Un estimateur de l'indice de valeur extrême du type "Pickands"

Laurent Gardes [1] & Stéphane Girard [2]

SMS/LMC, BP 53, 38041 Grenoble Cedex 9, France

RESUME – Nous proposons un nouvel estimateur de l'indice de valeur extrême ξ. Alors que de nombreux auteurs se sont focalisés sur le cas particulier $\xi > 0$, nous considérons ici le cas général $\xi \in \mathbb{R}$. Pour construire l'estimateur, nous utilisons une approche similaire à celle considérée par Pickands (1975). Nous établissons sa convergence en probabilité et sa convergence en loi. A l'aide de ce dernier résultat, nous proposons une méthode de débiaisage.

Mots-clés – Indice de valeur extrême, estimateur de Pickands.

ABSTRACT – We introduce a new estimator of the extreme value index ξ. Contrary to many others, our estimator is adapted to the general case $\xi \in \mathbb{R}$. Its construction is similar to the one of Pickands estimator (1975). Its weak consistency and its asymptotic distribution are established and a bias reduction method is proposed.

Key-words – Extreme value index, Pickands estimator.

1 Introduction

Considérons n variables aléatoires indépendantes $X_1, \ldots, X_n$ de fonction de répartition commune F et supposons qu'il existe deux suites $a_n > 0$ et b_n et un réel ξ tels que :

$$\lim_{n\to\infty} \mathrm{P}\left[\frac{\max(X_1,\ldots,X_n) - b_n}{a_n} \leq x\right] = G_\xi(x), \tag{1}$$

avec $G_\xi(x) = \exp[-(1+\xi x)_+^{-1/\xi}]$ si $\xi \neq 0$ et $G_0(x) = \exp[-\mathrm{e}^{-x}]$, où $y_+ = \max(0, y)$. Gnedenko (1943) montre que la relation (1) est satisfaite pour un très grand nombre de fonctions de répartition. G_ξ est la fonction de répartition de la loi des valeurs extrêmes et

[1] E-mail:`Laurent.Gardes@imag.fr`
[2] E-mail:`Stephane.Girard@imag.fr`

le paramètre ξ est appelé indice de valeur extrême. Ce paramètre caractérise le type de décroissance de la queue de distribution de F. Si $\xi > 0$, la décroissance est polynomiale (lois de Pareto, Burr, Student, log-gamma, etc ...), si $\xi = 0$, la décroissance est exponentielle (lois normale, log-normale, exponentielle, gamma, etc ...) et enfin, si $\xi < 0$, la fonction de répartition F admet un point terminal fini (loi uniforme, beta, etc ...). Disposer de la valeur de cet indice de valeur extrême est donc primordial dans de nombreuses applications comme par exemple en hydrologie (pour la construction de digues capable de résister aux crues centenales), en finances, dans les assurances, etc ...

2 Principe de construction de notre estimateur

On trouve dans la littérature de nombreux estimateurs de ce paramètre ξ, principalement dans le cas où $\xi > 0$ qui est fréquemment rencontré dans les applications. L'estimateur le plus connu dans ce cadre est l'estimateur de Hill (1975) défini par :

$$\hat{\xi}^H_{k,n} = \frac{1}{k}\sum_{i=1}^{k} \ln(X_{n-i+1,n}) - \ln(X_{n-k,n}), \text{ pour } k = 1, \dots, n-1,$$

où $X_{1,n} \leq \dots \leq X_{n,n}$ sont les statistiques ordonnées associées aux variables aléatoires $X_1, \dots, X_n$. On trouve aussi des estimateurs adaptés au cas $\xi < 0$ (voir par exemple Gardes, Chapitre 3 (2003)). Le cas général $\xi \in \mathbb{R}$ a été par contre moins étudié. Dekkers, Einmahl et de Haan (1989) ont adapté l'estimateur de Hill au cas $\xi \in \mathbb{R}$. Un autre estimateur a été proposé par Pickands (1975) :

$$\hat{\xi}^P_{k,n} = \frac{1}{\ln(2)} \ln\left(\frac{X_{n-k+1,n} - X_{n-2k+1,n}}{X_{n-2k+1,n} - X_{n-4k+1,n}}\right), \text{ pour } k = 1, \dots, \lfloor n/4 \rfloor,$$

où $\lfloor x \rfloor$ est la partie entière de x. On peut justifier la définition de cet estimateur de la manière suivante : soit $U(.)$ la fonction définie par :

$$U(x) = \left(\frac{1}{1-F(x)}\right)^{\leftarrow},$$

où le symbole "$\leftarrow$" représente l'inverse généralisé et soit $\varphi_t(.)$ la fonction définie par :

$$\varphi_t(x) = \int_1^x u^{t-1} du, \ x > 0, \ t \in \mathbb{R}.$$

De Haan (1984) montre que la relation (1) est vérifiée si et seulement si pour tout x, $y > 0$, $y \neq 1$,

$$\lim_{t\to\infty} \frac{U(tx) - U(t)}{U(ty) - U(t)} = \frac{\varphi_\xi(x)}{\varphi_\xi(y)}. \tag{2}$$

En remplaçant dans (2) U par $\hat{U}_n = [1/(1-\hat{F}_n)]^{\leftarrow}$ ($\hat{F}_n$ étant la fonction de répartition empirique), t par $n/(2k)$, x par $1/2$ et y par 2, on a pour n "suffisamment" grand,

$$2^{-\xi}\frac{X_{n-k+1,n} - X_{n-2k+1,n}}{X_{n-2k+1,n} - X_{n-4k+1,n}} = 1. \tag{3}$$

L'estimateur de Pickands $\hat{\xi}^P_{k,n}$ est alors la solution de l'équation (3). Les simulations montrent que cet estimateur possède un biais important. Une explication possible à cela est que $\hat{\xi}^P_{k,n}$, qui utilise l'information apportée par des distances entre deux statistiques ordonnées, n'utilise pas le maximum de l'échantillon ($X_{n,n}$), ce qui implique une perte d'information sur la queue de distribution. Nous avons donc proposé un estimateur tenant compte de l'information apportée par la distance entre une statistique ordonnée et $X_{n,n}$. Cet estimateur, noté $\hat{\xi}_{k,n}$ est défini comme étant la solution de l'équation en θ :

$$\left[\frac{\varphi_\theta(1/k')}{\varphi_\theta(1/k)}\right]\frac{X_{n-k+1,n} - X_{n,n}}{X_{n-k'+1,n} - X_{n,n}} = 1, \text{ pour } 1 < k' < k < n. \tag{4}$$

Cet estimateur, comme l'estimateur de Pickands, est utilisable quel que soit le signe de ξ et présente l'avantage d'être invariant par translation et changement d'échelle. Pour définir cet estimateur, nous avons généralisé le résultat (2) au cas où x et y tendent vers zéro lorsque t tend vers l'infini. Plus précisément, soit h_1 et h_2 deux fonctions positives (avec $h_2(t) \neq 1$ pour tout t) telles que :

$$\lim_{t\to\infty} h_1(t) = \lim_{t\to\infty} h_2(t) = 0, \ \lim_{t\to\infty} th_1(t) = \lim_{t\to\infty} th_2(t) = +\infty \text{ et } \lim_{t\to\infty}\frac{h_1(t)}{h_2(t)} = c > 0, c \neq 1.$$

Alors, si la relation (1) est vérifiée, on a :

$$\lim_{t\to\infty}\left\{\frac{\varphi_\xi[h_2(t)]}{\varphi_\xi[h_1(t)]}\right\}\frac{U[th_1(t)] - U[t]}{U[th_2(t)] - U[t]} = 1. \tag{5}$$

En remplaçant dans (5), U par $\hat{U}_n$, t par n, $h_1(t)$ par $1/k'$ et $h_2(t)$ par $1/k$, on obtient, pour n "suffisamment" grand, l'équation (4) définissant l'estimateur $\hat{\xi}_{k,n}$.

3 Résultats asymptotiques

Nous donnons ici les résultats de convergence en probabilité et de convergence en loi de l'estimateur $\hat{\xi}_{k,n}$. La connaissance de la distribution asymptotique de l'estimateur nous permet de proposer un estimateur débiaisé.

3.1 Convergence en probabilité

Théorème 1 – Supposons que F vérifie la relation (1). Si $k/k' \to c > 1$, $k \to \infty$ et $k/n \to \infty$ lorsque $n \to \infty$ alors, $\hat{\xi}_{k,n} \xrightarrow{\mathrm{P}} \xi$.

Remarque – Les hypothèses de la convergence en probabilité sont les mêmes que celles utilisées par Dekkers et de Haan (1989) pour démontrer la convergence en probabilité de l'estimateur de Pickands.

3.2 Convergence en loi

Les hypothèses nécessaires à la démonstration de la convergence en loi sont :

H1 – U est dérivable de dérivée U' avec $U'(x) = x^{\xi-1}\ell(x)$, où ℓ est une fonction à variations lentes.

H2 –

$$\varphi_\delta(k) \sup_{t\in[1,K_n]} \left| \frac{\ell(tN_n/K_n)}{\ell(N_n/K_n)} - 1 \right| \xrightarrow{\mathrm{P}} 0,$$

où $\delta = \min(-\xi, 1/2)$, $K_n = \bar{F}(X_{n-k+1,n})/\bar{F}(X_{n,n})$ et $N_n = 1/\bar{F}(X_{n,n})$, où $\bar{F} = 1-F$.

Remarques –
a) L'hypothèse H1 est une hypothèse du second ordre classique sur la fonction U (voir par exemple Dekkers et de Haan (1989)).
b) On peut montrer que, sous les hypothèses du Théorème 1, l'hypothèse H2 est vérifiée notamment pour des fonctions ℓ de la forme :

- $\ell(x) = \alpha + \theta x^{-\beta} + o(x^{-\beta})$ avec $\alpha > 0$, $\beta \geq 0$, $\theta \in \mathbb{R}$ et avec la condition supplémentaire, si β et/ou $\theta \neq 0$, $\varphi_\delta(k)(n/k)^{-\beta} \to 0$ lorsque $n \to \infty$.

- $\ell(x) = \alpha + \theta[\ln(x)]^{-\beta} + o[(\ln(x))^{-\beta}]$ avec $\alpha > 0$, $\beta \geq 0$, $\theta \in \mathbb{R}$ et avec la condition supplémentaire, si β et/ou $\theta \neq 0$, $\varphi_\delta(k)[\ln(n/k)]^{-\beta} \to 0$ lorsque $n \to \infty$.

- $\ell(x) = \theta[\ln(x)]^{-\beta}$ avec $\theta \in \mathbb{R}\backslash\{0\}$, $\beta \in \mathbb{R}$ et avec la condition supplémentaire, si $\beta \neq 0$, $\varphi_\delta(k)\ln(k)/\ln(n) \to 0$ lorsque $n \to \infty$.

Posons $V_k(\xi) = [(\ln(k) - 1)\mathbb{1}\{\xi \geq 0\} + 1]\varphi_\delta(k)$.

Théorème 2 – Sous les hypothèses du Théorème 1 (avec $k = ck'$) et sous les hypothèses

H1 et H2, on a pour tout $t \in \mathbb{R}$,

$$\lim_{n\to\infty} \mathrm{P}[V_k(\xi)(\hat{\xi}_{k,n} - \xi) \leq t] = \begin{cases} \exp(-\mathrm{e}^{-t}) & \text{si } \xi > 0, \\ \exp(-\mathrm{e}^{-t/2}) & \text{si } \xi = 0, \\ \exp\left[-\left(1 + \frac{t\xi\ln(c)}{c^{-\xi}-1}\right)^{-1/\xi}\right] & \text{si } \xi < 0 \text{ et } \xi > -1/2, \\ \Phi\left(-\frac{t\ln(c)}{2\xi\sqrt{c-1}}\right) & \text{si } \xi < -1/2. \end{cases}$$

où Φ est la fonction de répartition d'une loi normale centrée et réduite.

Remarques –
a) Dans le cas $\xi = -1/2$, on montre que $V_k(\xi)(\hat{\xi}_{k,n} - \xi)$ converge en loi vers une limite non dégénérée dont on ne peut déterminer explicitement l'expression.
b) L'estimateur $\hat{\xi}_{k,n}$ converge vers une loi des valeurs extrêmes pour $\xi > -1/2$ et vers une loi normale pour $\xi < -1/2$.

3.3 Débiaisage de l'estimateur

Posons

$$\mu(\xi) = \gamma \mathbb{I}\{\xi > 0\} + [1 - \Gamma(1-\xi)]\frac{1 - c^{-\xi}}{\xi \ln(c)}\mathbb{I}\{-1/2 < \xi < 0\},$$

où γ est la constante d'Euler et Γ est la fonction Gamma. Le Théorème 2 nous assure que $V_k(\xi)(\hat{\xi}_{k,n} - \xi)$ converge vers une loi d'espérance $\mu(\xi)$ pour $\xi \neq 0$ et $\xi \neq -1/2$. Nous proposons donc l'estimateur débiaisé défini par :

$$\hat{\xi}^*_{k,n} = \hat{\xi}_{k,n} - \frac{\mu(\hat{\xi}_{k,n})}{V_k(\hat{\xi}_{k,n})}.$$

L'efficacité de cette correction de biais apparaît sur des simulations (voir Gardes et Girard (2004)) que l'on ne présente pas ici faute de place.

Bibliographie

[1] A.L.M. Dekkers and L. de Haan. (1989) On the estimation of the extreme value index and large quantile estimation. *Annals of Statistics*, 17, p. 1795-1832.

[2] A.L.M. Dekkers, J.H.J. Einmahl and L. de Haan. (1989) A moment estimator for the index of an extreme value distribution. *Annals of Statistics*, 17, p. 1833-1855.

[3] L. Gardes. (2003) *Estimation d'une fonction quantile extrême.* Thèse de l'Université Montpellier 2.

[4] L. Gardes and S. Girard. (2004) A Pickands type estimator of the extreme value index. *Rapport de Recherche.*

[5] B. Gnedenko. (1943) Sur la distribution limite du terme maximum d'une série aléatoire. *Annals of Mathematics*, 44, p. 423-453.

[6] L. de Haan. (1984) Slow variation and characterization of domains of attraction. In *Statistical Extremes and Applications (J. Tiago de Oliveira ed.)*, p. 31-48.

[7] B.M. Hill. (1975) A simple general approach to inference about the tail of a distribution. *Annals of Statistics*, 3, p. 1163-1174.

[8] J. Pickands III. (1975) Statistical inference using extreme-order statistics. *Annals of Statistics*, 3, p. 119-131.