M2 Statistique
Parcours "Biostatistique et Statistiques Industrielles" et "Actuariat"

Examen de Statistique des valeurs extrêmes

année 2013-2014

(*Durée : 2h00. Seule la calculatrice est autorisée.*)

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.
- Pour toute fonction $\Phi(.)$ croissante, $\Phi^{\leftarrow}(y) := \inf\{x | \Phi(x) \geq y\}$.
- $\epsilon_n = o_{\mathbb{P}}(1)$ 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.

Questions de cours [5 pts] –

1) Donner la définition d'une fonction à variations régulières d'indice $\alpha \in \mathbb{R}$ ainsi que celle d'une fonction à variations lentes.

2) Soient $F(.)$ et $G(.)$ deux fonctions de répartition. Rappeler la définition de $F(.)$ appartient au domaine d'attraction de $G(.)$.

3) Compléter l'énoncé du théorème fondamental de la théorie des valeurs extrêmes : *Si* $F(.)$ *appartient au domaine d'attraction de* $G(.)$*, alors* $G(.)$ *est de même type que l'une des trois fonctions de répartition suivantes :* ...

4) Donner la représentation de Rényi pour des variables aléatoires uniformes. En déduire que si $U_1, \ldots, U_n$ sont des variables indépendantes de loi uniforme sur $[0,1]$ et si $(k_n) \in \{1, \ldots, n\}$ est une suite d'entiers telle que $k_n \to \infty$, alors $U_{k_n+1,n} = k_n/n(1 + o_{\mathbb{P}}(1))$.

Exercice 1 [2 pts] – On considère la fonction de répartition d'une loi de Fréchet donnée par

$$F(x) = \exp(-1/x)\mathbb{I}_{\{x \geq 0\}}.$$

1) Donner l'expression du quantile $q(\alpha) := F^{\leftarrow}(1-\alpha)$, $\alpha \in [0,1]$ de la loi de Fréchet.

2) Montrer que $q(\alpha) = \alpha^{-1}\ell(\alpha^{-1})$, où $\ell(.)$ est une fonction à variations lentes. Qu'en déduisez vous ?

Exercice 2 [3 pts] – Soit $\beta > 0$. On considère la fonction

$$f(x) = \beta(1-x)^{\beta-1}\mathbb{I}_{\{x\in[0,1]\}}.$$

1) Montrer que $f(.)$ est une densité.

2) Calculer la fonction de survie $\bar{F}(.) = 1 - F(.)$ associée.

3) Montrer que $F(.)$ appartient au domaine d'attraction de Weibull et donner un choix possible pour les suites de normalisation.

Exercice 3 [10 pts] – On considère un échantillon $(X_1, \dots, X_n)$ de variables aléatoires indépendantes et de même fonction de répartition $F(.)$ telle que pour $\alpha \in [0,1]$:

$$q(\alpha) := F^{\leftarrow}(1-\alpha) = \alpha^{-\gamma}(1+\alpha),\ \gamma \geq 1/2.$$

1) Montrer que la fonction $q(.)$ est bien décroissante sur $]0,1[$.

2) A quel domaine d'attraction appartient la fonction de répartition F (justifier).

3) Rappeler la définition de l'estimateur de Hill $\hat{\gamma}_n^{(H)}$ de l'indice des valeurs extrêmes $\gamma > 0$. Cet estimateur dépend d'une suite d'entiers $(k_n) \in \{1, \dots, n\}$ que l'on supposera telle que $k_n \to \infty$ et $n/k_n \to \infty$ lorsque $n \to \infty$.

4) Soient $U_1, \dots, U_n$ des variables aléatoires indépendantes de loi uniforme sur $[0,1]$. Montrer que

$$\{\log X_{n-i+1,n},\ i = 1, \dots, k_n\} \stackrel{\mathcal{L}}{=} \{\gamma \log U_{i,n}^{-1} + \log(1+U_{i,n}),\ i = 1, \dots, k_n\}.$$

5) En déduire que

$$\hat{\gamma}_n^{(H)} \stackrel{\mathcal{L}}{=} \gamma \frac{1}{k_n}\sum_{i=1}^{k_n} \log \frac{U_{i,n}^{-1}}{U_{k_n+1,n}^{-1}} + \frac{1}{k_n}\sum_{i=1}^{k_n} \log \frac{1+U_{i,n}}{1+U_{k_n+1,n}}.$$

6) En utilisant la représentation de Rényi pour des lois uniformes, montrer que

$$\frac{1}{k_n}\sum_{i=1}^{k_n} \log \frac{U_{i,n}^{-1}}{U_{k_n+1,n}^{-1}} \stackrel{\mathcal{L}}{=} \frac{1}{k_n}\sum_{i=1}^{k_n} E_i,$$

où $E_1, \dots, E_n$ sont des variables aléatoires indépendantes de loi exponentielle de paramètre 1.

7) On considère la fonction définie par $\ell(x) = 1 + x^{-1}$ pour tout $x \geq 1$. Montrer que $\ell(.)$ est une fonction à variations lentes et, en vous aidant de l'expression de $\ell'(x)/\ell(x)$, montrer que

$$\ell(x) = 2\exp\left(-\int_1^x \frac{1}{t(1+t)}dt\right).$$

8) Montrer que

$$\frac{1}{k_n}\sum_{i=1}^{k_n} \log \frac{1+U_{i,n}}{1+U_{k_n+1,n}} = \frac{1}{k_n}\sum_{i=1}^{k_n} \int_{U_{k_n+1,n}^{-1}}^{U_{i,n}^{-1}} \frac{-1}{t(1+t)}dt,$$

et en déduire que

$$\left|\frac{1}{k_n}\sum_{i=1}^{k_n} \log \frac{1+U_{i,n}}{1+U_{k_n+1,n}}\right| \leq \frac{1}{1+U_{k_n+1,n}^{-1}} \frac{1}{k_n}\sum_{i=1}^{k_n} \log \frac{U_{i,n}^{-1}}{U_{k_n+1,n}^{-1}}.$$

9) Déduire de ce qui précède que l'estimateur de Hill converge en probabilité vers γ.