Master Math-Info 2ème année
Spécialité Recherche en Mathématiques Appliquées

Examen de Statistique des valeurs extrêmes

6 mars 2006

Notations utilisées dans le sujet :

- On note $X_{1,n} \leq \ldots \leq X_{n,n}$ l'échantillon ordonné associé à l'échantillon $X_1, \ldots, X_n$.
- $Z_n = O_{\mathrm{P}}(1)$ signifie que la variable aléatoire Z_n est bornée en probabilité. Autrement dit, $\forall \epsilon > 0$, $\exists c > 0$, $N \in \mathbb{N}$ tels que $\forall n \geq N$, $\mathrm{P}(|Z_n| > c) < \epsilon$.
- $\epsilon_n = o_{\mathrm{P}}(1)$ signifie que ϵ_n converge en probabilité vers zéro lorsque $n \to \infty$.
- $X \stackrel{\mathcal{L}}{=} Y$ signifie que X et Y ont même loi.
- La convergence en probabilité est notée $\stackrel{\mathrm{P}}{\to}$ et la convergence en loi $\stackrel{\mathcal{L}}{\to}$.

On considère un échantillon $X_1, \ldots, X_n$ de variables aléatoires indépendantes et identiquement distribuées (i.i.d.) de fonction de répartition F. On suppose que F est bijective.
On définit les variables aléatoires $Y_1, \ldots, Y_n$ par $Y_i = V(X_i),\ i = 1, \ldots, n$ où $V = 1/(1-F)$.

Question 1 : Montrer que les variables aléatoires $Y_1, \ldots, Y_n$ sont i.i.d. de fonction de répartition

$$\Phi(y) = \begin{cases} 0 & \text{si } y < 1, \\ 1 - 1/y & \text{si } y \geq 1. \end{cases}$$

Quelle loi reconnaissez-vous ?

On définit la statistique

$$M_n(k_n) = \frac{1}{k_n} \sum_{i=1}^{k_n} \ln^2 \left(\frac{X_{n-i+1,n}}{X_{n-k,n}} \right),$$

où $1 \leq k_n \leq n$. Pour simplifier, on notera dans la suite $k = k_n$.

Question 2 : Déduire de la question 1 que

$$M_n(k_n) \stackrel{\mathcal{L}}{=} \frac{1}{k_n} \sum_{i=1}^{k_n} \ln^2 \left(\frac{U(Y_{n-i+1,n})}{U(Y_{n-k,n})} \right),$$

où U est l'inverse de la fonction V.

A présent, on suppose que F appartient au domaine d'attraction de Fréchet avec un indice de valeur extrême $\gamma > 0$.

Question 3 : En utilisant un résultat du cours, justifier le fait que U est à variations régulières d'indice γ (i.e. $U(x) = x^\gamma \ell(x)$ où ℓ est une fonction à variations lentes).

On suppose que ℓ vérifie la condition du second ordre (C_1) avec $\rho < 0$.

Question 4 : Montrer que $\forall \epsilon > 0$, $\exists x_0$ tels que $\forall x \geq x_0$, $\lambda \geq 1$,

$$\frac{(1+\epsilon)\lambda^{\rho+\epsilon} - 1}{\rho} \leq \frac{\ln(U(\lambda x)/U(x)) - \gamma \ln(\lambda)}{b(x)} \leq \frac{(1-\epsilon)\lambda^{\rho-\epsilon} - 1}{\rho}.$$

Question 5 : En déduire que $\forall 0 < \epsilon < \min(1, -\rho)$, $\exists x_0$ tels que $\forall x \geq x_0$, $\lambda \geq 1$,

$$\frac{1}{\rho} < \frac{\epsilon}{\rho} \leq \frac{\ln(U(\lambda x)/U(x)) - \gamma \ln(\lambda)}{b(x)} \leq -\frac{1}{\rho}.$$

On suppose que $k \to \infty$ et $k/n \to 0$ lorsque $n \to \infty$.

Question 6 : En remarquant que $\Phi(Y)$ suit une loi uniforme sur l'intervalle $[0, 1]$, montrer que $Y_{n-k,n} \xrightarrow{\mathrm{P}} +\infty$.

Question 7 : En utilisant les questions 5 et 6, montrer que pour $i = 1, \ldots, k$,

$$\ln\left(\frac{U(Y_{n-i+1,n})}{U(Y_{n-k,n})}\right) = \gamma \ln\left(\frac{Y_{n-i+1,n}}{Y_{n-k,n}}\right) + Z_i b(Y_{n-k,n}),$$

où Z_i est une variable aléatoire telle que, pour tout $i = 1, \ldots, k$, $|Z_i| \leq -1/\rho$.

On admet que pour $i = 1, \ldots, k$,

$$\frac{Y_{n-i+1,n}}{Y_{n-k,n}} \stackrel{\mathcal{L}}{=} Y_{i,k}.$$

On rappelle de plus que $U_{k,n} = k/n(1 + o_{\mathrm{P}}(1))$.

Question 8 : Montrer que

$$\ln\left(\frac{U(Y_{n-i+1,n})}{U(Y_{n-k,n})}\right) = \gamma \ln(Y_{i,k}) + Z_i b(n/k)(1 + o_{\mathrm{P}}(1)),$$

où la variable aléatoire $o_{\mathrm{P}}(1)$ ne dépend pas de i.

Question 9 : En déduire que

$$\begin{aligned} M_n(k) &\overset{\mathcal{L}}{=} \gamma^2 \frac{1}{k}\sum_{i=1}^{k} \ln^2(Y_i) + b^2(n/k)(1+o_{\mathrm{P}}(1))\frac{1}{k}\sum_{i=1}^{k} Z_i^2 + \gamma b(n/k)\frac{1}{k}\sum_{i=1}^{k}\ln(Y_{i,k})Z_i. \\ &= M_{n,1}(k) + M_{n,2}(k) + M_{n,3}(k). \end{aligned}$$

On admet que $\mathrm{E}(\ln^2(Y_1)) = 2$ et $\mathrm{Var}(\ln^2(Y_1)) = 20$.

Question 10 : Montrer que :

a) $\sqrt{k}(M_{n,1}(k) - 2\gamma^2) \xrightarrow{\mathcal{L}} \mathcal{N}(0, 20\gamma^4)$.

b) $M_{n,2}(k) = O_{\mathrm{P}}(b^2(n/k))$.

c) $M_{n,3}(k) = O_{\mathrm{P}}(b(n/k))$.

Indication : Utiliser, en le justifiant, le fait que $1/k \sum \ln(Y_i) \xrightarrow{\mathrm{P}} E(\ln(Y))$.

Question 11 : Donner les conditions sur k pour que

$$\sqrt{k}(M_n(k) - 2\gamma^2) \xrightarrow{\mathcal{L}} \mathcal{N}(0, 20\gamma^4).$$

Question 12 : En déduire un estimateur de γ et comparer le avec l'estimateur de Hill.