EXAMEN (ESTIMATION) : CORRIGÉ SUCCINCT

I.Nikiforov

5 décembre 2002

Exercice 1 (2 points) *Soit ξ une variable aléatoire continue issue d'une loi P dont la fonction de répartition est $F_\xi(x)$ et la densité de probabilité est $f_\xi(x)$. Définir la classe des toutes les variables aléatoires ξ telles que*

$$\mathbb{E}\left(\xi^{2002}\right) = 0.$$

Réponses : Cette classe est définie par les fonctions de répartition $F_\xi(x)$ ou les densités de probabilité $f_\xi(x)$ satisfaisant les équations suivantes

$$\mathbb{E}\left(\xi^{2002}\right) = \int_{-\infty}^{\infty} x^{2002} dF_\xi(x) = \int_{-\infty}^{\infty} x^{2002} f_\xi(x)dx = 0 \;\text{ et }\; \int_{-\infty}^{\infty} f_\xi(x)dx = 1.$$

On peut remarquer que cette classe est très restreinte : les variables aléatoires ξ de cette classe ont toute la «masse» probabiliste concentrée en 0. Donc, elles sont des constantes : $\xi = 0$ avec $\mathbb{P}(\xi = 0) = 1$.

Exercice 2 (9 points) *À partir des observations $y_1, y_2, \ldots, y_n$ et $x_1, x_2, \ldots, x_n$, on retient le modèle non linéaire suivant*

$$y_i = x_i \log\theta + \xi_i, \quad i = 1, 2, \ldots, n,$$

où $\theta > 0$ est le paramètre du modèle et $\xi_i \sim \mathcal{N}(0, \sigma^2)$, $\sigma^2 > 0$, est un bruit normal indépendant. On considère que les valeurs (déterministes) $x_1, x_2, \ldots, x_n$ sont choisies de telle sorte que $\sum_{i=1}^n x_i^2 > 0$.

1. *Donner l'expression de la log-vraisemblance pour le modèle retenu.*
2. *Déterminer l'estimateur $\widehat{\theta}$ du maximum de vraisemblance (MV) du paramètre θ.*
3. *L'estimateur MV $\widehat{\theta}$ est-il biaisé ?*
4. *L'estimateur MV $\widehat{\theta}$ est-il asymptotiquement biaisé ?*
5. *Vérifier les conditions de régularité $\mathsf{r1}$ et $\mathsf{r2}$ du Théorème de Rao-Cramer. Calculer (si les conditions $\mathsf{r1}$ et $\mathsf{r2}$ sont satisfaites) la borne de Rao-Cramer (en précisant la classe d'estimateurs).*

Réponses :

1. En utilisant le fait que
$$y_i - x_i \log\theta = \xi_i \; \left(\sim \mathcal{N}(0, \sigma^2)\right)$$
on trouve logarithme de la fonction de vraisemblance
$$\begin{aligned}\log f(\theta; y_1, \ldots, y_n) &= \log\left\{\frac{1}{\left(\sqrt{2\pi}\right)^n \sigma^n} \prod_{i=1}^n \exp\left[-\frac{(y_i - x_i\log\theta)^2}{2\sigma^2}\right]\right\} \\ &= n\log\frac{1}{\sigma\sqrt{2\pi}} - \frac{1}{2\sigma^2}\sum_{i=1}^n (y_i - x_i\log\theta)^2.\end{aligned}$$

2. Pour déterminer l'estimateur MV de θ il faut maximiser la fonction $\theta \mapsto \log f(\theta; y_1, \ldots, y_n)$.

– Condition nécessaire d'extremum local : pour trouver le (les) point critique $\widehat{\theta}$ on résoudra l'equation normale

$$\frac{\partial \log f(\theta; y_1, \dots, y_n)}{\partial \theta} = \frac{1}{\sigma^2} \sum_{i=1}^{n} (y_i - x_i \log \theta) \frac{x_i}{\theta} = 0.$$

D'où on obtient l'estimateur MV $\widehat{\theta}$ du paramètre θ

$$\widehat{\theta} = \exp \left\{ \frac{\sum_{i=1}^{n} y_i x_i}{\sum_{i=1}^{n} x_i^2} \right\}.$$

– Condition suffisante d'extremum local :

$$\begin{aligned}
\frac{\partial^2}{\partial \theta^2} \{\log f(\theta; y_1, \dots, y_n)\} &= \frac{\partial}{\partial \theta} \left\{ \frac{1}{\sigma^2} \sum_{i=1}^{n} (y_i - x_i \log \theta) \frac{x_i}{\theta} \right\} \\
&= -\frac{1}{\sigma^2 \theta^2} \sum_{i=1}^{n} y_i x_i - \frac{1}{\sigma^2 \theta^2} \sum_{i=1}^{n} x_i^2 + \frac{\log \theta}{\sigma^2 \theta^2} \sum_{i=1}^{n} x_i^2.
\end{aligned}$$

En remplaçant θ par $\widehat{\theta}$ on obtient

$$\frac{\partial^2}{\partial \theta^2} \left\{ \log f(\widehat{\theta}; y_1, \dots, y_n) \right\} = -\frac{1}{\sigma^2 \widehat{\theta}^2} \sum_{i=1}^{n} x_i^2 < 0.$$

Donc, la fonction $\theta \mapsto \log f(\theta; y_1, \dots, y_n)$ admet en $\widehat{\theta}$ un maximum. Maintenant il est clair pourquoi nous avons exigé que $\sum_{i=1}^{n} x_i^2 > 0$.

3. Soit $\zeta = \frac{\sum_{i=1}^{n} y_i x_i}{\sum_{i=1}^{n} x_i^2}$. L'espérance mathématique de ζ est

$$\mathbb{E}(\zeta) = \mathbb{E} \left(\frac{\sum_{i=1}^{n} y_i x_i}{\sum_{i=1}^{n} x_i^2} \right) = \frac{\sum_{i=1}^{n} \mathbb{E}(y_i) x_i}{\sum_{i=1}^{n} x_i^2} = \frac{\sum_{i=1}^{n} x_i^2 \log \theta}{\sum_{i=1}^{n} x_i^2} = \log \theta.$$

Donc, on a $\zeta \sim \mathcal{N}(\log \theta, \sigma_\zeta^2)$. Maintenant on calcule l'espérance de $\widehat{\theta}$:

$$\begin{aligned}
\mathbb{E}(\widehat{\theta}) &= \mathbb{E}(\exp\{\zeta\}) = \int_{-\infty}^{\infty} e^x \frac{1}{\sigma_\zeta \sqrt{2\pi}} e^{-\frac{(x - \log \theta)^2}{2\sigma_\zeta^2}} dx \\
&= \int_{-\infty}^{\infty} \frac{1}{\sigma_\zeta \sqrt{2\pi}} e^{-\frac{x^2 - 2x(\log \theta + \sigma_\zeta^2) + (\log \theta + \sigma_\zeta)^2 - (\log \theta + \sigma_\zeta)^2 + (\log \theta)^2}{2\sigma_\zeta^2}} dx \\
&= e^{\frac{(\log \theta + \sigma_\zeta)^2 - (\log \theta)^2}{2\sigma_\zeta^2}} \overbrace{\int_{-\infty}^{\infty} \dots dx}^{=1} = e^{\log \theta + \sigma_\zeta^2/2} = \theta e^{\sigma_\zeta^2/2} \neq \theta
\end{aligned}$$

Donc, l'estimateur MV $\widehat{\theta}$ est biaisé.

4. La variance σ_ζ^2 de la variable aléatoire ζ est

$$\sigma_\zeta^2 = \mathrm{var}(\zeta) = \mathrm{var} \left(\frac{\sum_{i=1}^{n} y_i x_i}{\sum_{i=1}^{n} x_i^2} \right) = \frac{\sum_{i=1}^{n} \mathrm{var}(y_i) x_i^2}{\left(\sum_{i=1}^{n} x_i^2 \right)^2} = \frac{\sigma^2 \sum_{i=1}^{n} x_i^2}{\left(\sum_{i=1}^{n} x_i^2 \right)^2} = \frac{\sigma^2}{\sum_{i=1}^{n} x_i^2}.$$

Soit $\lim_{n\to\infty}\sum_{i=1}^n x_i^2=\infty$, alors $\sigma_\zeta^2\to 0$. On obtient

$$\lim_{n\to\infty} e^{\log\theta + \sigma_\zeta^2/2} = e^{\log\theta} = \theta$$

et donc l'estimateur MV $\hat\theta$ est asymptotiquement non biaisé. (Par contre, si $\lim_{n\to\infty}\sum_{i=1}^n x_i^2<\infty$, alors l'estimateur MV $\hat\theta$ est même asymptotiquement biaisé !)

5. Soit

$$f(\theta;y_i)=\frac{1}{\sigma\sqrt{2\pi}}\exp\left\{-\frac{(y_i-x_i\log\theta)^2}{2\sigma^2}\right\}.$$

La conditions de régularité $\mathsf{r1}$ du Théorème de Rao-Cramer : la fonction $\theta\mapsto\sqrt{f(\theta;y_i)}$ est dérivable car c'est un produit ou composition des fonctions élémentaires. La dérivée

$$\frac{\partial\sqrt{f(\theta;y_i)}}{\partial\theta}=\frac{1}{2\sqrt{f(\theta;y_i)}}\frac{1}{\sigma\sqrt{2\pi}}\exp\left\{-\frac{(y_i-x_i\log\theta)^2}{2\sigma^2}\right\}\frac{y_i-x_i\log\theta}{\sigma^2}\left(\frac{x_i}{\theta}\right)$$

est continue comme produit ou composition des fonctions élémentaires ($\theta>0$).

La conditions de régularité $\mathsf{r2}$ du Théorème de Rao-Cramer :

$$\mathcal{F}(\theta)=\mathbb{E}_\theta\left[\left(\frac{\partial\log f(\theta;y_1,\ldots,y_n)}{\partial\theta}\right)^2\right]=\frac{1}{\theta^2\sigma^2}\sum_{i=1}^n x_i^2.$$

La fonction $\theta\mapsto\mathcal{F}(\theta)$ existe, strictement positive et continue par rapport à θ (car $\theta>0$ et $\sum_{i=1}^n x_i^2>0$). Soit

$$b(\theta)=\mathbb{E}(\hat\theta)-\theta=\theta\left(e^{\sigma_\zeta^2/2}-1\right),\quad \text{alors}\quad b'(\theta)=e^{\sigma_\zeta^2/2}-1.$$

On obtient la borne de Rao-Cramer dans la classe $\mathcal{K}_b$:

$$\mathbb{E}\left[(\hat\theta-\theta)^2\right]=\frac{\theta^2\sigma^2}{\sum_{i=1}^n x_i^2}\exp\left\{\frac{\sigma^2}{2\sum_{i=1}^n x_i^2}\right\}+\theta^2\left(\exp\left\{\frac{\sigma^2}{2\sum_{i=1}^n x_i^2}\right\}-1\right)^2.$$

Exercice 3 (5 points) *Soit le modèle autorégressif (AR) suivant*

$$y_t=ay_{t-1}+(1-a)m+\xi_t,\quad t\geq 1$$

où $\xi_t\sim\mathcal{N}(0,\sigma_\xi^2)$ *est un bruit normal indépendant et* $y_0\sim\mathcal{N}\left(m,\frac{\sigma_\xi^2}{1-a^2}\right)$. *Ce modèle AR a trois paramètres :* m, $|a|<1$ *et* $\sigma_\xi^2>0$. *Supposons que ces paramètres sont connus.*

1. *Calculer l'espérance mathématique et la variance de* y_t *pour* $t\geq 1$.
2. *On cherche un estimateur de* y_{t+1} *à base de* y_t *(problème d'extrapolation). Pour résoudre ce problème on utilise l'approche bayésienne.*

   Déterminer l'estimateur bayésien $\hat y_{t+1}$ *de* y_{t+1}*, c'est-à-dire l'espérance mathématique conditionnelle de la variables aléatoire* y_{t+1} *par rapport à* y_t.
3. *L'estimateur bayésien* $\hat y_{t+1}$ *est-il biaisé ?*
4. *Calculer l'erreur moyenne quadratique* $\mathbb{E}(\hat y_{t+1}-y_{t+1})^2$.

Réponses :

1. Soit $t = 1$. En utilisant le fait que $y_1 = ay_0 + (1-a)m + \xi_1$ et $y_0 \sim \mathcal{N}\left(m, \frac{\sigma_\xi^2}{1-a^2}\right)$, on obtient $\mathbb{E}(y_1) = a\mathbb{E}(y_0) + (1-a)m + 0$. Donc, $\mathbb{E}(y_1) = m$. Ensuite, pour $t > 1$, on a $\mathbb{E}(y_t) = a\mathbb{E}(y_{t-1}) + (1-a)m$ et finalement $\mathbb{E}(y_t) = m$ pour $t \geq 1$.
   Soit $t = 1$. Alors var $(y_1) = a^2$ var (y_0)+var(ξ_1) (car le bruit est indépendant et donc $\mathrm{cov}(y_{t-1}, \xi_t) = 0$ pour $t \geq 1$). On obtient $\mathrm{var}(y_1) = \frac{\sigma_\xi^2}{1-a^2}$. Ensuite, pour $t > 1$, on a $\mathrm{var}(y_t) = a^2\,\mathrm{var}(y_{t-1}) + \sigma_\xi^2$ et finalement $\mathrm{var}(y_t) = \frac{\sigma_\xi^2}{1-a^2}$ pour $t \geq 1$.
2. L'estimateur bayésien $\hat{y}_{t+1}$ de y_{t+1} est donné par l'équation suivante

$$\hat{y}_{t+1} = \mathbb{E}(y_{t+1}|y_t) = \mathbb{E}(ay_t + (1-a)m + \xi_{t+1}|y_t) = ay_t + (1-a)m + \overbrace{\mathbb{E}(\xi_{t+1}|y_t)}^{=0,\ \text{car le bruit est indépendant}} = ay_t + (1-a)m.$$

3. $$\mathbb{E}(y_{t+1} - \hat{y}_{t+1}) = \mathbb{E}(y_{t+1} - ay_t + (1-a)m) = \mathbb{E}(\xi_{t+1}) = 0.$$

   Donc l'estimateur bayésien $\hat{y}_{t+1}$ de y_{t+1} est non biaisé.
4. D'après la question 3 l'erreur moyenne quadratique est $\mathbb{E}(\hat{y}_{t+1} - y_{t+1})^2 = \mathbb{E}(\xi_{t+1})^2 = \sigma_\xi^2$.

Exercice 4 (4 **points)** *Soit* $\xi_1, \ldots, \xi_n$ n *variables aléatoires indépendantes. La variable aléatoire* ξ_i *est issue de la loi* P_θ *dont la densité de probabilité est*

$$f_\theta(y) = \begin{cases} \frac{3y^2}{8\theta^3} & si \quad y \in [0, 2\theta] \\ 0 & si \quad y \notin [0; 2\theta] \end{cases} \qquad \theta \in]0; \infty[,$$

où θ *est un paramètre inconnu.*

1. *Trouver un estimateur du paramètre* θ *par la méthode de moments. Utiliser le moment d'ordre* 1.
2. *L'estimateur par la méthode de moments est-il sans biais ?*
3. *Est-il asymptotiquement normal ?*

Réponses :

1. $$\mathbb{E}(\xi) = \int_0^{2\theta} y\frac{3y^2}{8\theta^3}dy = \left.\frac{3y^4}{32\theta^3}\right|_0^{2\theta} = \frac{3}{2}\theta.$$

   L'estimateur du paramètre θ par la méthode de moments est :

$$\hat{\theta} = \frac{2}{3n}\sum_{i=1}^{n}\xi_i$$

2. $$\mathbb{E}(\hat{\theta}) = \mathbb{E}\left(\frac{2}{3n}\sum_{i=1}^{n}\xi_i\right) = \theta.$$

3. L'estimateur $\hat{\theta}$ est asymptotiquement normal puisque la fonction $u \mapsto 2/3u$ est dérivable, $0 < 2/3 < \infty$ et le moment d'ordre 2 est fini :

$$\mathbb{E}(\xi^2) = \int_0^{2\theta} y^2\frac{3y^2}{8\theta^3}dy = \left.\frac{3y^5}{40\theta^3}\right|_0^{2\theta} = \frac{12}{5}\theta^2 < \infty.$$