M2 Statistique 2015-2016, Semestre 3
Parcours « Biostatistique et Statistiques Industrielles » et « Actuariat »

Statistique des valeurs extrêmes *(1ère Session)*

(*Durée : 2h00. Aucun document n'est autorisé. Usage de la calculatrice interdit.*)

Notations utilisées dans le sujet :

- Pour tout échantillon $W_1, \ldots, W_N$, on note $W_{1,N} \leq \ldots \leq W_{N,N}$ l'échantillon ordonné associé.
- $X \overset{\mathcal{L}}{=} Y$ signifie que X et Y ont même loi.
- Si (W_n) et (Z_n) sont deux suites de variables aléatoires réelles, $W_n \overset{\mathbb{P}}{\sim} Z_n$ signifie que W_n/Z_n converge en probabilité vers 1.

Exercice 1 [6 pts] – On rappelle que la fonction de répartition d'une loi de Pareto Généralisée est donnée pour $\gamma \in \mathbb{R}$ et $\theta > 0$ par

$$G_{\gamma,\theta}(x) = \begin{cases} 0 & \text{si } x \leq 0, \\ 1 - (1 + \gamma x/\theta)_+^{-1/\gamma} & \text{si } x \geq 0. \end{cases}$$

1) Donner la densité d'une loi de Pareto Généralisée de paramètres $\gamma \in \mathbb{R}$ et $\theta > 0$.

Dans le reste de cet exercice, on considère une variable aléatoire X suivant une loi de Pareto Généralisée de paramètres $\gamma \neq 0$ et $\theta = |\gamma|$.

2) Pour $s \in \mathbb{N}$, on pose $\nu_s = \mathbb{E}[X(1 - G_{\gamma,|\gamma|}(X))^s]$. Donner les valeurs de γ pour lesquelles $\nu_s < +\infty$ pour tout $s \in \mathbb{N}$ et calculer la valeur de ν_s.

Pour les questions restantes, on suppose que $\gamma < 0$.

3) En utilisant la question précédente, proposer un estimateur $\hat{\gamma}_n$ de γ construit à partir de n variables aléatoires $X_1, \ldots, X_n$ indépendantes et de même loi que X.

4) Montrer que $\hat{\gamma}_n$ converge presque-sûrement vers $\gamma < 0$.

On rappelle le résultat suivant. Soient (Z_n) une suite de variables aléatoires réelles et soit Z une variable aléatoire réelle. On suppose qu'il existe une suite (v_n) convergeant vers $+\infty$ et $\xi \in \mathbb{R}$ tels que $v_n(Z_n - \xi) \xrightarrow{\mathcal{L}} Z$ lorsque $n \to \infty$. Pour toute fonction $g(\cdot)$ telle que $g'(\xi)$ existe, on a $v_n(g(Z_n) - g(\xi)) \xrightarrow{\mathcal{L}} g'(\xi)Z$.

5) Montrer que $\sqrt{n}(\hat{\gamma}_n - \gamma) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \sigma^2)$. Vous préciserez l'expression de σ^2 en fonction de γ.

Exercice 2 [14 pts] – On considère une variable aléatoire X suivant une loi de Pareto de paramètres $\gamma > 0$ et $c > 1$. On rappelle que la fonction de répartition d'une loi de Pareto est $F(x) = 1 - (x/c)^{-1/\gamma}$ pour $x \geq c$ et $F(x) = 0$ pour $x < c$.

1) Soit $\theta > 0$. On pose $Y = \log^\theta(X)$. Calculer la fonction de répartition de Y. Cette fonction de répartition sera notée $G_\theta(\cdot)$ dans la suite de cet exercice.

2) Rappeler la définition de l'appartenance d'une fonction de répartition au domaine d'attraction de Gumbel.

3) **Dans le cas particulier où $\theta = 1$**, montrer que $G_1(\cdot)$ appartient au domaine d'attraction de Gumbel. Vous préciserez votre choix pour les suites de normalisation. (*Indication : une des deux suites est constante, l'autre converge vers* $+\infty$).

4) Montrer que la fonction de répartition $G_\theta(\cdot)$ peut s'écrire sous la forme $G_\theta(y) = 1 - \exp(-y^{1/\theta} L(y))$ où $L(\cdot)$ est une fonction à valeurs positives que vous préciserez. Montrer également que $L(\cdot)$ est une fonction à variations lentes.

5) Rappeler la définition d'une fonction à variations régulières d'indice $\xi \in \mathbb{R}$. Pour une fonction croissante $F(\cdot)$, rappeler la définition de l'inverse généralisée $F^{\leftarrow}(\cdot)$ de $F(\cdot)$.

6) On rappelle que si une fonction est à variations régulières d'indice $\xi > 0$ alors son inverse généralisée est à variations régulières d'indice $1/\xi$. Utiliser ce rappel pour montrer

que pour tout $\alpha \in [0,1]$,

$$q(\alpha) = G_\theta^{\leftarrow}(1-\alpha) = \log^\theta(1/\alpha)\ell(\log(1/\alpha)),$$

où $\ell(\cdot)$ est une fonction à variations lentes.

On admettra dans la suite de cet exercice que $\ell(\cdot)$ converge vers une constante strictement positive à l'infini.

7) Soient (α_n) et (β_n) deux suites convergeant vers 0 lorsque $n \to \infty$. Montrer que

$$\log q(\alpha_n) - \log q(\beta_n) \sim \theta \log \frac{\log(1/\alpha_n)}{\log(1/\beta_n)},$$

lorsque $n \to \infty$.

On considère à présent les variables aléatoires $Y_1, \dots, Y_n$ que l'on suppose indépendantes et de même fonction de répartition $G_\theta(\cdot)$. Dans toute la suite, (k_n) est une suite vérifiant $k_n \to \infty$ et $n/k_n \to \infty$.

8) En vous aidant de la question précédente et en vous inspirant de la construction de l'estimateur de Hill vue en cours, proposer un estimateur $\hat{\theta}_n$ de θ. (*Indication : prendre $\beta_n = k_n/n$ et $\alpha_n = (2i-1)/(2n)$ pour $i = 1, \dots, k_n$.*)

9) En vous aidant de la question 7), justifier pourquoi

$$\hat{q}_n(\alpha_n) := Y_{n-k_n,n} \left(\frac{\log(1/\alpha_n)}{\log(n/k_n)} \right)^{\hat{\theta}_n}$$

est un estimateur possible de $q(\alpha_n)$.

10) Soient $U_1, \dots, U_n$ des variables aléatoires indépendantes de loi uniforme sur $[0,1]$. Montrer que $\log(U_{k_n+1,n}) \overset{\mathbb{P}}{\sim} \log(k_n/n)$. (*Indication : utiliser la représentation de Rényi*).

11) Montrer que $\log(Y_{n-k_n,n}) \overset{\mathcal{L}}{=} \theta \log\log(1/U_{k_n+1,n}) + \log \ell(\log(1/U_{k_n+1,n}))$.

12) En utilisant les questions 6) et 11), montrer que

$$\log \frac{\hat{q}_n(\alpha_n)}{q(\alpha_n)} - (\hat{\theta}_n - \theta) \log \frac{\log(1/\alpha_n)}{\log(n/k_n)} \overset{\mathcal{L}}{=} \theta \log \frac{\log(U_{k_n+1,n})}{\log(k_n/n)} + \log \frac{\ell(\log(1/U_{k_n+1,n}))}{\ell(\log(1/\alpha_n))}.$$

13) En supposant que

$$(\hat{\theta}_n - \theta) \log \frac{\log(1/\alpha_n)}{\log(n/k_n)} \xrightarrow{\mathbb{P}} 0,$$

montrer que $\hat{q}_n(\alpha_n)/q(\alpha_n) \xrightarrow{\mathbb{P}} 1$.

14) On souhaite ici démontrer que la fonction à variations lentes $\ell(\cdot)$ apparaissant à la question 6) converge vers une constante. En partant du fait que $\bar{G}_\theta(\bar{G}_\theta^{\leftarrow}(\alpha)) = \alpha$, montrer tout d'abord que $\bar{G}_\theta^{\leftarrow}(\alpha) = (\log(1/\alpha))^\theta L^{-\theta}(\bar{G}_\theta^{\leftarrow}(\alpha))$. En déduire ensuite que $\ell(x) \to \gamma^\theta$ lorsque $x \to \infty$.