A Note on the Adaptive Choice of the Optimal Threshold in Extreme Value Analysis

Gomes, M. Ivette
Faculdade de Ciências de Lisboa (DEIO) and CEAUL, Universidade de Lisboa
Bloco C6, Piso 4, Cidade Universitária, Campo Grande
1749-016 Lisboa, Portugal
E-mail: ivette.gomes@fc.ul.pt

Pestana, Dinis
Faculdade de Ciências de Lisboa (DEIO) and CEAUL, Universidade de Lisboa
Bloco C6, Piso 4, Cidade Universitária, Campo Grande
1749-016 Lisboa, Portugal
E-mail: dinis.pestana@fc.ul.pt

Abstract

The main objective of Statistics of Extremes is the prediction of rare events, and thus the need for an adequate estimation of parameters related to natural disasters. The primary parameter is the extreme value index (EVI). One of the most recent and general approaches is the semi-parametric one, where it is merely assumed that the model underlying the random sample is in the domain of attraction of a (unified) extreme value distribution. The EVI estimation is then based on the largest k order statistics in the sample or on the excesses over a high level u. The question that has been often addressed in the practical applications of Extreme Value Theory is the choice of either k or u. A great variety of semi-parametric EVI-estimators have the same type of problems, consistency for intermediate ranks, high variance for small values of k, and high asymptotic bias for large values of k. In this paper we shall concentrate on heavy tails, i.e., a positive EVI. The most common methods of adaptive choice of the threshold k, among which we mention the bootstrap methods, are based on the minimization of some kind of mean squared error estimator. Here we advance with a methodology based on bias properties for the selection of the optimal sample fraction. The methodology depends on a tuning parameter, and we provide a choice for such a tuning parameter. In order to achieve our objectives we derive the asymptotic behavior of an adequate linear combination of the scaled log-spacings, suggest the use of two alternative auxiliary statistics, and draw some overall comments.

1 Introduction

Let $X_1, X_2, \ldots, X_n$ be independent, identically distributed (i.i.d.) random variables (r.v.'s) from an underlying population with unknown distribution function (d.f.) $F(\cdot)$, and let $X_{1:n} \le \cdots \le X_{n:n}$ denote the associated sample of ascending order statistics (o.s.). Then, if there exist attraction coefficients $\{a_n > 0\}_{n\ge1}$ and $\{b_n\}_{n\ge1}$, and a non-degenerate d.f. $G(\cdot)$ such that the distribution of the maximum $X_{n:n}$, linearly normalized, converges to $G(\cdot)$, the d.f. $G(\cdot)$ is of the type of an *Extreme Value* (EV) d.f.,

$$(1) \quad G_\gamma(x) \equiv EV_\gamma(x) := \exp\big(-(1+\gamma x)^{-1/\gamma}\big),\ 1+\gamma x > 0,\ \gamma \in \mathbb{R}.$$

We then say that F belongs to the domain of attraction of EV_γ, and write $F \in \mathcal{D}_{\mathcal{M}}(EV_\gamma)$. The main objective of *Statistics of Extremes* is the prediction of rare events, and thus the need for an adequate estimation of parameters related to natural "disasters". The primary parameter is however the *extreme value index* (EVI), the shape parameter γ in (1). One of the most recent and general approaches is the semi-parametric one, where it is merely assumed that $F \in \mathcal{D}_{\mathcal{M}}(EV_\gamma)$, the estimation of γ being then based on the largest k o.s. in the sample or on the excesses over a high level u. The question that has been often addressed in the practical applications of *Extreme Value Theory* (EVT) is the choice of either k or u.

A great variety of semi-parametric EVI-estimators, $\hat{\gamma}_{n,k} \equiv \hat{\gamma}_n(k)$, have the same type of problems — consistency for intermediate ranks, i.e., we need to have $k = k_n \to \infty$, and $k_n/n \to 0$, as $n \to \infty$, high variance for small values of k, and high bias for large values of k. In this paper we shall concentrate on Hill's estimator (Hill, 1975),

$$(2) \quad \hat{\gamma}^H_{n,k} \equiv \hat{\gamma}^H_n(k) := \frac{1}{k}\sum_{i=1}^{k}\left\{\ln\frac{X_{n-i+1:n}}{X_{n-k:n}}\right\} = \frac{1}{k}\sum_{i=1}^{k} U_i, \quad U_i := i\left\{\ln\frac{X_{n-i+1:n}}{X_{n-i:n}}\right\}, \quad 1 \le i \le k.$$

Let us put $U(t) := F^{\leftarrow}(1-1/t) := \{\inf x : F(x) > 1-1/t\},\ t > 1$. We shall assume that there exists a function $A(t)$ of constant sign and going to 0 as $t \to \infty$, such that

$$(3) \quad \lim_{t\to\infty} \frac{U(tx)/U(t) - x^\gamma}{A(t)} = x^\gamma \frac{x^\rho - 1}{\rho},$$

for every $x > 0$, where ρ (≤ 0) is a *second order parameter*. Then we have, asymptotically, the following distributional representation for the Hill estimator in (2),

$$(4) \quad \hat{\gamma}^H_{n,k} \stackrel{d}{=} \gamma + \frac{\gamma P_k}{\sqrt{k}} + \frac{1}{1-\rho}A(n/k) + o_p(1/\sqrt{k}) + o_p(A(n/k)),$$

where P_k is a standard Normal r.v. (de Haan and Peng, 1998). The limit in (3) must be of the stated form, and $|A(t)| \in RV_\rho$ (from Theorem 1.9 of Geluk and de Haan, 1987), where RV_a stands for the class of regularly varying functions at infinity with an index of regular variation a, i.e., positive measurable functions g such that $g(tx)/g(t) \to x^a$, as $t \to \infty$ and for all $x > 0$ (for details on regular variation, see Bingham *et al.*, 1987). It thus follows that if k is such that $\sqrt{k}A(n/k) \to \lambda$, finite, as $n \to \infty$, then $\sqrt{k}\big(\gamma^H_{n,k} - \gamma\big) \xrightarrow{d} N\big(\lambda/(1-\rho), \gamma^2)\big)$ as $n \to \infty$, i.e. we may have a non-null asymptotic bias. Moreover, if $\sqrt{k}A(n/k) \to \infty$, $\big(\hat{\gamma}^H_{n,k} - \gamma\big)/A(n/k) \xrightarrow{p} 1/(1-\rho)$. So we see that whether the limit of $\sqrt{k}A(n/k)$ is either finite or infinite, we can always specify the rate at which $\hat{\gamma}^H_{n,k}$ converges to γ.

A possible substitute for the *mean square error* (MSE) is (cf. equation (4))

$$AMSE\big(\hat{\gamma}^H_{n,k}\big) := \mathbb{E}\left(\frac{\gamma\ P_k}{\sqrt{k}} + \frac{1}{1-\rho}A(n/k)\right)^2 = \frac{\gamma^2}{k} + \frac{A^2(n/k)}{(1-\rho)^2},$$

depending on n and k. It is then possible to see that if $AMSE\big(\hat{\gamma}^H_{n,k}\big)$ is minimal at $k_0 = k_0(n)$, then $\sqrt{k_0}\ A(n/k_0) \to \gamma(1-\rho)/\sqrt{-2\rho}$, as $n \to \infty$.

Since we often need to impose the extra condition that the function $A(t)$ in (3) can be chosen as a power of t, i.e.,

$$(5) \quad A(t) = C\ t^\rho, \text{ with } C \neq 0 \text{ and } \rho < 0,$$

we shall assume throughout the paper that we are working in Hall-Welsh class of d.f.'s (Hall, 1982; Hall and Welsh, 1985), with a right tail of the type

$$1 - F(x) = \alpha x^{-1/\gamma}\left(1 + \beta x^{\rho/\gamma} + o(x^{\rho/\gamma})\right), \text{ as } x \to \infty, \quad \alpha > 0,\ \gamma > 0,\ \beta \neq 0,\ \rho < 0,$$

which is equivalent to (3) jointly with (5) and with $C = \gamma\rho\beta\alpha^\rho$.

The most common methods of adaptive choice of the threshold k are based on the minimization of some kind of MSE-estimator. We mention the bootstrap methodology for the selection of $k_0(n)$

(Hall, 1990; Draisma *et al.*, 1999; Danielsson *et al.*, 2001; Gomes and Oliveira, 2001). The adaptive choice of the threshold in Hall and Welsh (1985), as well as the methodology developed in Beirlant *et al.* (1996a, 1996b) have also the objective of minimization of a MSE-estimator. Drees and Kaufmann (1998) proposal, as well as the methodology in Guillou and Hall (2001), for the selection of the optimal sample fraction are based on bias properties. In this note, we are particularly interested in the approach followed in the paper by Guillou and Hall. In Section 2, we describe the method, and provide a choice for the tuning parameter under play. Section 3 is devoted to the asymptotic behavior of a linear combination of the scaled log-spacings. Finally, in Section 4, we suggest the use of two alternative auxiliary statistics, and draw some overall comments.

2 An adaptive choice of the threshold based on bias behaviour

Guillou and Hall (2001) suggested an ingenious approach for choosing the threshold when fitting the Hill estimator of a tail exponent to extreme value data. Here, we merely propose a more objective choice of the nuisance parameter, c_{crit}, there considered. The argument is the following: for the Hill estimator we have the validity of the distributional representation (4), and the asymptotic mean squared error of $\sqrt{k}\ \gamma^H_{n,k}$ is equal to $\gamma^2(1+\mu_k^2)$ where

$$\mu_k = \sqrt{k}\ A(n/k)/(\gamma(1-\rho)).$$

Then, the value k_0 of k that is optimal in the sense of minimizing the asymptotic MSE of $\gamma^H_{n,k}$ may be taken as a solution of

$$\mu^2_{k_0} = 1/\sqrt{-2\rho}.$$

Similarly to what happens in the bootstrap methodology of Draisma *et al.* (1999), Danielsson *et al.* (2001) and Gomes and Oliveira (2001), the idea underlying Guillou and Hall's paper is to replace the r.v. $\sqrt{k}\{\gamma^H_{n,k}-\gamma\}/\gamma$ by a statistic which also converges towards 0 and has similar bias properties. Such a statistic is merely a linear combination of the log-spacings U_i, $1 \le i \le k$, in (2), with a null mean value. More precisely, the main role in this diagnostic technique is played by the statistics

$$T_k \equiv T_n(k) := \sqrt{\frac{3}{k^3}}\ \sum_{i=1}^{k} w_i\ U_i / \sum_{i=1}^{k} U_i,$$

with $w_i = \mathrm{sgn}(k-2i+1)|k-2i+1|$, $1 \le i \le k$, together with a moving average of its squares, which dampens stochastic fluctuations of $T_n(k)$, the set of statistics

$$Q_k \equiv Q_n(k) := \left\{ \frac{1}{2[k/2]+1} \sum_{i=k-[k/2]}^{k+[k/2]} T_i^2 \right\}^{1/2}.$$

3 The asymptotic behaviour of a linear combination of the scaled log-spacings

Let us consider the statistic

$$\sum_{i=1}^{k} w_i\ U_i, \quad w_i = \mathrm{sgn}(k-2i+1)|k-2i+1|,\ 1 \le i \le k.$$

Since $\sum_{i=1}^k w_i = 0$, we obviously have $\mathbb{E}\big(\sum_{i=1}^k w_i\ U_i\big) = 0$. But from the fact that, under the second order framework in (3), the U_i's are approximately exponential and given by

$$U_i = \gamma\ E_i + i\ A(n/k)\,(k/i))^{\rho}\left(e^{\rho E_i/i} - 1\right)(1+o_p(1))/\rho,$$

where $\{E_i\}_{i\geq 1}$ is a sequence of i.i.d. exponential r.v.'s, we may have a non-null dominant component of asymptotic bias, related to the behaviour of the term $A(n/k)\ \sum_{i=1}^{k}\left(\frac{i}{k}\right)^{-\rho} w_i\ E_i$. For the choice $w_i = \mathrm{sgn}(k-2i+1)|k-2i+1|,\ 1 \leq i \leq k$, and with the usual notation $a_n \sim b_n$ whenever $a_n/b_n \to 1$, as $n \to \infty$, we have

$$\sum_{i=1}^{k}\left(\frac{i}{k}\right)^{-\rho} w_i \sim \frac{\rho}{(1-\rho)(2-\rho)}\, k^2 \qquad \text{and} \qquad \sum_{i=1}^{k} w_i^2 \sim \frac{k^3}{3}.$$

Hence, we can write, for intermediate k,

$$T_n(k) \stackrel{d}{=} P_k + \frac{\sqrt{3}\rho}{\gamma(1-\rho)(2-\rho)}\sqrt{k}\ A(n/k)\ (1+o_p(1)) = P_k + \frac{\sqrt{3}\ \rho\ \mu_k}{2-\rho}\ (1+o_p(1))$$

where P_k is the standard normal r.v. in (4). If we recall that the optimal choice of k should be such that $\mu_{k_0} \underset{n\to\infty}{\longrightarrow} 1/\sqrt{-2\rho}$, a sensible adaptive choice of k by means of $T_n(k)$ should be

$$k_0^T := \inf\{k:\ |T_n(j)| \geq c_T,\ \forall j \geq k\}, \quad c_T = \frac{-\sqrt{3}\ \rho}{(2-\rho)\sqrt{-2\rho}}.$$

We have asymptotically,

$$\mathbb{E}\big(T_n^2(k)\big) = 1 + \frac{3\rho^2\mu_k^2}{(2-\rho)^2} \qquad \text{and} \qquad \mathbb{E}\big(Q_n(k)\big) = 1 + \frac{3\rho^2\ (3^{2(1-\rho)}-1)}{2^{2(2-\rho)}(1-\rho)(2-\rho)^2}\ \mu_k^2.$$

The adaptive choice of k_0 through the help of the statistic $Q_n(k)$ should be based on the critical value

$$c_Q = 1 + \frac{3\rho^2\ (3^{2(1-\rho)}-1)}{2^{2(2-\rho)}(1-\rho)(2-\rho)^2\sqrt{-2\rho}},$$

and the choice

$$k_0^Q := \inf\big\{k:\ |Q_n(j)| \geq c_Q,\ \forall j \geq k\big\}.$$

In Figure 1 we picture both the values of c_T and c_Q as a function of $|\rho|$. Whereas c_T converges towards 0 both as $\rho \to 0$ and $\rho \to -\infty$, attaining a maximum value, equal to 0.43 for $\rho = -2$, c_Q converges towards 1 as $\rho \to 0$, diverging to $+\infty$ as $\rho \to -\infty$. The critical value c_Q depends strongly on the value of ρ, and it thus seems not to be advisable in practice, where all values of ρ may appear, to consider a fixed c_Q, equal for instance to c_{crit}=1.25, as suggested by Guillou and Hall (2001). It seems we should choose the critical level as a function of an adequate estimate of the second order parameter ρ. Notice however that for $|\rho| \leq 2$ we have not a great variation in c_Q, as a function of ρ.

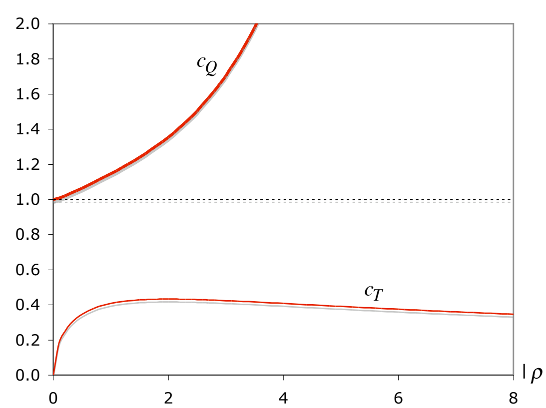

Figure 1: Values of c_T and c_Q as a function of $|\rho|$.

4 The use of alternative auxiliary statistics

We here suggest the use of two alternative auxiliary statistics: either the statistic

$$T_n^*(k) = T_n^2(k) \quad \text{or} \quad Q_n^*(k) := \left\{ \frac{1}{k} \sum_{i=1}^{k} T_n^2(i) \right\}^{1/2}.$$

The main reason for these choices is that we have asymptotically

$$\mathbb{E}\big(T_n^*(k)\big) = 1 + \frac{3\rho^2}{(2-\rho)^2}\,\mu_k^2 \qquad \text{and} \quad \mathbb{E}\big(Q_n^*(k)\big) = 1 + \frac{3\rho^2}{4(1-\rho)(2-\rho)^2}\,\mu_k^2,$$

and the choices

$$k_0^{T^*} := \inf\{k:\ |T_n^*(j)| \geq c_{T^*},\ \forall j \geq k\} \quad \text{and} \quad k_0^{Q^*} := \inf\{k:\ |Q_n^*(j)| \geq c_{Q^*},\ \forall j \geq k\},$$

where

$$c_{T^*} := 1 + \frac{3\rho^2}{\sqrt{-2\rho}\,(2-\rho)^2} \quad \text{and} \quad c_{Q^*} := 1 + \frac{3\rho^2}{4\sqrt{-2\rho}\,(1-\rho)(2-\rho)^2},$$

respectively. From a theoretical point of view, the main advantage of these choices seems to be the fact that both c_{T^*} and c_{Q^*} converge towards 1 either as $\rho \to 0$ or as $\rho \to -\infty$. More than that: c_{T^*} attains a maximum value, equal to 1.487 for $\rho = -6$ and c_{Q^*} attains a maximum equal to 1.032 for $\rho = -1.59$, being thus "almost invariant" with ρ. In Figure 2 we picture both the values of c_{T^*} and c_{Q^*} as a function of $|\rho|$.

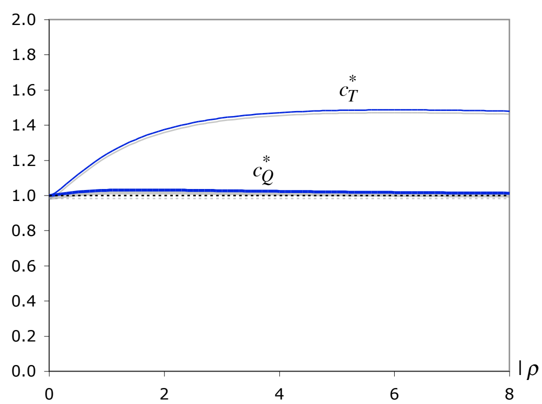

Figure 2: Values of c_{T^*} and c_{Q^*} as a function of $|\rho|$.

Figure 2, as well as a small-scale Monte-Carlo simulation, show that it seems preferable to work with the statistic $Q_n^*(k)$ rather than with the other statistics — a smaller range means better final results, as expected, and possibly with a c_{crit} independent on ρ, and equal to 1. Notice also that whereas $Q_n(k)$ is defined only for $k+[k/2] \leq n-1$, $Q_n^*(k)$ can be defined for all $1 \leq k \leq n-1$, another point in favour of $Q_n^*(k)$.

REFERENCES

1. Beirlant, J., Vynckier, P. and J.L. Teugels (1996a). Excess function and estimation of the extreme-value index. *Bernoulli* **2**, 293-318.

2. Beirlant, J., Vynckier, P. and J.L. Teugels (1996b). Tail index estimation, Pareto quantile plots, and regression diagnostics. *J. American Statististical Association* **91**, 1659-1667.

3. Bingham, N., Goldie, C.M. and J.L. Teugels (1987). *Regular Variation.* Cambridge Univ. Press, Cambridge.
4. Danielsson, J., de Haan, L., Peng, L. and C.G. de Vries (2001). Using a bootstrap method to choose the sample fraction in the tail index estimation. *J. Multivariate Analysis* **76**:2, 226-248.
5. Draisma, G., de Haan, L., Peng, L. and T.T. Pereira (1999). Estimating the index of a stable distribution. *Extremes* **2**:4, 367-404.
6. Drees, H. and E. Kaufmann (1998). Selecting the optimal sample fraction in univariate extreme value estimation. *Stochastic Processes and Applications* **75**, 149-172.
7. Geluk, J. and L. de Haan (1987). *Regular Variation, Extensions and Tauberian Theorems.* CWI Tract 40, Center for Mathematics and Computer Science, Amsterdam, Netherlands.
8. Gomes, M.I. and O. Oliveira (2001). The bootstrap methodology in statistical extremes — the choice of the optimal sample fraction. *Extremes* **4**:4, 331-358.
9. Guillou, A. and Hall, P. (2001). A diagnostic for selecting the threshold in extreme-value analysis. *J. Royal Statistical Society* **B 63**, 293- 305.
10. Haan, L. de and L. Peng (1998). Comparison of tail index estimators. *Statistica Neerlandica* **52**, 60-70.
11. Hall, P. (1982). On some simple estimates of an exponent of regular variation. *J. Royal Statistical Society* **B44**, 37-42.
12. Hall, P. (1990). Using bootstrap to estimate mean squared error and selecting parameter in nonparametric problems. *J. Multivariate Analysis* **32**, 177-203.
13. Hall, P. and A.H. Welsh (1985). Adaptive estimates of parameters of regular variation. *Annals of Statistics* **13**, 331-341.
14. Hill, B.M. (1975). A simple general approach to inference about the tail of a distribution. *Annals of Statistics* **3**, 1163-1174.

RÉSUMÉ

L'objectif principal de la Statistique des Valeurs Extrêmes est la prédiction d' événements rares; il en résulte le besoin d'estimation appropriée des paramètres connexes aux désastres naturels. Le paramètre principal est l'index de valeur extrême (Extreme Value Index, EVI). La récente méthodologie demi-paramétrique, où l'on présume simplement que le modèle sous-jacent à l'échantillon aléatoire est dans le domaine d'attraction de la distribution (unifiée) de valeurs extrêmes, est très générale. L'estimation de l'EVI a donc comme point de départ les k statistiques d'ordre les plus grandes dans l'échantillon, ou les excès d'un niveau élevé u. Dans l'application pratique de la Théorie des Valeurs Extrêmes, une des questions envisagées très souvent est le choix soit de k soit de u. Une grande variété d'estimateurs demi-paramétriques de l'EVI, présentent le même type de problèmes, notamment consistance seulement pour les rangs intermédiaires, variance élevée en cas de petites valeurs de k, et biais asymptotique élevé en cas de grandes valeurs de k. Dans cette communication nous envisageons le cas de queues lourdes, c'est à dire EVI positif. Les méthodes les plus usuelles de choix adaptatif du seuil k, parmi lesquelles nous mentionnons les méthodes bootstrap, sont basées dans la minimisation d'un estimateur quelconque de l'erreur carrée moyenne. Nous proposons une méthode pour la sélection de la fraction d'échantillon optimale qui a pour fondement des propriétés du biais. Cette méthode dépend d'un paramètre d'ajustement, et nous proposons un choix pour ce paramètre d'ajustement. Pour atteindre nos objectifs, nous dérivons le comportement asymptotique d'une combinaison linéaire appropriée des log-espacements, nous suggérons le recours à deux statistiques supplétives alternatives, et nous présentons quelques réflexions générales.