MAP433 Statistique
PC4: méthodes d'estimation

18 septembre 2015

1. Paramètre vectoriel - vitesses de convergence différentes

Soient $X_1, \ldots, X_n$ des variables aléatoires i.i.d. de loi exponentielle translatée dont le densité est de la forme :

$$f(x, \theta, \alpha) = \frac{1}{\theta} \exp\left[-\frac{(x-\alpha)}{\theta}\right] I(x \geq \alpha),$$

où $\theta > 0$ et $\alpha \in \mathbb{R}$ sont deux paramètres inconnus.

1. Donner les estimateurs du maximum de vraisemblance $\hat{\alpha}_n$ et $\hat{\theta}_n$ de α et θ.
2. Quelle est la loi de $X_i - \alpha$? Calculer la loi (exacte) de $n(\hat{\alpha}_n - \alpha)$.
3. Déterminer la loi limite de $\sqrt{n}(\hat{\theta}_n - \theta)$.

Corrigé :

1. La fonction de vraisemblance est donnée par

$$\theta \mapsto \begin{cases} \theta^{-n} \exp\left(-\sum_{k=1}^n (X_k - \alpha)/\theta\right) & \text{si } \alpha \leq \inf_k X_k = X_{(1)} \\ 0 & \text{sinon} \end{cases}$$

On en déduit que l'estimateur du maximum de vraisemblance de $\psi = (\alpha, \theta)$ est unique et est égal à

$$\widehat{\psi}_n = \begin{bmatrix} X_{(1)} \\ \bar{X}_n - X_{(1)} \end{bmatrix}$$

où $X_{(1)} = \min_{1 \leq k \leq n} X_k$.

2. $X_i - \alpha \sim \mathcal{E}(1/\theta)$. De plus, $\hat{\alpha}_n - \alpha = \min_{1 \leq k \leq n}(X_k - \alpha)$ dont on déduit que $\hat{\alpha}_n - \alpha \sim \mathcal{E}(n/\theta)$. Par suite, pour toute fonction f continue bornée, on a

$$\mathbb{E}\left[f(n(\hat{\alpha}_n - \alpha))\right] = \int_0^\infty f(ny)\frac{n}{\theta}\ \exp(-ny/\theta)\ dy = \int_0^\infty f(x)\frac{1}{\theta}\ \exp(-x/\theta)\ dx$$

dont on déduit que

$$n\left(\hat{\alpha}_n - \alpha\right) \sim \mathcal{E}(1/\theta).$$

3. On écrit

$$\hat{\theta}_n - \theta = \bar{X}_n - \hat{\alpha}_n - \theta = \bar{X}_n - (\theta + \alpha) + (\alpha - \hat{\alpha}_n).$$

On a

$$\sqrt{n}\left(\bar{X}_n - (\theta+\alpha)\right) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \theta^2\right) \qquad \sqrt{n}\left(\hat{\alpha}_n - \alpha\right) \xrightarrow{\mathrm{P}} 0$$

dont on déduit par Slutsky que $\sqrt{n}\left(\hat{\theta}_n - \theta\right) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \theta^2)$.

2. Information de Fisher : entraînement !

Dans les modèles suivants, calculer l'information de Fisher associée aux n observations (si elle est bien définie), l'estimateur du maximum de vraisemblance et sa loi asymptotique :

1. $X_1, \ldots, X_n \overset{i.i.d}{\sim} \mathcal{B}(\theta)$.
2. $X_1, \ldots, X_n \overset{i.i.d}{\sim} \mathcal{N}(m, v)$.
3. $X_1, \ldots, X_n \overset{i.i.d}{\sim} \mathcal{U}[0, \theta]$.

Corrigé :

1. L'information de Fisher est $I_n(\theta) = n\theta^{-1}(1-\theta)^{-1}$. L'estimateur de MV est $\bar{X}_n$ et on a
$$\sqrt{n}\left(\bar{X}_n - \theta\right) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \theta(1-\theta))$$
2. Attention, dans ce corrigé, v désigne l'écart-type et v^2 la variance. L'information de Fisher est
$$I_n(\theta) = \frac{n}{v^4}\begin{bmatrix} v^2 & 0 \\ 0 & 1/2 \end{bmatrix}$$
L'estimateur MV est unique et vaut
$$\hat{\theta}_n^{MV} = \begin{bmatrix} \bar{X}_n \\ \frac{1}{n}\sum_{k=1}^n \left(X_k - \bar{X}_n\right)^2 \end{bmatrix}$$
On établit un TCl pour le couple $(\bar{X}_n, n^{-1}\sum_{k=1}^n (X_k - m)^2)$; puis on applique la méthode delta avec $g(x,y) = (x; y - (x-m)^2)$. On obtient
$$\sqrt{n}\left(\begin{bmatrix} \bar{X}_n \\ \frac{1}{n}\sum_{k=1}^n (X_k - \bar{X}_n)^2 \end{bmatrix} - \begin{bmatrix} m \\ v^2 \end{bmatrix}\right) \xrightarrow{\mathcal{L}} \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}; \begin{bmatrix} v^2 & 0 \\ 0 & 2v^4 \end{bmatrix}\right)$$
3. Ce n'est pas un modèle statistique régulier. L'estimateur MV est donné par
$$\hat{\theta}_n^{MV} = \max_{1 \le k \le n} X_k = X_{(n)}$$
On a établi (voir PC1) que la loi de $X_{(n)}$ admettait $F(t)^n$ comme fonction de répartition, en notant F la fonction de répartition de X_1. On en déduit la densité de $X_{(n)}$ puis celle de $n(X_{(n)} - \theta)$; et par le lemme de Scheffé, on montre que
$$n\left(\hat{\theta}_n^{MV} - \theta\right) \xrightarrow{\mathcal{L}} \frac{1}{\theta}\exp(x/\theta)\mathbb{1}_{x \le 0}$$

3. Borne de Cramer-Rao

On considère un vecteur aléatoire $(X_1, \ldots, X_n) \in \mathbb{R}^n$ de loi appartenant à une famille $\{\mathbb{P}_\theta,\ \theta \in \Theta\}$ de lois sur $\mathbb{R}^n$, avec Θ intervalle ouvert de $\mathbb{R}$. On suppose que $d\mathbb{P}_\theta(x) = p(\theta, x)\, d\mu(x)$ avec μ mesure σ-finie sur $\mathbb{R}^n$ et on note $l_X(\theta) = \log(p(\theta, X))$. On suppose que la famille de lois $\{\mathbb{P}_\theta,\ \theta \in \Theta\}$ est régulière ; l'information de Fisher $I_n(\theta)$ est donc bien définie et on pourra intervertir intégrales et dérivations à notre de guise.

Pour un estimateur $\hat{\theta}$ donné, on note $R_\theta(\hat{\theta}) := \mathbb{E}_\theta\left[(\hat{\theta} - \theta)^2\right]$ son risque quadratique et $b(\theta) := \mathbb{E}_\theta[\hat{\theta}] - \theta$ son biais (qu'on suppose dérivable).

1. Montrez que $R_\theta(\hat{\theta}) = b(\theta)^2 + \mathrm{Var}_\theta(\hat{\theta})$.
2. Montrez que $\mathbb{E}_\theta\left[l'_X(\theta)\right] = 0$.
3. Montrez que $b'(\theta) = \mathbb{E}_\theta\left[\hat{\theta} l'_X(\theta)\right] - 1$.
4. Déduire des deux questions précédentes l'égalité $1 + b'(\theta) = \mathbb{E}_\theta\left[(\hat{\theta} - \mathbb{E}_\theta(\hat{\theta})) l'_X(\theta)\right]$.
5. En déduire la borne de Cramer-Rao :
$$\boxed{R_\theta(\hat{\theta}) \ge \frac{(1 + b'(\theta))^2}{I_n(\theta)} + b(\theta)^2.}$$

6. Quel est le risque quadratique minimal d'un estimateur sans biais ?

Corrigé :

1. On écrit $\hat{\theta} - \theta = \hat{\theta} - \mathbb{E}\left[\hat{\theta}\right] + \mathbb{E}\left[\hat{\theta}\right] - \theta$ puis on développe le carré et on applique l'espérance.
2. Conséquence de la permutation dérivée/intégrale et de la propriété $\int p(x;\theta)\mu(dx) = 1$.
3. Conséquence de la permutation dérivée/intégrale et du fait que $\partial_\theta p(x;\theta) = p(x;\theta)\ \partial_\theta \ln p(x;\theta)$.
4. Conséquence des questions 2 et 3.
5. En utilisant la question 1, il suffit de minorer la variance. Ce que l'on fait en utilisant l'inégalité de Hölder et la question 4.

4. Phénomène de Stein

On considère le modèle

$$Y_j = \theta_j + \xi_j,\ j = 1, \ldots, d,$$

avec les ξ_j iid gaussiennes centrées de variance 1. On pose $Y = (Y_1, \ldots, Y_d)$ et $\theta = (\theta_1, \ldots, \theta_d)$. On s'intéresse à l'estimation de θ et on suppose $d \geq 3$.

Definition : Un estimateur θ^* de θ est dit admissible sur $\Theta \subset \mathbb{R}^d$ par rapport au risque quadratique s'il n'existe pas d'estimateur $\hat{\theta}$ tel que pour tout $\theta \in \Theta$

$$\mathbb{E}_\theta[\|\hat{\theta} - \theta\|^2] \leq \mathbb{E}_\theta[\|\theta^* - \theta\|^2]\,,$$

avec inégalité stricte en au moins un $\theta_0 \in \Theta$.

Lemme (admis) Soit $d \geq 3$. Pour tout $\theta \in \mathbb{R}^d$, on a $0 < \mathbb{E}_\theta[\|Y\|^{-2}] < \infty$.

1. Donner l'estimateur intuitif de θ. Calculer son risque quadratique.
2. Soit $\xi = (\xi_1, \ldots, \xi_d)$. Soit $f : \mathbb{R}^d \to \mathbb{R}$ telle que

   - pour presque tout $(x_2, \ldots, x_d)$, la fonction $x_1 \to f(x_1, \ldots, x_d)$ est dérivable et $\lim_{|x_1|\to\infty} f(x_1, \ldots, x_d)e^{-x_1^2/2} = 0$,

   - $\mathbb{E}[|\frac{\partial f}{\partial x_1}(\xi)|] < +\infty$, $\mathbb{E}[|\xi_1 f(\xi)|] < +\infty$.

   Montrer que
   $$\mathbb{E}\left[\frac{\partial f}{\partial x_1}(\xi)\right] = \mathbb{E}[\xi_1\ f(\xi)].$$
3. Montrer que si $\tilde{f} : \mathbb{R}^d \to \mathbb{R}$ vérifie

   - $\tilde{f}(y_1, \ldots, y_d)$ admet des dérivées partielles par rapport à chaque composante pour presque toutes les valeurs des autres composantes.

   - $\lim_{|y_i|\to\infty} \tilde{f}(y_1, \ldots, y_d)e^{-(y_i-\theta_i)^2/2} = 0$ pour presque tout $(y_1, \ldots, y_{i-1}, y_{i+1}, \ldots, y_n)$, et tout $i = 1, \ldots, d$,

   - $\mathbb{E}_\theta[|\frac{\partial \tilde{f}}{\partial y_i}(Y)|] < +\infty,\ \mathbb{E}_\theta[|(Y_i - \theta_i)\ \tilde{f}(Y)|] < +\infty,\ i = 1, \ldots, d,$

   alors
   $$\mathbb{E}_\theta\left[(Y_i - \theta_i)\tilde{f}(Y)\right] = \mathbb{E}_\theta\left[\frac{\partial \tilde{f}}{\partial y_i}(Y)\right],\ i = 1, \ldots, d$$
4. Soit $g : \mathbb{R}^d \to \mathbb{R}$ telle que les conditions de la question précédente sont vérifiées par les $\tilde{f}_i(y) = (1 - g(y))y_i,\ i = 1, \ldots, d$. On considère un estimateur de la forme $\hat{\theta} = g(Y)Y$ (i.e. $\hat{\theta}_j = g(Y)Y_j$). Montrer que
   $$\mathbb{E}_\theta[\|\hat{\theta} - \theta\|^2] = d + \mathbb{E}_\theta[W(Y)],$$
   avec
   $$W(Y) = -2d(1 - g(Y)) + 2\sum_{i=1}^{d} Y_i \frac{\partial g}{\partial y_i}(Y) + \|Y\|^2(1 - g(Y))^2.$$

5. Soit $g(y)$ de la forme $1 - \frac{c}{\|y\|^2}$. Dans ce cas les $\tilde{f}_i(y) = (1 - g(y))y_i$ vérifient les hypothèses de la question 4. Trouver c tel que $\mathbb{E}_\theta[W(Y)] < 0$.
6. L'estimateur intuitif est-il admissible ?

Corrigé :

1. Estimateur intuitif : $\hat{\theta}^\star = Y$. Risque quadratique : $R_\theta(\hat{\theta}^\star) = d$.
2. Sous les hypothèses, on peut appliquer Fubini pour montrer que

$$\mathbb{E}\left[\xi_1 \ f(\xi)\right] = \frac{1}{\sqrt{2\pi}^d} \int_{\mathbb{R}^{d-1}} \left(\int_{\mathbb{R}} z_1 f(z_1, z_{2:d}) \exp(-z_1^2/2) dz_1 \right) \ \exp(-\sum_{k=2}^d z_k^2/2) \ dz_{2:d}$$

$$\mathbb{E}\left[\partial_1 f(\xi)\right] = \frac{1}{\sqrt{2\pi}^{d-1}} \int_{\mathbb{R}^{d-1}} \mathbb{E}\left[\partial_1 f(\xi_1, z_2, \cdots, z_d)\right] \ \prod_{k=2}^d \exp(-z_k^2/2) dz_k.$$

Par une intégration par parties, on montre

$$\mathbb{E}\left[\partial_1 f(\xi_1, z_2, \cdots, z_d)\right] = \int_{\mathbb{R}} z_1 f(z_1, z_{2:d}) \exp(-z_1^2/2) dz_1$$

ce qui conclut la démonstration.
3. On applique la question précédente avec $f = \tilde{f}(\cdot + \theta)$ et on utilise le fait que sous $\mathbb{P}_\theta$, Y_i a même loi que $\theta_i + \xi_i$.
4. On écrit $\|\hat{\theta} - \theta\|^2 = \left\|(g(Y)\, Y - Y) + (Y - \theta)\right\|^2$; on développe le carré puis prend l'espérance. On obtient le résultat en utilisant la question précédente et en notant que $\mathbb{E}_\theta\left[\|Y - \theta\|^2\right] = d$.
5. Avec ce choix de g, on a

$$\mathbb{E}_\theta\left[W(Y)\right] = \mathbb{E}_\theta\left[-\frac{2dc}{\|Y\|^2} + \frac{4c}{\|Y\|^2} + \frac{c^2}{\|Y\|^2} \right] = \left(c^2 + 4c - 2dc\right) \mathbb{E}_\theta\left[\|Y\|^{-2}\right].$$

Cette quantité est minimale (atteignant une valeur négative) pour $c = d - 2$.
6. Non.