Master “Mathématiques et Applications”, parcours “Statistique” et M2 “Actuariat”

Statistique des valeurs extrêmes

(1ère Session, année 2020-2021, Semestre 3)

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.
- $U_1, \ldots$ sont des variables aléatoires indépendantes de loi uniforme standard et $E_1, \ldots$ des variables aléatoires indépendantes de loi exponentielle standard.
- Pour tout $n \in \mathbb{N}^*$, $k_n \in \{1, \ldots, n\}$ est une suite telle que $k_n \to \infty$ et $n/k_n \to \infty$ lorsque n tend vers l'infini.

Partie I Soit $\varphi : [0, \infty[\to [0, \infty[$ une fonction continue, strictement croissante et telle que $\varphi(0) = 0$ et $\varphi(x) \to \infty$ lorsque $x \to \infty$. On considère la fonction de répartition définie pour tout $x \geq 0$ par $F(x) := 1 - \exp[-(\varphi(x))^\theta]$, où $\theta > 0$.

1) Expliquer pourquoi F n'appartient pas au domaine d'attraction de Weibull.
2) Montrer que si F vérifie la condition suivante : il existe $\xi > 0$ tel que pour tout $t > 0$,
$$\lim_{x \to \infty} \frac{(\varphi(tx))^\theta - (\varphi(x))^\theta}{\ln(t)} = \xi, \tag{1}$$
alors F appartient au domaine d'attraction de Fréchet. Vous préciserez la valeur de l'indice des valeurs extrêmes.
3) Proposer une fonction φ vérifiant la condition (1), cette fonction pouvant dépendre de θ.
4) Démontrer que si φ est une fonction à variations régulières d'indice $\beta > 0$ alors F n'appartient pas au domaine d'attraction de Fréchet.

On suppose dans toute la suite que φ est deux fois dérivable. On notera φ' sa dérivée et φ'' sa dérivée seconde.
On rappelle qu'une fonction de répartition G de point terminal x_G appartient au domaine d'attraction de Gumbel si et seulement si il existe $x_0 < x_G$, une fonction positive c avec $c(x) \to d > 0$

lorsque $x \to x_G$ et une fonction positive et dérivable a de dérivée a' vérifiant $a'(x) \to 0$ lorsque $x \to x_G$ tels que

$$1 - G(x) = c(x) \exp\left\{-\int_{x_0}^{x} \frac{1}{a(t)} dt\right\}.$$

5) Donner, en fonction de φ' et θ, l'expression de la fonction a telle que pour tout $x > 0$,

$$1 - F(x) = \exp\left\{-\int_{0}^{x} \frac{1}{a(t)} dt\right\}.$$

6) En déduire qu'une condition suffisante pour que F appartienne au domaine d'attraction de Gumbel est que

$$\lim_{x\to\infty} \frac{\varphi''(x)\varphi(x)}{(\varphi'(x))^2} < \infty.$$

7) Montrer que si $\varphi(x) = x^\beta$ avec $\beta > 0$, alors F appartient au domaine d'attraction de Gumbel. Dans ce cas, donner un choix possible des suites de normalisation $a_n > 0$ et $b_n \in \mathbb{R}$ telles que $F^n(a_n x + b_n) \to \exp(-\exp(-x))$. Vous aurez besoin d'utiliser le développement de Taylor $(1+u)^p = 1 + pu(1 + O(u))$ lorsque $u \to 0$.

Partie II – Inférence statistique Soient $X_1, \ldots, X_n$ des variables aléatoires indépendantes de fonction de répartition commune définie pour $x \geq 0$ par

$$F(x) = 1 - \exp[-(\varphi(x))^\theta],$$

avec $\theta > 0$ et φ une fonction continue, strictement croissante et telle que $\varphi(0) = 0$. On supposera dans toute la suite que φ est une fonction à variations régulières d'indice $1/(\theta\xi)$ où $\xi > 0$. L'objectif de cette partie est de proposer et d'étudier un estimateur de ξ. On notera $q(u) := F^{\leftarrow}(1-u)$ le quantile associé à la fonction de répartition F.

8) Montrer que la fonction $x \mapsto q(1/x)$ est une fonction à variations lentes.
9) Montrer que la fonction $x \mapsto q(\exp(-x))$ est à variations régulières d'indice ξ.
10) Donner l'idée conduisant à l'approximation

$$\xi \approx \sum_{i=1}^{k_n-1} \ln\left(\frac{q(i/n)}{q(k_n/n)}\right) \Bigg/ \sum_{i=1}^{k_n-1} \ln\left(\frac{\ln(n/i)}{\ln(n/k_n)}\right)$$

pour k_n suffisamment grand. (Indication : partir de l'approximation $q(\exp(-tx))/q(\exp(-x)) \approx t^\xi$ pour x "grand" et prendre $x = \ln(n/k_n)$).

On rappelle que l'estimateur empirique de $q(u)$ est donné pour tout $u \in [0,1[$ par $X_{n-\lfloor nu \rfloor, n}$. On propose donc d'estimer ξ par

$$\hat{\xi}_n = \sum_{i=0}^{k_n-1} \ln\left(\frac{X_{n-i,n}}{X_{n-k_n,n}}\right) \Bigg/ \sum_{i=1}^{k_n-1} \ln\left(\frac{\ln(n/i)}{\ln(n/k_n)}\right)$$

11) Démontrer que $(n/k_n)U_{k_n+1,n} \stackrel{\text{p.s.}}{\to} 1$.

12) En déduire que $X_{n-k_n,1}/q(k_n/n) \stackrel{\text{p.s.}}{\to} 1$.

13) Montrer que le vecteur aléatoire $\{\ln X_{n-i,n} - \ln X_{n-k_n,n};\ i = 0, \ldots, k_n - 1\}$ a même loi que le vecteur

$$\left\{\xi \ln\left(1 + \frac{\ln(U_{i+1,n}/U_{k_n+1,n})}{\ln(U_{k_n+1,n})}\right) + \ln\left(\frac{\ell\left([\ln(U_{i+1,n}^{-1})]^{1/\theta}\right)}{\ell\left([\ln(U_{k_n+1,n}^{-1})]^{1/\theta}\right)}\right);\ i = 0, \ldots, k_n - 1\right\},$$

où ℓ est une fonction à variations lentes.

14) Montrer que

$$\left\{\frac{\ln(U_{i+1,n}/U_{k_n+1,n})}{\ln(U_{k_n+1,n})};\ i = 0, \ldots, k_n - 1\right\} \stackrel{\mathcal{L}}{=} \left\{\frac{E_{n-i,n} - E_{n-k_n,n}}{E_{n-k_n,n}};\ i = 0, \ldots, k_n - 1\right\}$$

On rappelle que

$$\{E_{j,n},\ j = 1, \ldots, n\} \stackrel{\mathcal{L}}{=} \left\{\sum_{r=1}^{j} \frac{F_r^*}{n - r + 1},\ j = 1, \ldots, n\right\},$$

où $F_1^*, \ldots, F_n^*$ sont des variables aléatoires indépendantes de loi exponentielle standard.

15) Montrer que le vecteur aléatoire $\{E_{n-i,n} - E_{n-k_n,n},\ i = 0, \ldots, k_n - 1\}$ a même loi que le vecteur $\{F_{k_n-i,k_n},\ i = 0, \ldots, k_n - 1\}$ où $F_1, \ldots, F_n$ sont des variables aléatoires indépendantes de loi exponentielle standard. Montrer également que les variables $\{E_{n-i,n} - E_{n-k_n,n},\ i = 0, \ldots, k_n - 1\}$ sont indépendantes de la variable $E_{n-k_n,n}$.

16) En utilisant les questions précédentes, montrer que

$$\left\{\ln\left(1 + \frac{\ln(U_{i+1,n}/U_{k_n+1,n})}{\ln(U_{k_n+1,n})}\right);\ i = 0, \ldots, k_n - 1\right\} \stackrel{\mathcal{L}}{=} \left\{\ln\left(1 + \frac{F_{k_n-i,k_n}}{E_{n-k_n,n}}\right);\ i = 0, \ldots, k_n - 1\right\},$$

où les variables $F_1, \ldots, F_{k_n}$ et $E_{n-k_n,n}$ sont indépendantes.

Nous supposons à présent que la fonction à variations lentes ℓ qui apparait dans la question 13) est de la forme

$$\ell(x) = d \exp\left[\int_0^x \frac{\Delta(t)}{t} dt\right],$$

où $d > 0$ est $|\Delta|$ est une fonction à variations régulières d'indice $\rho < 0$ qui est asymptotiquement décroissante.

17) Montrer que pour n assez grand

$$\left|\sum_{i=0}^{k_n-1} \ln\left(\frac{\ell\left([\ln(U_{i+1,n}^{-1})]^{1/\theta}\right)}{\ell\left([\ln(U_{k_n+1,n}^{-1})]^{1/\theta}\right)}\right)\right| \leq \frac{1}{\theta}\left|\Delta([\ln(U_{k_n+1,n}^{-1})]^{1/\theta})\right| \sum_{i=0}^{k_n-1} \ln\left(\frac{\ln(U_{i+1,n})}{\ln(U_{k_n+1,n})}\right).$$

18) En admettant que

$$\sum_{i=1}^{k_n} \ln\left(1+\frac{F_i}{E_{n-k_n,n}}\right) \Bigg/ \sum_{i=1}^{k_n-1} \ln\left(\frac{\ln(n/i)}{\ln(n/k_n)}\right) \xrightarrow{\mathbb{P}} 1,$$

où les variables $F_1, \ldots, F_{k_n}$ et $E_{n-k_n,n}$ sont indépendantes, utiliser les résultats obtenus précédemment pour montrer que $\hat{\xi}_n$ converge en probabilité vers ξ.