MAP433 Statistique
PC4: métodes d'estimation
Correction

19 septembre 2014

4. La statistique d'ordre

Soient $X_1, \ldots, X_n$ des variables aléatoires réelles i.i.d. de fonction de répartition F. On suppose que F admet une densité f par rapport à la mesure de Lebesgue. On note $X_{(1)} \leq X_{(2)} \leq \ldots \leq X_{(n)}$ les variables aléatoires $X_1, \ldots, X_n$ réordonnées par ordre croissant.

1. **Déterminer la fonction de répartition $F_k(x)$ puis la densité $f_k(x)$ de $X_{(k)}$.**
   Comme $\mathbb{P}(X_i \leq t) = F(t)$ la variable aléatoire

$$S_n(t) = \sum_{i=1}^{n} \mathbf{1}_{\{X_i \leq t\}}$$

suit une loi binomiale de paramètre $(n, F(t))$. En conséquence

$$F_k(t) = \mathbb{P}(X_{(k)} \leq t) = \mathbb{P}(S_n(t) \geq k) = \sum_{j=k}^{n} \binom{n}{j} \ (F(t))^j \ (1 - F(t))^{n-j} .$$

- Dans le cas $k = n$ (resp. $k = 1$) on retrouve que la fonction de répartition est resp.

$$F_n(t) = F(t)^n \qquad F_1(t) = 1 - (1 - F(t))^n$$

- Par dérivation, on obtient la densité f_k satisfaisant

$$
\begin{aligned}
\frac{f_k(t)}{f(t)} &= \sum_{j=k}^{n} \tbinom{n}{j} j \,(F(t))^{j-1}\ (1-F(t))^{n-j} - \sum_{j=k}^{n} \tbinom{n}{j}\ (n-j)\,(F(t))^{j}\ (1-F(t))^{n-j-1} \\
&= \sum_{j=k}^{n} \frac{n!}{(j-1)!(n-1-(j-1))!} (F(t))^{j-1}\ (1-F(t))^{n-1-(j-1)} \\
&- \sum_{j=k}^{n-1} \frac{n!}{j!(n-1-j)!}\ (F(t))^{j}\ (1-F(t))^{n-j-1} \\
&= \sum_{j=k-1}^{n-1} \frac{n(n-1)!}{j!(n-1-j)!} (F(t))^{j}\ (1-F(t))^{n-1-j} \\
&- \sum_{j=k}^{n-1} \frac{n(n-1)!}{j!(n-1-j)!}\ (F(t))^{j}\ (1-F(t))^{n-1-j} \\
&= \frac{n!}{(k-1)!(n-1-(k-1))!} F(t)^{k-1}\ (1-F(t))^{n-k}
\end{aligned}
$$

i.e.

$$f_k(t) = n\tbinom{n-1}{k-1} F(t)^{k-1}\ (1-F(t))^{n-k}\ f(t)$$

Dans le cas de v.a. uniformes sur $[0,1]$, f_k est donnée par

$$\frac{1}{\theta^n}\tbinom{n-1}{k-1} t^{k-1}(\theta-t)^{n-k} 1_{[0,\theta]}(t)$$

C'est une Loi Beta de paramètres (k ; n-k+1).

2. **Donner l'expression de la loi de la statistique d'ordre $(X_{(1)},\ldots,X_{(n)})$ en fonction de f.**
Notons $\mathcal{T}$ l'ensemble des permutations d'entiers $\{1,\cdots,n\}$, de cardinal $n!$. Pour tout n-uplet $(X_1,\cdots,X_n)$ il existe une unique permutation S (aléatoire) telle que

$$X_{(i)} = X_{S(i)} \qquad \forall i$$

Soit h une fonction mesurable bornée définie sur $\mathbb{R}^n$. On a

$$
\begin{aligned}
\mathbb{E}\left[h\left(X_{(1)},\cdots,X_{(n)}\right)\right] &= \sum_{\sigma\in\mathcal{T}} \mathbb{E}\left[h\left(X_{(1)},\cdots,X_{(n)}\right) 1_{S=\sigma}\right] \\
&= \sum_{\sigma\in\mathcal{T}} \mathbb{E}\left[h\left(X_{\sigma(1)},\cdots,X_{\sigma(n)}\right) 1_{S=\sigma}\right]
\end{aligned}
$$

Or, pour toute permutation déterministe σ, la loi de $(X_{\sigma(1)},\cdots,X_{\sigma(n)})$ est $\prod_{k=1}^{n} f(x_{\sigma(k)})$ puisque les composantes sont i.i.d. De plus,

$$1_{S=\sigma} = 1_{X_{\sigma(1)}<\cdots<X_{\sigma(n)}}.$$

Par suite

$$\begin{aligned}\mathbb{E}\left[h\left(X_{(1)},\cdots,X_{(n)}\right)\right] &= \sum_{\sigma\in\mathcal{T}}\int_{\mathbb{R}^n} h\left(x_{\sigma(1)},\cdots,x_{\sigma(n)}\right)\,\mathbb{1}_{x_{\sigma(1)}<\cdots<x_{\sigma(n)}}\prod_{k=1}^n f(x_{\sigma(k)})dx_{(k)}\\ &= \sum_{\sigma\in\mathcal{T}}\int_{\mathbb{R}^n} h\left(y_1,\cdots,y_n\right)\,\mathbb{1}_{y_1<\cdots<y_n}\prod_{k=1}^n f(y_k)dy_k\\ &= n!\int_{\mathbb{R}^n} h\left(y_1,\cdots,y_n\right)\,\mathbb{1}_{y_1<\cdots<y_n}\prod_{k=1}^n f(y_k)dy_k.\end{aligned}$$

On en déduit que la loi de la statistique d'ordre possède une densité donnée par

$$\bar{f}(y_1,\cdots,y_n) = n!\ \mathbb{1}_{y_1<\cdots<y_n}\prod_{k=1}^n f(y_k).$$

3. **Sans utiliser les résultats des questions précédentes, calculer les fonctions de répartition de $X_{(1)}$, $X_{(n)}$, du couple $(X_{(1)},X_{(n)})$ et la loi de la statistique $W=X_{(n)}-X_{(1)}$ (on appelle W *étendue*). Les variables $X_{(1)}$ et $X_{(n)}$ sont elles indépendantes ?**

• Fonction de répartition de $X_{(1)}$ et de $X_{(n)}$

$$\begin{aligned}F_1(t) &= \mathbb{P}(X_{(1)}\le t) = 1-\mathbb{P}(X_{(1)}>t)\\ &= 1-\mathbb{P}(X_1>t,X_2>t,\cdots,X_n>t) = 1-(1-F(t))^n.\\ F_n(t) &= \mathbb{P}(X_{(n)}\le t) = \mathbb{P}(X_1\le t,\cdots,X_n\le t) = F(t)^n.\end{aligned}$$

• Loi du couple

$$\begin{aligned}\mathbb{P}(X_{(1)}\le x,X_{(n)}\le y) &= \mathbb{P}(X_{(n)}\le y)-\mathbb{P}(X_{(1)}>x,X_{(n)}\le y)\\ &= F(y)^n-\mathbb{P}\left(\bigcap_{j=1}^n\{X_j\in]x,y]\}\right)\\ &= F(y)^n-\left(F(y)-F(x)\right)^n\mathbb{1}_{x\le y}\end{aligned}$$

Les v.a. ne sont pas indépendantes puisque $F_1(x)F_n(y)\neq\mathbb{P}(X_{(1)}\le x,X_{(n)}\le y)$.

• Loi de l'étendue $W=X_{(n)}-X_{(1)}$ On déduit de la question précédente que le couple $(X_{(1)},X_{(n)})$ a pour densité jointe

$$n(n-1)f(y)f(x)\left(F(y)-F(x)\right)^{n-2}\mathbb{1}_{x\le y}$$

dont on en déduit la loi de l'étendue

$$\begin{aligned}\mathbb{E}\left[h(W)\right] &= \int_{\mathbb{R}^2} h(y-x)n(n-1)f(y)f(x)\left(F(y)-F(x)\right)^{n-2}\mathbb{1}_{x\le y}dxdy\\ &= \int_{\mathbb{R}^+} h(w)\left(\int_{\mathbb{R}} n(n-1)f(w+x)f(x)\left(F(w+x)-F(x)\right)^{n-2}dx\right)dw\end{aligned}$$

La densité de W est donc donnée par

$$w\mapsto\mathbb{1}_{\mathbb{R}^+}(w)\ n(n-1)\ \int_{\mathbb{R}} f(w+x)f(x)\ \left(F(w+x)-F(x)\right)^{n-2}\ dx.$$

Dans le cas de v.a. uniformes sur $[0,1]$, cette denseté vaut

$$n(n-1)w^{n-2}(1-w)1_{[0,1]}(w).$$

C'est une Loi Beta de paramètres (n-1 ; 2).

4. **Nous supposons dans la suite que les $X_1, \cdots, X_n$ suivent une loi exponentielle de paramètre 1. On note $Y_j = (n-j+1)(X_{(j)} - X_{(j-1)})$ pour $j = 1, \cdots, n$. Quelle est la loi de $(Y_1; \cdots, Y_n)$.**
Cherchons la loi de $(X_{(1)}, \cdots, X_{(n)})$ dans ce cas exponentiel. D'après la question 2, elle est donnée par

$$\bar{f}(z_1, \cdots, z_n) = n!\ 1_{z_1 < \cdots < z_n} \prod_{k=1}^{n} f(z_k) = n!\ 1_{0 < z_1 < \cdots < z_n} \exp(-\sum_{k=1}^{n} z_k)$$

Or

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \cdots \\ Y_j \\ \cdots \\ Y_n \end{bmatrix} = \begin{bmatrix} nX_{(1)} \\ (n-1)(X_{(2)} - X_{(1)}) \\ \cdots \\ (n-j+1)(X_{(j)} - X_{(j-1)}) \\ \cdots \\ X_{(n)} - X_{(n-1)} \end{bmatrix} = A \begin{bmatrix} X_{(1)} \\ X_{(2)} \\ \cdots \\ X_{(j)} \\ \cdots \\ X_{(n)} \end{bmatrix}$$

avec

$$A = \begin{bmatrix} n & 0 & \cdots & & \cdots \\ -(n-1) & (n-1) & 0 & \cdots & \\ 0 & -(n-2) & (n-2) & 0 & \cdots \\ \cdots & & & & \\ \cdots & 0 & -2 & 2 & 0 \\ & \cdots & 0 & -1 & 1 \end{bmatrix}$$

Par la formule du changement de variable, en observant que le déterminant de A est $n!$, et que $\sum_{k=1}^{n} z_k = \sum_{k=1}^{n} y_k$ on obtient

$$f_{(Y_{1:n})}(y_1, \cdots, y_n) = \prod_{k=1}^{n} 1_{\mathbb{R}^+}(y_k) \exp(-\sum_{k=1}^{n} y_k)$$

i.e. les v.a. $(Y_k)_k$ sont i.i.d. exponentielles.

5. **Montrer que $X_{(1)}$ est indépendant de $\bar{X}_n - X_{(1)}$**
On remarque que

$$Y_2 + \cdots + Y_n = \sum_{k=2}^{n} X_{(k)} - (n-1)X_{(1)} = \sum_{k=1}^{n} X_{(k)} - nX_{(1)} = n\left(\bar{X}_n - X_{(1)}\right)$$

Or, la question précédente entraine que Y_1 est indépendant de $(Y_2, \cdots, Y_n)$ donc $Y_1 = nX_{(1)}$ est indépendant de $\sum_{k=2}^{n} Y_k = n\left(\bar{X}_n - X_{(1)}\right)$. Ce qui conclut la démonstration.

5. Borne de Cramer-Rao

On considère un vecteur aléatoire $X = (X_1, \cdots, X_n) \in \mathbb{R}^n$ de loi appartenant à une famille $\{\mathbb{P}_\theta, \theta \in \Theta\}$ de lois sur $\mathbb{R}^n$, avec Θ intervalle ouvert de $\mathbb{R}$. Soit le modèle statistique dominé $\{\mathbb{P}_\theta, \theta \in \Theta\}$ où $d\mathbb{P}_\theta(x) = f(X;\theta)\mu(dx)$ et μ est une mesure σ-finie sur $\mathbb{R}^n$. On suppose que la famille de lois $\{\mathbb{P}_\theta, \theta \in \Theta\}$ est régulière ; l'information de Fisher, notée $I_n(\theta)$, est donc bien définie et on pourra intervertir intégrales et dérivations à notre guise. Pour un estimateur $\hat{\theta}$ donné, on note $R(\theta, \hat{\theta}) := \mathbb{E}_\theta[(\hat{\theta}(X) - \theta)^2]$ son risque quadratique et $b(\theta) := \mathbb{E}_\theta[\hat{\theta}(X)] - \theta$ son biais (qu'on suppose dérivable).

1. **Montrer que $R(\theta, \hat{\theta}) = b(\theta)^2 + \mathrm{Var}_\theta(\hat{\theta}(X))$.**

$$
\begin{aligned}
R(\theta, \hat{\theta}) &= \mathbb{E}_\theta\left[(\hat{\theta} - \theta)^2\right] \\
&= \mathbb{E}_\theta\left[(\hat{\theta} - \mathbb{E}_\theta[\hat{\theta}(X)] + \mathbb{E}_\theta[\hat{\theta}(X)] - \theta)^2\right] \\
&= \mathbb{E}_\theta\left[(\hat{\theta} - \mathbb{E}_\theta[\hat{\theta}(X)])^2\right] + b(\theta)^2 - 2b(\theta)\ \mathbb{E}_\theta\left[\hat{\theta} - \mathbb{E}_\theta[\hat{\theta}(X)]\right] \\
&= \mathrm{Var}_\theta(\hat{\theta}(X)) - b(\theta)^2 + 2b(\theta)^2
\end{aligned}
$$

2. **Montrer que $\mathbb{E}_\theta\left[\partial_\theta \ln f(X;\theta)\right] = 0$**
Puisque le modèle est régulier, on peut permuter intégrale et dérivation. Ce qui donne

$$
\begin{aligned}
\mathbb{E}_\theta\left[\partial_\theta \ln f(X;\theta)\right] &= \int_{\mathbb{R}^n} \partial_\theta \ln f(x;\theta)\, f(x;\theta)\mu(dx) \\
&= \int_{\mathbb{R}^n} \partial_\theta f(x;\theta)\mu(dx) \\
&= \partial_\theta \int_{\mathbb{R}^n} f(x;\theta)\mu(dx) = 0, \quad \text{car} \int_{\mathbb{R}^n} f(x;\theta)\mu(dx) = 1.
\end{aligned}
$$

3. **Montrer que $b'(\theta) = \mathbb{E}_\theta\left[\hat{\theta}(X)\ \partial_\theta f(X;\theta)\right] - 1$.** Puisque l'on peut permuter intégrale et dérivation, on a

$$
\begin{aligned}
b'(\theta) &= \partial_\theta\left(\int_{\mathbb{R}^n} \hat{\theta}(x) f(x;\theta)\mu(dx) - \theta\right) \\
&= \int_{\mathbb{R}^n} \hat{\theta}(x)\ \partial_\theta f(x;\theta)\ \mu(dx) - 1 \\
&= \int_{\mathbb{R}^n} \hat{\theta}(x)\ \partial_\theta \ln f(x;\theta)\ f(x;\theta)\, \mu(dx) - 1 \\
&= \mathbb{E}_\theta\left[\hat{\theta}(X)\ \partial_\theta \ln f(X;\theta)\right] - 1.
\end{aligned}
$$

4. **Déduire des deux questions précédentes l'égalité** $1 + b'(\theta) = \mathbb{E}_\theta\left[\left(\hat{\theta}(X) - \mathbb{E}_\theta\left[\hat{\theta}\right]\right)\ \partial_\theta \ln f(X;\theta)\right]$
D'après la question 3

$$
\begin{aligned}
1 + b'(\theta) &= \mathbb{E}_\theta\left[\hat{\theta}(X)\ \partial_\theta \ln f(X;\theta)\right] \\
&= \mathbb{E}_\theta\left[\left(\hat{\theta}(X) - \mathbb{E}_\theta\left[\hat{\theta}(X)\right]\right)\ \partial_\theta \ln f(X;\theta)\right] + \mathbb{E}_\theta\left[\hat{\theta}(X)\right] \mathbb{E}_\theta\left[\partial_\theta \ln f(X;\theta)\right] \\
&= \mathbb{E}_\theta\left[\left(\hat{\theta}(X) - \mathbb{E}_\theta\left[\hat{\theta}(X)\right]\right)\ \partial_\theta \ln f(X;\theta)\right] + \mathbb{E}_\theta\left[\hat{\theta}\right] \times 0
\end{aligned}
$$

en utilisant la question 2 dans la dernière égalité.

5. **En déduire la borne de Cramer-Rao :**

$$R(\theta,\hat{\theta}) \geq b(\theta)^2 + \frac{(1+b'(\theta))^2}{I_n(\theta)}$$

En utilisant la question 1, il suffit de minorer la variance de $\hat{\theta}$. D'une part,

$$\mathrm{Var}_\theta(\hat{\theta}) = \mathbb{E}_\theta\left[\left(\hat{\theta}(X) - \mathbb{E}_\theta\left[\hat{\theta}(X)\right]\right)^2\right].$$

D'autre part, en utilisant la question 4 et l'inégalité de Cauchy–Schwartz il vient

$$\begin{aligned}(1+b'(\theta))^2 &= \left(\mathbb{E}_\theta\left[\left(\hat{\theta}(X) - \mathbb{E}_\theta\left[\hat{\theta}(X)\right]\right)\ \partial_\theta \ln f(X;\theta)\right]\right)^2 \\ &\leq \mathbb{E}_\theta\left[\left(\hat{\theta}(X) - \mathbb{E}_\theta\left[\hat{\theta}(X)\right]\right)^2\right] \mathbb{E}_\theta\left[(\partial_\theta \ln f(X;\theta))^2\right] \\ &\leq \mathbb{E}_\theta\left[\left(\hat{\theta}(X) - \mathbb{E}_\theta\left[\hat{\theta}(X)\right]\right)^2\right]\ I_n(\theta)\end{aligned}$$

Ces deux relations donnent

$$\mathrm{Var}_\theta(\hat{\theta}) \geq \frac{(1+b'(\theta))^2}{I_n(\theta)}$$

ce qui conclut la démonstration.

6. **Quel est le risque quadratique minimal d'un estimateur sans biais ?**
Si l'estimateur est sans biais, $b(\theta) = 0$ pour tout $\theta \in \Theta$ ce qui donne

$$R(\theta,\hat{\theta}) \geq \frac{1}{I_n(\theta)}.$$