Corrigés du final A06 : OS02 – Théorie de l'estimation

I.Nikiforov

10 janvier 2007

Sujet 1 [5 points] : Soit $\xi_1, \xi_2, \ldots, \xi_n$ une suite de variables aléatoires indépendantes, $\xi_i \sim U(0, 3\theta)$, $\theta > 0$, où $U(a, b)$ est la distribution uniforme sur $[a; b]$, $a < b$. On cherche à estimer le paramètre θ à base de mesures $\xi_1, \xi_2, \ldots, \xi_n$.

1. Trouver un estimateur du paramètre θ par la méthode du maximum de vraisemblance (MV).
2. Cet estimateur MV est-il biaisé ?
3. Cet estimateur MV est-il convergent (en probabilité) ?

Réponses :

1) Tout d'abord, définissons la densité de la loi $U(0, 3\theta)$ pour une variable :

$$f(x; \theta) = \begin{cases} \dfrac{1}{3\theta} & \text{si} \quad x \in [0; 3\theta] \\ 0 & \text{si} \quad x \notin [0; 3\theta] \end{cases}$$

et pour n variables

$$f(x_1, \ldots, x_n; \theta) = \begin{cases} \dfrac{1}{(3\theta)^n} & \text{si} \quad x_i \in [0; 3\theta] \text{ pour } i \in [1; n] \\ 0 & \text{si} \quad \exists i \in [1; n] \text{ t.q. } x_i \notin [0; 3\theta] \end{cases}.$$

Soit $\xi_{(n)} = \max(\xi_1, \ldots, \xi_n)$. La fonction de vraisemblance $\theta \mapsto f(\xi_1, \ldots, \xi_n; \theta)$ pour les valeurs $\xi_1, \ldots, \xi_n$ données

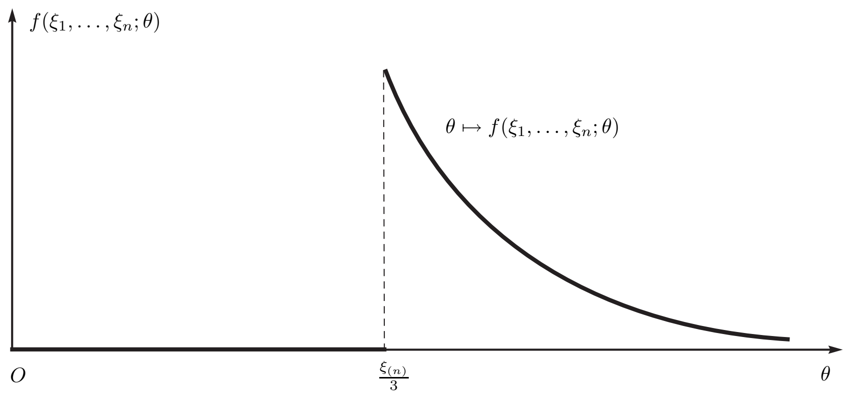

FIG. 1 – La fonction de vraisemblance.

est représentée sur la figure 1. Il est evident que l'estimateur MV du paramètre θ est

$$\widehat{\theta} = \arg\max_{\theta} f(\xi_1, \ldots, \xi_n; \theta) = \frac{\xi_{(n)}}{3}.$$

2) La fonction de répartition d'une loi uniforme $U(a,b)$ est définie par :

$$F(x) = \begin{cases} 0 & \text{si} \quad x < a \\ \dfrac{x-a}{b-a} & \text{si} \quad x \in [a;b] \\ 1 & \text{si} \quad x > b \end{cases}, a < b.$$

La fonction de répartition de la variable $\xi_{(n)} = \max(\xi_1, \ldots, \xi_n)$ est (pour les $x \in [a;b]$)

$$F_{\xi_{(n)}}(x) = \mathbb{P}(\xi_{(n)} < x) = \prod_{i=1}^{n} \mathbb{P}(\xi_i < x) = [\mathbb{P}(\xi_i < x)]^n = \left(\frac{x-a}{b-a}\right)^n.$$

La densité de $\xi_{(n)}$ est

$$f_{\xi_{(n)}}(x) = \begin{cases} \dfrac{d}{dx}\left[\dfrac{x-a}{b-a}\right]^n = n\dfrac{(x-a)^{n-1}}{(b-a)^n} & \text{si} \quad x \in [a;b] \\ 0 & \text{si} \quad x \notin [a;b] \end{cases}, a < b.$$

L'espérance mathématique de la variable aléatoire $\xi_{(n)}$ est

$$\mathbb{E}\left(\xi_{(n)}\right) = \int_a^b x f_{\xi_{(n)}}(x) dx = \int_a^b xn\frac{(x-a)^{n-1}}{(b-a)^n}\, dx = b - \frac{b-a}{n+1}.$$

donc

$$\mathbb{E}\left(\widehat{\theta}\right) = \mathbb{E}\left(\frac{\xi_{(n)}}{3}\right) = \theta - \frac{\theta}{n+1}$$

et l'estimateur MV est biaisé.

3) On dit que l'estimateur $\hat{\theta}_n = \hat{\theta}(\xi_1, \ldots, \xi_n)$ est convergent si $\forall \theta \in \Theta$

$$\forall \varepsilon > 0 \ : \lim_{n \to \infty} \mathbb{P}_\theta(|\hat{\theta}_n - \theta| > \varepsilon) = 0$$

D'après le paragraph 2),

$$\mathbb{P}_\theta(|\hat{\theta}_n - \theta| > \varepsilon) = \mathbb{P}_\theta(\hat{\theta}_n < \theta - \varepsilon) = \mathbb{P}_\theta(\xi_{(n)} < 3\theta - 3\varepsilon) = \int_0^{3\theta - 3\varepsilon} n\frac{x^{n-1}}{(3\theta)^n} dx = \left(1 - \frac{\varepsilon}{\theta}\right)^n.$$

Donc,

$$\forall \theta > 0 \text{ et } \forall\ 0 < \varepsilon < \theta \lim_{n \to \infty} \left(1 - \frac{\varepsilon}{\theta}\right)^n = 0$$

et l'estimateur est convergent.

Sujet 2 [5 points] : Soit $\xi_1, \xi_2, \ldots, \xi_n$ une suite de variables aléatoires indépendantes issues d'une loi binomiale $B(m,p)$. Le paramètre m est connu. On cherche à estimer le paramètre $0 < p < 1$ au vu de l'échantillon $\xi_1, \xi_2, \ldots, \xi_n$.

1. Trouver le biais et la variance de l'estimateur $\hat{p}_n = \xi_1$.
2. Trouver le biais et la variance de l'estimateur $\hat{p}_n = \xi_1/m$.
3. Trouver le biais et la variance des estimateurs «améliorés» $\hat{p}_n = \overline{\xi}$ et $\hat{p}_n = \overline{\xi}/m$, où $\overline{\xi} = \frac{1}{n}\sum_{i=1}^{n} \xi_i$.

1) Le biais et la variance de l'estimateur $\hat{p}_n = \xi_1$ sont

$$b(p) = \mathbb{E}_p\left(\xi_1\right) - p = mp - p, \ \ \mathrm{var}_p\left(\xi_1\right) = mp(1-p).$$

2) Le biais et la variance de l'estimateur $\hat{p}_n = \xi_1/m$ sont

$$b(p) = \mathbb{E}_p\left(\xi_1/m\right) - p = 0, \quad \mathrm{var}_p\left(\xi_1/m\right) = \frac{p(1-p)}{m}.$$

3) Le biais et la variance de l'estimateur $\hat{p}_n = \overline{\xi}$ sont

$$b(p) = \mathbb{E}_p\left(\overline{\xi}\right) - p = mp - p, \quad \mathrm{var}_p\left(\overline{\xi}\right) = \frac{mp(1-p)}{n}.$$

4) Le biais et la variance de l'estimateur $\hat{p}_n = \overline{\xi}/m$ sont

$$b(p) = \mathbb{E}_p\left(\overline{\xi}\right) - p = 0, \quad \mathrm{var}_p\left(\overline{\xi}\right) = \frac{p(1-p)}{mn}.$$

Sujet 3 [5 points] : Soit $\xi_1, \xi_2, \ldots, \xi_n$ une suite de variables aléatoires indépendantes issues d'une loi binomiale $B(5,p)$. On cherche à estimer le paramètre $0 < p < 1$ à l'aide de l'estimateur suivant :

$$\hat{p}_n = \frac{n+1}{5n}\overline{\xi}, \quad \overline{\xi} = \frac{1}{n}\sum_{i=1}^{n}\xi_i$$

1. Déterminer le biais $b(p)$ et la variance de l'estimateur $\hat{p}_n$.
2. Vérifier les conditions de régularité r1 et r2 du Théorème de Rao-Cramer pour $f_p(x) = \mathbb{P}(\xi = x) = C_5^x p^x (1-p)^{5-x}$ pour $x = 0, 1, 2, 3, 4, 5$ (c'est la loi $B(5,p)$).
3. Si les conditions sont satisfaites, calculer la borne de Rao-Cramer dans la classe d'estimateurs biaisés $\mathcal{K}_{b(p)}$ et répondre à la question : $\hat{p}_n$ est-il efficace dans cette classe ?

1) On calcule le biais

$$b(p) = \mathbb{E}_p\left(\hat{p}_n\right) - p = \frac{n+1}{5n}\mathbb{E}_p\left(\overline{\xi}\right) - p = \frac{n+1}{5n}\mathbb{E}_p\left(\xi_i\right) - p = \frac{n+1}{5n}5p - p = \frac{p}{n}.$$

et la variance de l'estimateur $\hat{p}_n$:

$$\mathrm{var}_p\left(\hat{p}_n\right) = \frac{(n+1)^2}{25n^2} \cdot \frac{\mathrm{var}_p\left(\overline{\xi}\right)}{n} = \frac{(n+1)^2}{25n^2} \cdot \frac{5p(1-p)}{n} = \frac{(n+1)^2 p(1-p)}{5n^3}$$

2) La condition de régularité r1 du Théorème de Rao-Cramer pour $f(x;p) = C_5^x p^x (1-p)^{5-x}$ est satisfaite car la fonction $p \mapsto \sqrt{f(x;p)}$ est dérivable par rapport à p et les dérivées sont continues pour $x = 0, 1, 2, 3, 4, 5$. Pour vérifier la condition de régularité r2 il faut calculer la quantité d'information de Fisher :

$$\mathcal{F}(p) = \mathbb{E}_p\left[\left(\frac{\partial \log f_p(\xi_i;p)}{\partial p}\right)^2\right] = \mathbb{E}_p\left[\left(\frac{\xi_i - 5p}{p(1-p)}\right)^2\right] = \frac{\mathrm{var}_p(\xi_i)}{p^2(1-p)^2} = \frac{5p(1-p)}{p^2(1-p)^2} = \frac{5}{p(1-p)}.$$

La fonction $p \mapsto \mathcal{F}(p) = \frac{5}{p(1-p)}$ est continue, finie et positive pour $0 < p < 1$.

3) L'estimateur $\hat{p}_n$ est efficace dans la classe d'estimateurs biaisés $\mathcal{K}_{b(p)}$ avec $b(p) = \frac{p}{n}$ car sa variance est égale à la borne inférieure de Rao-Cramer

$$\mathrm{var}_p(\hat{p}) \geq \frac{[1 + b'(p)]^2}{n\mathcal{F}(p)} = \frac{(1 + \frac{1}{n})^2}{n\frac{5}{p(1-p)}} = \frac{(n+1)^2 p(1-p)}{5n^3}$$

Sujet 4 [5 points] : Soit $\xi_1, \xi_2, \ldots, \xi_n$ une suite de variables aléatoires indépendantes issues d'une loi exponentielle $\Gamma\left(\frac{2}{\theta+1}, 1\right)$ dont la densité de probabilité est

$$f_\theta(y) = \begin{cases} \frac{2}{\theta+1} \exp\left\{-\frac{2}{\theta+1}\, y\right\} & \text{si} \quad y \geq 0 \\ 0 & \text{si} \quad y < 0 \end{cases},$$

où θ est un paramètre inconnu.

1. Trouver l'estimateur du maximum de vraisemblance (MV) du paramètre θ.
2. Trouver l'estimateur du paramètre θ par la méthode de moments (MM). Utiliser le moment d'order 1.
3. L'estimateur MV est-il sans biais ?
4. L'estimateur MM est-il sans biais ?

1) Tout d'abord, définissons la densité de la loi $\Gamma\left(\frac{2}{\theta+1}, 1\right)$ pour n variables :

$$f_\theta(y_1, \ldots, y_n) = \begin{cases} \frac{2^n}{(\theta+1)^n} \exp\left\{-\frac{2}{\theta+1} \sum_{i=1}^n y_i\right\} & \text{si} \quad y_i \geq 0 \text{ pour } i \in [1; n] \\ 0 & \text{si} \quad \exists i \in [1; n] \text{ t.q. } y_i < 0 \end{cases},$$

La fonction logarithmique de vraisemblance $\theta \mapsto \log f(\xi_1, \ldots, \xi_n; \theta)$ pour les valeurs $\xi_1, \ldots, \xi_n$ données positives est

$$\log f(\xi_1, \ldots, \xi_n; \theta) = n \log 2 - n \log(\theta + 1) - \frac{2}{\theta + 1} \sum_{i=1}^n \xi_i$$

On a

$$\frac{d(\log f(\xi_1, \ldots, \xi_n; \theta))}{d\theta} = -\frac{n}{\theta + 1} + \frac{2}{(\theta + 1)^2} \sum_{i=1}^n \xi_i = 0$$

et on obtient le point critique $\theta_0 = \frac{2}{n} \sum_{i=1}^n \xi_i - 1$. C'est un maximum car

$$\left.\frac{d^2\left(\log f(\xi_1, \ldots, \xi_n; \theta)\right)}{d\theta^2}\right|_{\theta=\theta_0} = -\frac{n}{\left(\frac{2}{n} \sum_{i=1}^n \xi_i\right)^2} < 0$$

donc, l'estimateur MV est

$$\widehat{\theta} = \frac{2}{n} \sum_{i=1}^n \xi_i - 1 = 2\overline{\xi} - 1, \quad \overline{\xi} = \frac{1}{n} \sum_{i=1}^n \xi_i$$

2) L'équation de la méthode de moments (en utilisant la fonction $g : x \mapsto x$, c'est-à-dire le moment d'order 1) est

$$\mathbb{E}(g(\xi)) = \mathbb{E}(\xi) = \frac{\theta + 1}{2}.$$

En remplaçant le moment d'order 1 théorique par le moment empirique $\overline{\xi}$, on obtient l'estimateur MM

$$\widehat{\theta} = \frac{2}{n} \sum_{i=1}^n \xi_i - 1 = 2\overline{\xi} - 1$$

On peut constater que les deux estimateurs sont équivalents MV=MM.

3-4) Le biais des estimateurs MV et MM $\widehat{\theta} = \frac{2}{n} \sum_{i=1}^n \xi_i - 1$ est nul car

$$b(\theta) = \mathbb{E}\left(\widehat{\theta}\right) - \theta = \mathbb{E}\left(\frac{2}{n} \sum_{i=1}^n y_i - 1\right) - \theta = 2\mathbb{E}(\xi) - 1 - \theta = (\theta + 1) - 1 - \theta = 0.$$