Risques extrêmes et dépendance entre risques

Année 2016-2017

Exercices pour le jeudi 9 mars 2017

Exercice 1

On dispose de deux échantillons de taille $n = 20\,000$ dont les lois de probabilité sont différentes. [1]

1 - Pour chaque échantillon $(X_1\,, \ldots, X_n)$, on effectuera les traitements suivants.

(a) Construction d'un histogramme. Le nombre de classes sera choisi convenablement. Pour cela, plusieurs essais pourront être nécessaires afin de trouver un compromis raisonnable entre l'information conservée et la précision de l'estimation de la fréquence de chaque classe.

(b) Calculer les valeurs de la moyenne empirique, de l'écart-type empirique, du minimum, du maximum, de l'étendue et des quartiles empiriques. Les résultats seront présentés dans un tableau.

(c) Construction de la courbe de concentration empirique et calcul de l'indice de Gini empirique $\hat{G}_n$. On rappelle que cet indice est donné par

$$\hat{G}_n = \frac{2}{n}\,\frac{\sum_{i=1}^n i\, X_{i:n}}{\sum_{i=1}^n X_i} - \frac{n+1}{n}$$

où $(X_{1:n}\,, \ldots, X_{n:n})$ est l'échantillon ordonné associé à l'échantillon initial $(X_1\,, \ldots, X_n)$.

(d) On se propose d'estimer le paramètre ξ de la loi des extrêmes généralisée (GEV) dont la fonction de répartition s'écrit

$$G_\xi(x) = \exp\big(-(1+\xi\, x)^{-1/\xi}\big) \qquad\qquad x \in \mathbf{R}$$

avec la condition

$$1 + \xi\, x > 0$$

Lorsque $\xi = 0$ le prolongement par continuité donne la formule $G_0(x) = \exp\big(-e^{-x}\big)$ valable pour tout réel x, ce qui correspond à la loi de Gumbel.

- À partir de l'échantillon de taille n, commencer par constituer k échantillons de taille m, où k et m sont des entiers suffisamment grands et qui vérifient $n = k\,m$. Par exemple on pourra prendre $m = 100$ et $k = 200$, mais d'autres choix sont possibles. Cela permettra d'obtenir un échantillon de taille k du maximum d'un $m-$échantillon. On note cet échantillon par $(M_1\,, \ldots, M_k)$.

- Grâce à cet échantillon de taille k, construire un diagramme de probabilité de retour ("return level plot"). Autrement dit, reporter sur un même graphique les points

$$\big(\ln(y_u), \hat{Q}_k(u)\big)$$

où $y_u = -\ln(1-u)$ avec $u = i/(k+1)$ pour $i = 1, 2, \ldots, k$. Cela donne-t-il une indication sur le domaine d'attraction ?

(d) En utilisant l'estimateur de Pickand, proposer une estimation de ξ. On rappelle que cet estimateur est donné par

[1] Le premier échantillon est inclus dans le fichier "fic_exo_9_3_17c", le second dans le fichier "fic_exo_9_3_17d" (les valeurs sont données sur 2000 lignes et 10 colonnes).

$$R_{i,k} = \frac{1}{\ln 2} \ln\left(\frac{M_{k-i+1:k} - M_{k-2i+1:k}}{M_{k-2i+1:k} - M_{k-4i+1:k}}\right)$$

où $(M_{1:k}\,, \ldots, M_{k:k})$ est l'échantillon ordonné associé à l'échantillon initial $(M_1\,, \ldots, M_k)$. La méthode consiste à reporter les points $\big(i, R_{i,k}\big)$ sur un graphique et à repérer une valeur de i à partir de laquelle les points sont sensiblement alignés.

2 - On se propose maintenant d'identifier la loi de chaque échantillon de taille n.

(a) Construire les diagrammes quantiles-quantiles afin de déterminer la loi de probabilité d'où proviennent les données sachant que les seules lois possibles sont les suivantes :

- la loi exponentielle de paramètre β (autrement dit d'espérance mathématique $1/\beta$),

- la loi normale d'espérance mathématique m et d'écart-type σ,

- la loi log-normale de paramètres m et σ ; on rappelle que si une variable aléatoire W suit une telle loi, alors $\ln(W)$ suit la loi normale de paramètres m et σ.

- la loi de Pareto de paramètres x_0 et γ, tous deux strictement positifs. On rappelle que la fonction de répartition de cette loi s'écrit

$$F(x) = 1 - (x_0/x)^{\gamma} \;\text{ si } x > x_0, \;\; = 0 \text{ sinon.}$$

(b) Estimer les paramètres de la loi ainsi trouvée. Les estimateurs utilisés seront explicités et les valeurs seront présentées dans un tableau. Le cas échéant donner les compléments d'information qui vous paraissent utiles.

Exercice 2

En statistique, il arrive fréquemment que l'on connaisse un estimateur sans biais (ou asymptotiquement sans biais) pour un paramètre θ, ainsi que sa loi limite lorsque la taille n de l'échantillon tend vers l'infini. Si l'on désigne un tel estimateur par T_n, on aura donc $E(T_n) = \theta$ et

$$\sqrt{n}\big(T_n - \theta\big) \xrightarrow{loi} N(0, \sigma^2)$$

où σ^2 est supposé fini. Cependant, le paramètre d'intérêt peut être, non pas θ, mais une fonction $g(\theta)$ de ce paramètre. Si la fonction g est continue, $g(T_n)$ est un estimateur convergent pour $g(\theta)$, ce qui est un résultat intéressant, mais insuffisant pour déterminer des intervalles de confiance asymptotiques pour le paramètre $g(\theta)$. Il est heureusement possible, sous des conditions assez générales, de déterminer la loi limite de $g(T_n)$ lorsque n tend vers l'infini en utilisant la méthode Delta présentée brièvement dans le document "Delta method wiki". Appliquer cette méthode pour déterminer un intervalle de de confiance pour le paramètre β de la loi exponentielle lorsque n est grand.