M2 Statistique 2014-2015
Parcours « Biostatistique et Statistiques Industrielles » et « Actuariat »

Correction de l'examen de Statistique des valeurs extrêmes

Questions de cours [4 pts] –

1) Le point terminal est la valeur maximale que l'on peut observer. Il est donné par $x_F = F^{\leftarrow}(1) = \bar{F}^{\leftarrow}(0)$.

2) Il est donné par $q(\alpha) = F^{\leftarrow}(1-\alpha) = \bar{F}^{\leftarrow}(\alpha)$.

3) L'indice des valeurs extrêmes γ est négatif et le point terminal est forcément fini. L'expression de $F(\cdot)$ est $F(x) = 1 - (x_F - x)^{-1/\gamma} L((x_F - x)^{-1})$.

4) La fonction de répartition des excès de X au dessus d'un seuil $u < x_F$ est

$$F_u(x) = \mathbb{P}(X - u \leq x | X \geq u) = 1 - \frac{\bar{F}(x+u)}{\bar{F}(u)}.$$

La fonction *moyenne des excès* est donnée par $e_u(X) = \mathbb{E}(X - u | X \geq u)$. Enfin, le théorème de Pickands assure que $F(\cdot)$ appartient au domaine d'attraction de la loi des valeurs extrêmes $H_\gamma(\cdot)$ si et seulement si

$$\lim_{u \uparrow x_F} \sup_{0 < x < x_F - u} \left| F_u(x) - G_{\gamma, a(u)}(x) \right| = 0,$$

où $G_{\gamma,\theta}(\cdot)$ est la fonction de répartition de la loi GPD de paramètres $\gamma \in \mathbb{R}$ et $\theta > 0$.

Exercice 1 [5 pts] –
1) Il est évident que $F(\cdot)$ est continue, croissante sur $[0, \infty[$. De plus, $F(x) \to 0$ lorsque $x \to 0$ et $F(x) \to 1$ lorsque $x \to \infty$. Le point terminal de $F(\cdot)$ est donc $+\infty$.

2) On a :
$$1 - F(x) = (1 + x^{\theta})^{-\lambda} = x^{-\theta\lambda}(1 + x^{-\theta})^{-\lambda} = x^{-1/\gamma}L(x),$$
avec $\gamma = 1/(\theta\lambda)$ et $L(x) = (1 + x^{-\theta})^{-\lambda} \to 1$ lorsque $x \to \infty$.

3) On a évidemment que $x^{-1/\gamma}(1 - F(x)) = L(x) = (1 + x^{-\theta})^{-\lambda}$. Cette fonction convergeant vers une constante, c'est une fonction à variations lentes.

4) On a :
$$\Delta(x) = \frac{xL'(x)}{L(x)} = \theta\lambda x^{-\theta}(1 + x^{-\theta})^{-1} = x^{-\theta}\ell(x),$$
avec $\ell(x) = \theta\lambda(1 + x^{-\theta})^{-1} \to \theta\lambda$ lorsque $x \to \infty$. Donc $\Delta(\cdot)$ est à variations régulières d'indice $-\theta$.

5) Calculons tout d'abord a_n. En résolvant l'équation $F(x) = \alpha$, on trouve que $F^{\leftarrow}(\alpha) = [(1 - \alpha)^{-1/\lambda} - 1]^{1/\theta}$. Ainsi, $a_n = (n^{1/\lambda} - 1)^{1/\theta}$.
On note ensuite que Y_n prend ses valeurs dans $[0, \infty[$. De plus, pour $x > 0$,
$$\mathbb{P}(Y_n \leq x) = F^n(a_n x) = \exp\left[n \log(1 - u_n)\right],$$
avec $u_n = (1 + (a_n x)^{\theta})^{-\lambda} \to 0$ car $a_n \to \infty$. Donc $\log(1 - u_n) \sim u_n \sim x^{-\theta\lambda}/n$. Ainsi, si $x > 0$,
$$\mathbb{P}(Y_n \leq x) \to \exp(-x^{-\theta\lambda}) = \exp(-x^{-1/\gamma}),$$
c'est-à-dire que $F(\cdot)$ appartient au domaine d'attraction de Fréchet.

Exercice 2 [11 pts] –

1) Sur $[0, \infty[$, $\bar{F}^*(\cdot)$ est évidemment décroissante puisque $F(\cdot)$ est une fonction de répartition (et donc une fonction croissante). De plus, $\bar{F}^*(x) = x^{1/\gamma}L(x)$ qui est à variations régulières d'indice $1/\gamma$.

2) En résolvant l'équation $\bar{F}^*(x) = \alpha$ c'est-à-dire l'équation $F(x_F - 1/x) = 1/\alpha$ on a bien que $(\bar{F}^*)^{\leftarrow}(\alpha) = (x_F - F^{\leftarrow}(1 - \alpha))^{-1}$. D'après le rappel précédent, on sait qu'il existe une fonction à variations lentes $\ell^*(\cdot)$ telle que
$$(\bar{F}^*)^{\leftarrow}(1/x) = x^{-\gamma}\ell^*(x) = (x_F - F^{\leftarrow}(1 - \alpha))^{-1}.$$
En posant $\alpha = 1/x$, on en déduit le résultat annoncé avec $\ell(\cdot) = 1/\ell^*(\cdot)$.

3) On sait que

$$\{X_1,\dots,X_n\} \stackrel{\mathcal{L}}{=} \{F^{\leftarrow}(U_1),\dots,F^{\leftarrow}(U_n)\} = \{x_F - (1-U_i)^{-\gamma}\ell((1-U_i)^{-1}),\ i=1,\dots,n\}.$$

En rangeant par ordre décroissant, il vient

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

4) La question 3) implique que

$$\left\{\frac{x_F - X_{n-i+1,n}}{x_F - X_{n-k,n}},\ i=1,\dots,k\right\} \stackrel{\mathcal{L}}{=} \left\{\frac{U_{i,n}^{-\gamma}}{U_{k+1,n}^{-\gamma}}\frac{\ell(U_{i,n}^{-1})}{\ell(U_{k+1,n}^{-1})},\ i=1,\dots,k\right\}.$$

En prenant le logarithme et en sommant sur $i=1,\dots,k$, on obtient le résultat.

5) La représentation de Rényi implique que

$$\left\{\frac{U_{i,n}^{-1}}{U_{k+1,n}^{-1}},\ i=1,\dots,k\right\} \stackrel{\mathcal{L}}{=} \left\{\frac{T_{k+1}}{T_i},\ i=1,\dots,k\right\} \stackrel{\mathcal{L}}{=} \{U_{i,k}^{-1},\ i=1,\dots,k\}.$$

En sommant, on a donc

$$\sum_{i=1}^{k}\log\frac{U_{i,n}^{-1}}{U_{k+1,n}^{-1}} \stackrel{\mathcal{L}}{=} \sum_{i=1}^{k} -\log(U_{i,k}) \stackrel{\mathcal{L}}{=} \sum_{i=1}^{k} -\log(1-U_{k-i+1,k}) = \sum_{i=1}^{k} -\log(1-U_i).$$

Comme $-\log(1-\cdot)$ est l'inverse de la fonction de répartition d'une loi exponentielle standard, on trouve le résultat annoncé.

6) C'est direct en utilisant la loi faible des grands nombres.
7) Pour $x>0$, on a

$$\begin{aligned}\mathbb{P}(n^{-\gamma}(x_F - X_{n,n}) \le x) &= \mathbb{P}(X_{n,n} \ge x_F - n^{\gamma}x) = 1 - F^n(x_F - n^{\gamma}x)\\ &= 1 - \exp\left\{n\log\left[1 - \frac{x^{-1/\gamma}}{n}L\left(\frac{n^{-\gamma}}{x}\right)\right]\right\}.\end{aligned}$$

Comme

$$\frac{x^{-1/\gamma}}{n}L\left(\frac{n^{-\gamma}}{x}\right) \sim \frac{1}{n}cx^{-1/\gamma} \to 0,$$

on en déduit que

$$\mathbb{P}(n^{-\gamma}(x_F - X_{n,n}) \le x) \to 1 - \exp(-cx^{-1/\gamma}),$$

pour $x > 0$. On déduit aisément de ce résultat que $X_{n,n} \xrightarrow{\mathrm{P}} x_F$.

8) On propose d'estimer γ par

$$\hat{\gamma}_n := \frac{1}{k} \sum_{i=1}^{k} \log \left(\frac{X_{n,n} - X_{n-i+1,n}}{X_{n,n} - X_{n-k,n}} \right).$$

Cet estimateur est connu dans la littérature sous le nom de *negative Hill estimator*.