}{\sum_{i=1}^{n} \log(X_i)}.$$

4 Soit $Z = \frac{n}{\hat{\theta}(X_1, \ldots, X_n)} = -\sum_{i=1}^{n} \log(X_i) = \sum_{i=1}^{n} -\log(X_i) = \sum_{i=1}^{n} Y_i$.
Dans la mesure où les variables aléatoires Y_i sont iid comme $Y \sim \mathcal{E}(\theta)$, alors $Z \sim \mathcal{G}(n, \theta)$ et ainsi :

$$f_\theta(z) = \frac{\theta^n}{(n-1)!} z^{n-1} \exp(-\theta z) 1_{]0;+\infty[}(z).$$

5 Soit $I = \frac{1}{\hat{\theta}(X_1, \ldots, X_n)} = -\frac{\sum_{i=1}^{n} X_i}{n}$.

$$\mathbb{E}(I) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}(-\log(X_i)) = \frac{1}{n} \sum_{i=1}^{n} \left(\frac{1}{\theta}\right).$$

Et ainsi I est un estimateur non biaisé de $\frac{1}{\theta}$.

6
$$\mathbb{V}(I) = \frac{1}{n^2}\sum_{i=1}^{n}\mathbb{V}\left(-\log\left(X_i\right)\right) = \frac{1}{n\theta^2}.$$

La variance minimale de tout estimateur non biaisé de $\frac{1}{\theta}$ est égale à $\frac{g'(\theta)}{I_n(\theta)}$ (borne FDCR) où $g(\theta) = \frac{1}{\theta}$ et $I_n(\theta) = -\mathbb{E}\left[\frac{d^2LV\left(\theta; X_1, \ldots, X_n\right)}{d\theta^2}\right]$ (information de Fisher apportée par l'échantillon sur θ).
Nous avons $I_n(\theta) = \frac{n}{\theta^2}$ et $g'(\theta) = -\frac{1}{\theta^2}$.
Ainsi la borne FDCR est égale à $\frac{1}{n\theta^2} = \mathbb{V}(i)$. Par conséquent, I est un estimateur efficace de $\frac{1}{\theta}$.

7 Soit $T = \frac{n\theta}{\hat{\theta}\left(X_1, \ldots, X_n\right)}$. Soit $f_T(t)$ la densité de probabilité de T. Nous avons :

$$f_T(t) = \frac{1}{(n-1)!}t^{n-1}\exp(-t)1_{]0,+\infty[}(t).$$

Ainsi, $T \sim \mathcal{G}(n, 1)$ loi de probabilité indépendante de θ.
Soient t_1 et t_2 les fractiles d'ordre $\alpha/2$ et $1-\alpha/2$ d'une loi $\mathcal{G}(n, 1)$:

$$\mathbb{P}\left(t_1 \leq T \leq t_2\right) = 1 - \alpha.$$

On en déduit un intervalle de confiance de niveau $1-\alpha$ pour le paramètre θ :

$$IC_{1-\alpha}(\theta) = \left[t_1\frac{\theta^*}{n}, t_2\frac{\theta^*}{n}\right].$$

Questions à choix multiples

1. $\alpha = \frac{1}{n}$.
2. $I = \exp(-1)$.
4. $\mathcal{E}(\theta)$.
5. $\left[\overline{X_n} - 1.96/\sqrt{n}, \overline{X_n} + 1.96/\sqrt{n}\right]$.