ECOLE NATIONALE SUPERIEURE DE LA STATISTIQUE ET D'ECONOMIE APPLIQUEE

OCTOBRE 2015

<u>CONCOURS D'ACCES EN DOCTORAT LMD</u>

<u>Matière</u>: Statistique Mathématique
<u>Durée:</u> 02 heures

Soit X une variable aléatoire de loi de Pareto $\mathcal{P}_{\theta,a}$ de paramètres $\theta > 0$ et $a > 1$ et $X_1, X_2, \cdots, X_n$ un n-échantillon de X. La densité d'une loi de Pareto est donnée par:

$$f_a(x,\theta) = \frac{\theta a^\theta}{x^{\theta+1}}, \quad x > a$$

1. Montrer que $Y = \ln\left(\frac{X}{a}\right)$ suit une loi exponentielle de paramètre θ.
2. Déterminer l'estimateur du maximum de vraisemblance de θ, noté $\widehat{\theta}_{MV}$.
3. Montrer que $\widehat{\theta}_{MV}$ s'écrit sous la forme $\widehat{\theta}_{MV} = \frac{n}{Z_n}$, où Z_n est une variable aléatoire de loi Gamma de paramètres n et θ. $(Z_n \sim \gamma(n,\theta))$
4. Calculez[1] la moyenne et la variance de $\widehat{\theta}_{MV}$. Déduisez-en un estimateur sans biais $\widehat{\theta}_n$.de θ. Quelle est la variance de $\widehat{\theta}_n$? Est-il convergent?
5. Notons $B^{(n)}(\theta)$ la borne de Cramer-Rao pour l'estimation de θ. Dans l'ensemble des estimateurs sans biais de θ, on définit l'efficacité relative de $\widehat{\theta}_n$ comme étant $e\left(\widehat{\theta}_n\right) := \frac{B^{(n)}(\theta)}{V\left(\widehat{\theta}_n\right)}$

   Calculez $e\left(\widehat{\theta}_n\right)$. L'estimateur$\widehat{\theta}_n$ est-il efficace? Est-il asymptôtiquement efficace?
6. On suppose a connu et on désire tester les deux hypothèses

   $$H_0 : \theta = \text{ vs } H_1 : \theta = \theta_1 \text{ avec } \theta_1 > \theta_0$$

   (a) Démontrer que la région critique du test de niveau α est alors de la forme

   $$W = \{Z_n < k_n\}$$

   (b) On considère la fonction Q_n définie par: $Q_n(x) = \frac{1}{\Gamma(n)} \int_0^x t^{n-1} e^{-t} dt$. Déterminer la valeur du seuil k_n en fonction de θ_0, α et de Q_n.

   (c) Pour $\theta_1 = 2\theta_0$, montrer que la puissance de ce test: $1 - \beta = Q_n\left(2Q_n^{-1}(\alpha)\right)$.

---

[1] <u>Loi Gamma</u>: Si $X \leadsto \gamma(a,b)$ alors $f_X(x) = \frac{b^a}{\Gamma(a)} x^{a-1} e^{-bx}, \; x > 0$

$\Gamma(a+1) = a\Gamma(a)$ et $\int_0^{+\infty} x^{a-1} e^{-bx} = \frac{\Gamma(a)}{b^a}$

ECOLE NATIONALE SUPERIEURE DE STATISTIQUE ET D'ECONOMIE APPLIQUEE

OCTOBRE 2015

CONCOURS D'ACCES EN DOCTORAT LMD
(Corrigé)

Matière: Statistique Mathématique
Durée: 02 heures

$X \sim \mathcal{P}_{\theta,a}$ (loi de Pareto) et $X_1, X_2, \cdots, X_n \sim X$: $X \sim f_a(x,\theta) = \dfrac{\theta a^{\theta}}{x^{\theta+1}}, \quad x > a$

1.

$$\begin{aligned} F_Y(y) &= P(Y \leq y) = P\left(\ln\left(\frac{X}{a}\right) \leq y\right) \ , \ y > 0 \\ &= P(X \leq ae^{y}) = F_X(ae^{y}) \end{aligned}$$

La densité de la variable aléatoire Y est:

$$f_Y(y) = \frac{\partial}{\partial y} F_Y(y) = \frac{\partial}{\partial y} F_X(ae^{y}) = ae^{y} f_a(ae^{y},\theta) = ae^{y} \frac{\theta a^{\theta}}{(ae^{y})^{\theta+1}} = \theta e^{-\theta y} \ , \ y > 0$$

2. La vraisemblance $L(\theta, a, x)$ s'écrit:

$$L(\theta,a,x) = \prod_{i=1}^{n} f_a(x_i,\theta) = \prod_{i=1}^{n} \frac{\theta a^{\theta}}{x_i^{\theta+1}} = \frac{\theta^n a^{n\theta}}{\prod_{i=1}^{n} x_i^{\theta+1}} = \frac{\theta^n a^{n\theta}}{\left(\prod_{i=1}^{n} x_i\right)^{\theta+1}}$$

et Log-vraisemblance $l(\theta)$ s'écrit:

$$l(\theta) = \ln L(\theta,a,x) = n\ln\theta + n\theta\ln a - (\theta+1)\sum_{i=1}^{n} \ln x_i$$

$$\frac{\partial}{\partial\theta} l(\theta) = 0 \Longleftrightarrow \frac{n}{\theta} + n\ln a - \sum_{i=1}^{n} \ln x_i = 0$$

et on a:

$$\theta^* = \frac{n}{\sum_{i=1}^{n} \ln x_i - n\ln a} \quad \text{et} \quad \left.\frac{\partial^2}{\partial\theta^2} l(\theta)\right|_{\theta=\theta^*} = -\frac{n}{\theta^{*2}} < 0$$

Donc l'EMV de θ

$$\widehat{\theta}_{MV} = \frac{n}{\sum_{i=1}^{n} \ln X_i - n\ln a}$$

3.

$$\widehat{\theta}_{MV} = \frac{n}{\sum_{i=1}^{n} \ln X_i - n \ln a} = \frac{n}{\sum_{i=1}^{n} \ln X_i - \sum_{i=1}^{n} \ln a} = \frac{n}{\sum_{i=1}^{n} (\ln X_i - \ln a)} = \frac{n}{\sum_{i=1}^{n} \ln\left(\frac{X_i}{a}\right)} = \frac{n}{Z_n}$$

Avec $Z_n = \sum_{i=1}^{n} \ln\left(\frac{X_i}{a}\right) = \sum_{i=1}^{n} Y_i \doteq \sum_{i=1}^{n} \zeta(\theta) \doteq \gamma(n,\theta)$ $(i.e\ Z_n \sim \gamma(n,\theta))$.

4.

$$\begin{aligned} E\left(\widehat{\theta}_{MV}\right) &= E\left(\frac{n}{Z_n}\right) = nE\left(\frac{1}{Z_n}\right) = n\int_0^{+\infty} \frac{1}{z}\frac{\theta^n}{\Gamma(n)} z^{n-1} e^{-\theta z} dz \\ &= n\frac{\theta^n}{\Gamma(n)} \int_0^{+\infty} z^{n-1-1} e^{-\theta z} dz = n\frac{\theta^n}{\Gamma(n)} \frac{\Gamma(n-1)}{\theta^{n-1}} = \frac{n}{n-1}\theta \end{aligned}$$

Avec le même calcul, on trouve

$$E\left(\widehat{\theta}^2_{MV}\right) = \frac{n^2}{(n-1)(n-2)}\theta^2$$

et par la suite, on a:

$$V\left(\widehat{\theta}_{MV}\right) = \frac{n^2}{(n-1)^2(n-2)}\theta^2$$

L'estimateur $\widehat{\theta}_{MV}$ est biaisé. $\widehat{\theta}_n = \frac{n-1}{n}\widehat{\theta}_{MV} = \frac{n-1}{Z_n}$ est un estimateur sans biais de θ et de variance

$$V\left(\widehat{\theta}_n\right) = \left(\frac{n-1}{n}\right)^2 V\left(\widehat{\theta}_{MV}\right) = \frac{\theta^2}{(n-2)}$$

Comme: $\begin{cases} \lim\limits_{n\to+\infty} E\left(\widehat{\theta}_n\right) = \theta \\ \text{et} \\ \lim\limits_{n\to+\infty} V\left(\widehat{\theta}_n\right) = 0 \end{cases}$. Alors $\widehat{\theta}_n$ est un estimateur convergent de θ

5.

$$e\left(\widehat{\theta}_n\right) := \frac{B^{(n)}(\theta)}{V\left(\widehat{\theta}_n\right)} = \frac{\left(\gamma'(\theta)\right)^2 / I_n(\theta)}{V\left(\widehat{\theta}_n\right)} = \frac{1/\left(n/\theta^2\right)}{\frac{\theta^2}{(n-2)}} = \frac{n-2}{n}$$

$e\left(\widehat{\theta}_n\right) := \frac{n-2}{n} \neq 1$, l'estimateur $\widehat{\theta}_n$ n'est pas efficace. Mais il est asymptôtiquement efficace car $e\left(\widehat{\theta}_n\right) := \frac{n-2}{n} \underset{n\to+\infty}{\longrightarrow} 1$

6. On suppose a connu et on désire tester les deux hypothèses

$$H_0 : \theta = \theta_0 \text{ vs } H_1 : \theta = \theta_1 \text{ avec } \theta_1 > \theta_0$$

(a) D'après le théorème de NEYMAN et PEARSON, $\forall\alpha$ donné, $0 \leq \alpha \leq 1$, il existe un test pur Φ de niveau α, de puissance maximale défini par la région critique W :

$$W = \left\{ x \in \chi^n : \frac{L(x, \theta_0)}{L(x, \theta_1)} \leq K \right\}$$

En effet

$$\frac{L(x,\theta_0)}{L(x,\theta_1)} = \frac{\theta_0^n a^{n\theta_0}}{\left(\prod_{i=1}^n x_i\right)^{\theta_0+1}} \frac{\left(\prod_{i=1}^n x_i\right)^{\theta_1+1}}{\theta_1^n a^{n\theta_1}} \leq K \Longleftrightarrow \left(\prod_{i=1}^n x_i\right)^{\theta_1-\theta_0} \leq K'$$

On conclut que

$$W = \left\{ x \in \chi^n : \sum_{i=1}^n \ln X_i \leq K'' \right\}$$

et finalement, on peut déduire que la région critique

$$W = \{Z_n < k_n\}$$

1. b.

$$\begin{aligned} \alpha = P_{H_0}(W) &= P_{H_0}(Z_n < k_n) = \int_0^{k_n} \tfrac{\theta_0^n}{\Gamma(n)} z^{n-1} e^{-\theta_0 z} dz \\ &\overset{t=\theta_0 z}{=} \int_0^{\theta_0 k_n} \tfrac{\theta_0^n}{\Gamma(n)} \left(\frac{t}{\theta_0}\right)^{n-1} e^{-t} \frac{dt}{\theta_0} = \tfrac{1}{\Gamma(n)} \int_0^{\theta_0 k_n} t^{n-1} e^{-t} dt = Q_n(\theta_0 k_n) \end{aligned}$$

alors $k_n = Q_n^{-1}(\alpha)/\theta_0$

1. c.

$$\begin{aligned} 1 - \beta = P_{H_1}(W) &= P_{H_1}(Z_n < k_n) = Q_n(\theta_1 k_n) = Q_n\left(\frac{\theta_1}{\theta_0} Q_n^{-1}(\alpha)\right) \\ &= Q_n\left(\frac{2\theta_0}{\theta_0} Q_n^{-1}(\alpha)\right) = Q_n\left(2Q_n^{-1}(\alpha)\right) \end{aligned}$$