PORTUGALIAE MATHEMATICA
Vol. 59 Fasc. 4 – 2002
Nova Série

GENERALIZED JACKKNIFE SEMI-PARAMETRIC ESTIMATORS OF THE TAIL INDEX *

M. Ivette Gomes, M. João Martins and Manuela Neves

Presented by J.P. Dias

Abstract: In this paper we shall consider a natural *Generalized Jackknife* estimator of the *index of regular variation* γ of a heavy right-tail $1-F(x)$, as $x \to +\infty$, associated to any adequate semi-parametric estimator of γ. Such an estimator is merely a linear combination of the original estimator at two different levels. We also study such a general linear combination asymptotically, and an illustration of how these results work in practice is provided.

1 – Introduction and overview of the subject

In *Statistical Extreme Value Theory* we are mainly interested in the estimation of parameters of rare events, the basic parameter being the *tail index* $\gamma = \gamma(F)$, directly related to the tail weight of the model $F(\cdot)$. The tail index γ is the shape parameter in the unified *Extreme Value* (EV) distribution function (d.f.),

$$G(x) \equiv G_\gamma(x) := \begin{cases} \exp\Big\{-(1+\gamma x)^{-1/\gamma}\Big\}, & 1+\gamma x > 0 \quad \text{if} \quad \gamma \neq 0, \\ \exp\Big\{-\exp(-x)\Big\}, & x \in \mathbb{R} \qquad\quad\ \text{if} \quad \gamma = 0 \end{cases} \tag{1.1}$$

(Gnedenko, 1943). This d.f. appears as the non-degenerate limiting d.f. of the sequence of maximum values, linearly normalized. Whenever there is a non-degenerate limit of the sequence of normalized maximum values towards a random

Received: May 9, 2001.
AMS Subject Classification: Primary 62G05, 62G32, 62E20; Secondary 62F40, 65C05.
Keywords and Phrases: Statistical Theory of Extremes; Semi-parametric estimation; Jackknife methodology.
* Research partially supported by POCTI / FCT / FEDER.

variable (r.v.), then necessarily with d.f. given by (1.1), we say that F is in the domain of attraction of G_γ, and write $F \in D(G_\gamma)$. As usual, let us put

$$U(t) := \begin{cases} 0 & t \le 1 \\ F^{\leftarrow}(1-1/t) & t > 1 \end{cases}, \tag{1.2}$$

for the quantile function at $1-1/t$, where $F^{\leftarrow}(t) = \inf\{x\colon F(x) \ge t\}$ is the generalized inverse function of $F(\cdot)$.
Then, for heavy tails ($\gamma > 0$) we have

$$F \in D(G_\gamma) \quad \text{iff} \quad 1 - F \in RV_{-1/\gamma} \quad \text{iff} \quad U \in RV_\gamma\ , \tag{1.3}$$

where RV_α stands for the class of regularly varying functions at infinity with index of regular variation equal to α, i.e., functions $g(\cdot)$ with infinite right end-point, and such that $\lim_{t\to\infty} g(tx)/g(t) = x^\alpha$, for all $x > 0$. The conditions in (1.3) characterize completely the first order behaviour of $F(\cdot)$ (Gnedenko, 1943; de Haan, 1970). The second order theory has been worked out in full generality by de Haan and Stadtmüller (1996). Indeed, for a large class of models there exists a function $A(t)$ of constant sign for large values of t, such that

$$\lim_{t\to\infty} \frac{U(tx)/U(t) - x^\gamma}{A(t)} = x^\gamma\, \frac{x^\rho - 1}{\rho} \tag{1.4}$$

for every $x > 0$, where ρ (≤ 0) is a *second order parameter*, which also needs to be properly estimated from the original sample. The right side of (1.4) is to be interpreted as $x^\gamma \ln x$ whenever $\rho = 0$. The limit function in (1.4) must be of the stated form, and $|A(\cdot)| \in RV_\rho$ (Geluk and de Haan, 1987).

In this paper, and with $X_{i:n}$ denoting the i-th ascending order statistic (o.s.), $1 \le i \le n$, we are going to work with any general semi-parametric estimator of the tail index,

$$\hat\gamma_n(k) \;=\; \Phi(X_{n-k:n}, X_{n-k+1:n}, ..., X_{n:n}) \tag{1.5}$$

for which the distributional representation

$$\hat\gamma_n(k) \overset{d}{=} \gamma + \frac{\sigma}{\sqrt{k}}\, Z_n + b\, A(n/k) + o_p(A(n/k)) + o_p(1/\sqrt{k})\,, \quad b\in\mathbb{R},\ \ \sigma>0 \tag{1.6}$$

holds for every intermediate k, i.e. a sequence $k = k_n$ such that

$$k_n \to \infty, \;\; k_n/n \to 0 \quad \text{as} \;\; n \to \infty\ . \tag{1.7}$$

The r.v. Z_n is asymptotically standard Normal, being a suitable linear combination of r.v.'s P_n and Q_n, where $P_n \overset{d}{=} \frac{1}{\sqrt{k}}(\sum_{i=1}^k W_i - k)$, and $Q_n \overset{d}{=} \frac{1}{\sqrt{k}}(\sum_{i=1}^k W_i^2 - 2k)$, with $\{W_i\}_{i\geq 1}$, a sequence of i.i.d. unit exponential r.v.'s. Since $E[W_i^r] = r!$, we have $Var[W_i] = 1$, $Var[W_i^2] = 20$, and consequently $Var[P_n] = 1$, $Var[Q_n] = 20$, $Cov[P_n, Q_n] = 4$.

Among the estimators with the above mentioned properties we mention Hill's estimator (Hill, 1975), with the functional form

$$\hat{\gamma}_n^H(k) := \frac{1}{k}\sum_{i=1}^{k}\Big[\ln X_{n-i+1:n} - \ln X_{n-k:n}\Big] , \tag{1.8}$$

and the Moment estimator (Dekkers *et al.*, 1989),

$$\hat{\gamma}_n^M(k) := M_n^{(1)}(k) + 1 - \frac{1}{2}\left\{1 - \frac{\Big(M_n^{(1)}(k)\Big)^2}{M_n^{(2)}(k)}\right\}^{-1} , \tag{1.9}$$

with

$$M_n^{(j)}(k) := \frac{1}{k}\sum_{i=1}^{k}\Big\{\ln X_{n-i+1:n} - \ln X_{n-k:n}\Big\}^j , \quad j = 1, 2 . \tag{1.10}$$

For heavy tails, $\gamma > 0$, σ and b in the representation (1.6) are to be replaced for the Hill estimator by $\sigma_H = \gamma$ and $b_H = \frac{1}{1-\rho}$, and for the Moment estimator, by $\sigma_M = \sqrt{\gamma^2+1}$ and $b_M = \frac{\gamma(1-\rho)+\rho}{\gamma(1-\rho)^2}$. Also $Z_n^H = P_n$ and $Z_n^M = \frac{1}{\sqrt{\gamma^2+1}}\Big(\frac{Q_n}{2} + (\gamma-2)P_n\Big)$. The distributional representation (1.6) for the Hill and the Moment estimators may be found in Dekkers *et al.*, 1989.

The *Generalized Jackknife* statistic was introduced by Gray and Shucany, 1972, with the purpose of bias reduction. Let $T_n^{(1)}$ and $T_n^{(2)}$ be two biased estimators of ξ with similar bias properties such that $Bias(T_n^{(i)}) = \phi(\xi)\, d_i(n)$, $i = 1, 2$; then

$$T_n^G := \frac{T_n^{(1)} - qT_n^{(2)}}{1-q}, \quad \text{with } q = q_n = d_1(n)/d_2(n) ,$$

is an unbiased estimator of ξ. In Extreme Value Theory, one usually has information about the asymptotic bias of the estimators, so one can use this information to build new estimators with a reduced asymptotic bias. We propose several Jackknife estimators for γ based on $\hat{\gamma}_n(\cdot)$ at two different levels. These estimators will be introduced and studied asymptotically in sections 2 and 3. Finally, in section 4, we illustrate their properties for finite samples.

2 – Asymptotic properties of the estimator at different levels, and the Generalized Jackknife statistics

For the estimators under study, i.e. estimators for which (1.6) holds, we have

$$\sqrt{k}\,[\widehat{\gamma}_n(k)-\gamma] \xrightarrow{d} N(\lambda b,\sigma^2), \quad \text{as } n\to\infty\,, \tag{2.1}$$

provided we choose k such that $\sqrt{k}\,A(n/k)\to\lambda$, finite, as $n\to\infty$. In this situation we say that $\widehat{\gamma}_n(k)$ is asymptotically normal with asymptotic bias $Bias_\infty[\widehat{\gamma}_n(k)] = bA(n/k)$ and asymptotic variance $Var_\infty[\widehat{\gamma}_n(k)] = \sigma^2/k$. On the other side, if $\sqrt{k}\;|A(n/k)|\to\infty$, then $\frac{\widehat{\gamma}_n(k)-\gamma}{A(n/k)} \xrightarrow{p} b$.

Under the validity of (1.6), and whenever $\sqrt{k}\,A(n/k)\to\lambda$, finite, as $n\to\infty$, we have an *Asymptotic Mean Square Error* (*AMSE*) given by

$$AMSE[\widehat{\gamma}_n(k)] = \frac{\sigma^2}{k} + b^2A^2(n/k)\;.$$

Then, since there exists a function $s(\cdot)\in RV_{2\rho-1}$, such that, as $t\to\infty$, $A^2(t) = \int_t^{+\infty} s(u)\,du\,(1+o(1))$, (Proposition 1.7.3 of Geluk and de Haan, 1987), we have, for $b\neq 0$ and $r:=\frac{n}{k}$, and by Lemma 2.9 of Dekkers and de Haan, 1993, $\inf_{r>0}\left\{r\,\frac{\sigma^2}{n} + b^2A^2(r)\right\} = \inf_{r>0}\left\{r\,\frac{\sigma^2}{n} + b^2\int_r^{+\infty} s(u)\,du\,(1+o(1))\right\} = \inf_{r>0}\left\{r\,\frac{\sigma^2}{n} + b^2\int_r^{+\infty} s(u)\,du\right\}(1+o(1))$; hence $\inf_{r>0}\left\{r\,\frac{\sigma^2}{n} + b^2A^2(r)\right\} = \int_0^{\frac{\sigma^2}{b^2n}} s^{\leftarrow}(u)\,du\,(1+o(1))$, and $r_0 = \arg\inf_{r>0}\left\{r\,\frac{\sigma^2}{n} + b^2A^2(r)\right\} = s^{\leftarrow}(\frac{\sigma^2}{b^2n})\,(1+o(1))$. We thus have

$$k_0(n) := \arg\inf_k AMSE[\widehat{\gamma}_n(k)] \;=\; \frac{n}{s^{\leftarrow}\left(\frac{\sigma^2}{b^2n}\right)}\,(1+o(1))\;, \tag{2.2}$$

and the *AMSE* of $\widehat{\gamma}_n(k)$ at the optimal level, i.e. the *AMSE* of $\widehat{\gamma}_{n,0} := \widehat{\gamma}_n(k_0(n))$ is such that

$$\lim_{n\to\infty} AMSE[\varphi(n)\,\widehat{\gamma}_{n,0}] \;=\; \frac{2\rho-1}{2\rho}\,(\sigma^2)^{-\frac{2\rho}{1-2\rho}}\,(b^2)^{\frac{1}{1-2\rho}} \;:=\; LMSE[\widehat{\gamma}_{n,0}]$$

where $\varphi(n) = (n/s^{\leftarrow}(1/n))^{1/2}$.

If we consider the original estimator, $\widehat{\gamma}_n(k)$, and two different intermediate levels k_1 and k_2, with $k_1<k_2$, $k_2-k_1\to\infty$, as $n\to\infty$, we have the asymptotic representations:

$$\widehat{\gamma}_n(k_j) \stackrel{d}{=} \gamma + \frac{\sigma}{\sqrt{k_j}}\,Z_{n,j} + b\,A(n/k_j) + o_p(A(n/k_j)) + o_p(1/\sqrt{k_j})\,, \qquad j=1,2\;,$$

where $(Z_{n,1}, Z_{n,2})$ is asymptotically Bivariate Normal with null mean and covariance matrix $\Sigma_{1,2} = [\sigma_{ij}]$, where $\sigma_{11} = \sigma_{22} = 1$, and $\sigma_{12} = \sigma_{21} = \sqrt{\frac{k_1}{k_2}}$. Let us put $\theta := \lim_{n\to\infty} \frac{k_1}{k_2}$. Since the function $|A(\cdot)|$ is of regular variation of index ρ, we have

$$A(n/k_2) = \left(\frac{k_1}{k_2}\right)^{\rho} A(n/k_1)\,(1 + o(1)), \qquad \text{as} \;\; n \to \infty, \quad \text{for } 0 < \theta \le 1 \;;$$

if $\theta = 0$, then the previous statement is only valid under the more restrictive situation $A(t) = Ct^{\rho}(1+o(1))$, which holds for Hall's class of distributions (Hall, 1982).

Let us consider an affine combination of $\widehat{\gamma}_n(k_1)$ and $\widehat{\gamma}_n(k_2)$,

$$\widehat{\gamma}^{a}_n(k_1, k_2) := a\,\widehat{\gamma}_n(k_1) + (1-a)\,\widehat{\gamma}_n(k_2) \;. \tag{2.3}$$

We may then write

$$\begin{aligned}\widehat{\gamma}^{a}_n(k_1, k_2) \;&\overset{d}{=}\; \gamma + \frac{\sigma}{\sqrt{k_1}}\sqrt{a^2 + (1-a^2)\,\frac{k_1}{k_2}}\; Z^a_n \;+\; b\,A\Big(\frac{n}{k_1}\Big)\,\left[a + (1-a)\Big(\frac{k_1}{k_2}\Big)^{\rho}\right] \\ &\quad + a\,o_p\Big(A\Big(\frac{n}{k_1}\Big)\Big) + (1-a)\,o_p\Big(\Big(\frac{k_1}{k_2}\Big)^{\rho} A\Big(\frac{n}{k_1}\Big)\Big) \\ &\quad + a\,o_p(1/\sqrt{k_1}) + (1-a)\,o_p(1/\sqrt{k_2})\end{aligned} \tag{2.4}$$

where $Z^a_n = \dfrac{a\,Z_{n,1} + (1-a)\sqrt{\frac{k_1}{k_2}}\,Z_{n,2}}{\sqrt{a^2 + (1-a^2)\,\frac{k_1}{k_2}}}$ is asymptotically standard Normal.

If $0 < \theta \le 1$ (i.e., k_1 and k_2 are of the same order), we have asymptotic normality whenever $\lim_{n\to\infty} \sqrt{k_1}\,A(n/k_1) = \lambda_1$ finite, i.e., as $n \to \infty$,

$$\sqrt{k_1}\Big(\widehat{\gamma}^{a}_n(k_1, k_2) - \gamma\Big) \;\xrightarrow{d}\; N\Big(\lambda_1\, b\Big(a + (1-a)\,\theta^{\rho}\Big),\, \sigma^2\Big(a^2 + (1-a^2)\,\theta\Big)\Big)$$

and convergence in probability whenever $\sqrt{k_1}\,|A(n/k_1)| \longrightarrow \infty$, i.e.,

$$\frac{\widehat{\gamma}^{a}_n(k_1, k_2) - \gamma}{A(n/k_1)} \;\xrightarrow{p}\; b\Big(a + (1-a)\,\theta^{\rho}\Big) \;.$$

But, if $\rho < 0$, we may choose $a = \big(1 - (k_1/k_2)^{-\rho}\big)^{-1}$ and we then obtain asymptotic normality with an asymptotic null bias if, as $n \to \infty$, $\sqrt{k_1}\,A(n/k_1) \to \lambda_1$ finite, and

$$\frac{\widehat{\gamma}^{a}_n(k_1, k_2) - \gamma}{A(n/k_1)} = o_p(1) \quad \text{ if } \;\; \sqrt{k_1}\,|A(n/k_1)| \to \infty \;. \tag{2.5}$$

Under the validity of (2.4) and whenever $\sqrt{k_1}\,|A(n/k_1)| \longrightarrow \lambda_1$ finite as $n \to \infty$, we have an *AMSE* given by

$$(2.6)\qquad AMSE[\widehat{\gamma}_n^a(k_1,k_2)] \;=\; \frac{\sigma^2\Big(a^2 + (1-a^2)\,\theta\Big)}{k_1} + b^2\Big(a + (1-a)\,\theta^\rho\Big)^2 A^2(n/k_1)\;.$$

If $\theta = 0$ (i.e. k_1 is of smaller order than k_2), and $\rho < 0$, then for a fixed a, and for a model in Hall's class, the third term of (2.4) dominates and we get a convergence in probability, i.e.

$$\frac{\widehat{\gamma}_n^a(k_1,k_2) - \gamma}{A(n/k_1)\left(\frac{k_1}{k_2}\right)^\rho} \xrightarrow{\;p\;} b(1-a)\;,$$

provided that $\sqrt{k_1}\,A(n/k_1)$ does not converge to zero.

If we choose $a \equiv a_n \to 1$, such that $(1-a_n)\,A(n/k_1)\left(\frac{k_1}{k_2}\right)^\rho = o(A(n/k_1))$, we have $\widehat{\gamma}_n^a(k_1,k_2) = \widehat{\gamma}_n(k_1) + o_p(1)$. But if we particularly choose $a_n = (1-(k_1/k_2)^{-\rho})^{-1}$, then we have again asymptotic normality with null bias if $\sqrt{k_1}\,A(n/k_1) \to \lambda_1$ finite, and (2.5) holds.

Convex mixtures of two Hill's estimators, with the form (2.3) with $k_1 = \theta k_2$ and $0 < a < 1$ were already considered in Martins *et al.* (1999).

Let us now think on the Generalized Jackknife r. v. associated to $(\widehat{\gamma}_n(k), \widehat{\gamma}_n(\theta k))$,

$$(2.7)\qquad \widehat{\gamma}_{n,\theta}^q(k) := \frac{\widehat{\gamma}_n(k) - q\,\widehat{\gamma}_n(\theta k)}{1-q}\,, \qquad 0<\theta<1, \quad \text{with } q = q_n = \frac{Bias_\infty[\widehat{\gamma}_n(k)]}{Bias_\infty[\widehat{\gamma}_n(\theta k)]}\;.$$

Since $Bias_\infty[\widehat{\gamma}_n(k)]/Bias_\infty[\widehat{\gamma}_n(\theta k)] = A(n/k)/A(n/\theta k)$ converges to θ^ρ as $n/k \to \infty$, we will consider $q = \theta^\rho$. Notice that $\widehat{\gamma}_{n,\theta}^q(k)$ has the functional form of $\widehat{\gamma}_n^a(k_1,k_2)$ in (2.3), with $k_1 = \theta k$, $k_2 = k$, $a = q/(q-1)$. If we put $\rho = -1$, we have $q = 1/\theta$, and the new estimator

$$(2.8)\qquad \widehat{\gamma}_{n,\theta}^{1/\theta}(k) := \frac{\widehat{\gamma}_n(\theta k) - \theta\,\widehat{\gamma}_n(k)}{1-\theta}\;.$$

We then have $Var_\infty\Big[\widehat{\gamma}_{n,\theta}^{1/\theta}(k)\Big] = \frac{\sigma^2}{k}\,\frac{\theta^2-\theta-1}{\theta(\theta-1)}$, and

$$\theta_0 := \arg\min_\theta\, Var_\infty\Big[\widehat{\gamma}_{n,\theta}^{1/\theta}(k)\Big] \;=\; \frac{1}{2}\;.$$

This is an argument that leads to the Generalized Jackknife estimator we are going to consider,

$$\widehat{\gamma}_n^G(k) := 2\,\widehat{\gamma}_n(k/2) - \widehat{\gamma}_n(k) \;, \tag{2.9}$$

with the distributional representation

$$\widehat{\gamma}_n^G(k) \overset{d}{=} \gamma + \frac{\sigma\sqrt{5}}{\sqrt{k}}\, Z_n^G + b(2^{\rho+1} - 1)\, A(n/k) + o_p(A(n/k)) + o_p(1/\sqrt{k}) \;,$$

where $Z_n^G = \frac{1}{\sqrt{5}}\Big[2\sqrt{2}\, Z_{n,1} - Z_{n,2}\Big]$ is asymptotically standard Normal. Notice that to reduce asymptotically the bias, we increase the variance 5 times!

For this estimator, we have the *AMSE*

$$AMSE[\widehat{\gamma}_n^G(k)] \;=\; \frac{5\,\sigma^2}{k} + \Big(b(2^{\rho+1} - 1)\Big)^2 A^2(n/k)$$

and consequently whenever $b \neq 0$ and $\rho \neq -1$,

$$\begin{aligned} k_0^G(n) &:= \arg\inf_k AMSE[\widehat{\gamma}_n^G(k)] \\ &= \frac{n}{s^{\leftarrow}\Big(\frac{5\sigma^2}{n(b(2^{\rho+1}-1))^2}\Big)} \;. \end{aligned}$$

Notice that the squared asymptotic bias of $\widehat{\gamma}_n^G(k)$ is always smaller than that of the original estimator.

The estimator in (2.9) is a particular case (assuming a known value $\rho = -1$) of the Generalized Jackknife estimator, to be studied elsewhere,

$$\widehat{\gamma}_{n,\widehat{\rho}}^G(k) := \frac{\widehat{\gamma}_n(k) - 2^{-\widehat{\rho}}\,\widehat{\gamma}_n(k/2)}{1 - 2^{-\widehat{\rho}}} \;, \tag{2.10}$$

where $2^{-\widehat{\rho}}$ is an estimator of $\lim_{n\to\infty} \frac{Bias[\widehat{\gamma}_n(k)]}{Bias[\widehat{\gamma}_n(k/2)]}$. $\widehat{\gamma}_{n,\widehat{\rho}}^G(k)$ is also a particular form of the more general estimator $\widehat{\gamma}_n^a(k_1, k_2)$ in (2.3) with $k = k_2 = 2\,k_1$ and the weight a given by $a = 1/(1 - 2^{\widehat{\rho}})$; this is the weight that provides the elimination of the asymptotic bias.

Since the estimation of the second order parameter ρ is still problematic, it is useful to analyse the behaviour of $\widehat{\gamma}_n^{(a)}(k) := \widehat{\gamma}_n^a(k/2, k)$ for a non-optimal choice of $a > 1$. The asymptotic properties of this class of estimators are discussed below, while some numerical results are shown in section 4.

From the asymptotic behaviour in (2.6) we can conclude that

(**i**) for a fixed level k, the reduction in the asymptotic bias of $\widehat{\gamma}_n^{(a)}(k)$ relatively to the original estimator $\widehat{\gamma}_n(k)$ is:

$$R := \lim_{n\to\infty} \left(\left| \frac{Bias_\infty[\widehat{\gamma}_n(k)]}{Bias_\infty[\widehat{\gamma}_n^{(a)}(k)]} \right| \right) = \frac{1}{|1-a(1-2^\rho)|} \ ;$$

Figure 1 shows the values of this quocient in the (a,ρ)-plane, independently of the tail index γ;

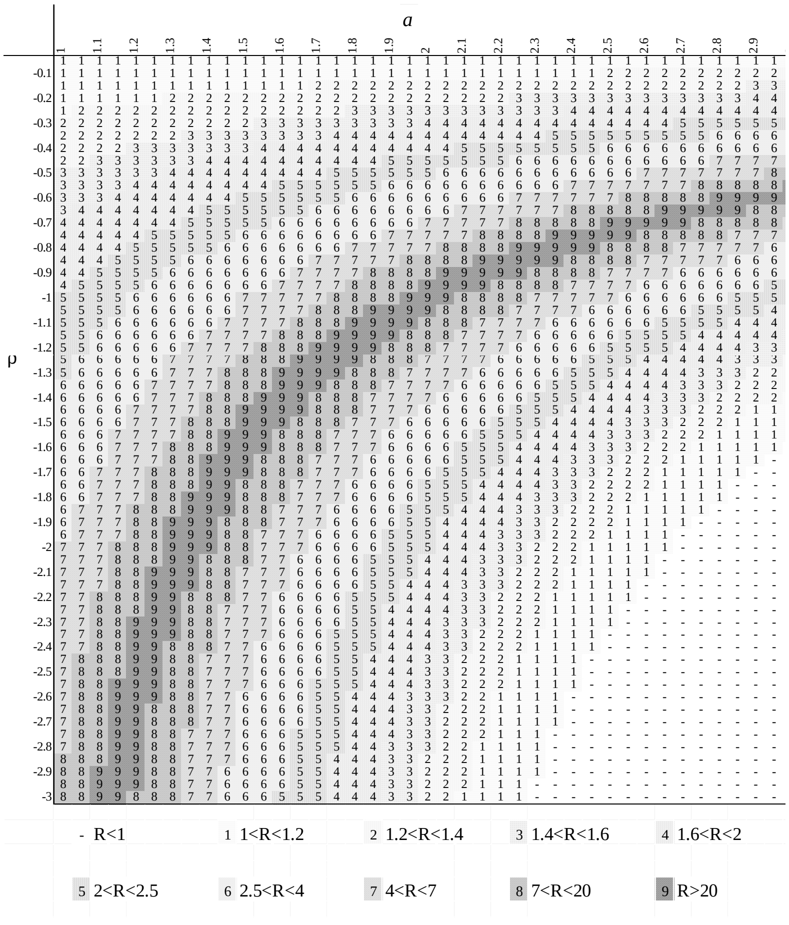

Fig. 1: Ratio between the asymptotic bias of $\widehat{\gamma}_n^{(a)}(k)$ and of $\widehat{\gamma}_n(k)$, in the (a,ρ)-plane (independently of the tail index γ).

(ii) defining asymptotic efficiency of $\widehat{\gamma}_n^{(a)}(k)$ relative to $\widehat{\gamma}_n(k)$ as the ratio between the asymptotic mean squared errors computed at the respective optimal levels, we have, provided that $a \neq 1/(1-2^\rho)$,

$$AEFF_a = AEFF_{\widehat{\gamma}_{n,0}^{(a)}|\widehat{\gamma}_{n,0}} = \frac{LMSE[\widehat{\gamma}_{n,0}]}{LMSE[\widehat{\gamma}_{n,0}^{(a)}]} = \left(\frac{(a^2+1)^\rho}{1-a(1-2^\rho)}\right)^{\frac{2}{1-2\rho}} ;$$

Figure 2 shows the values of the asymptotic efficiency in the (a,ρ)-plane.

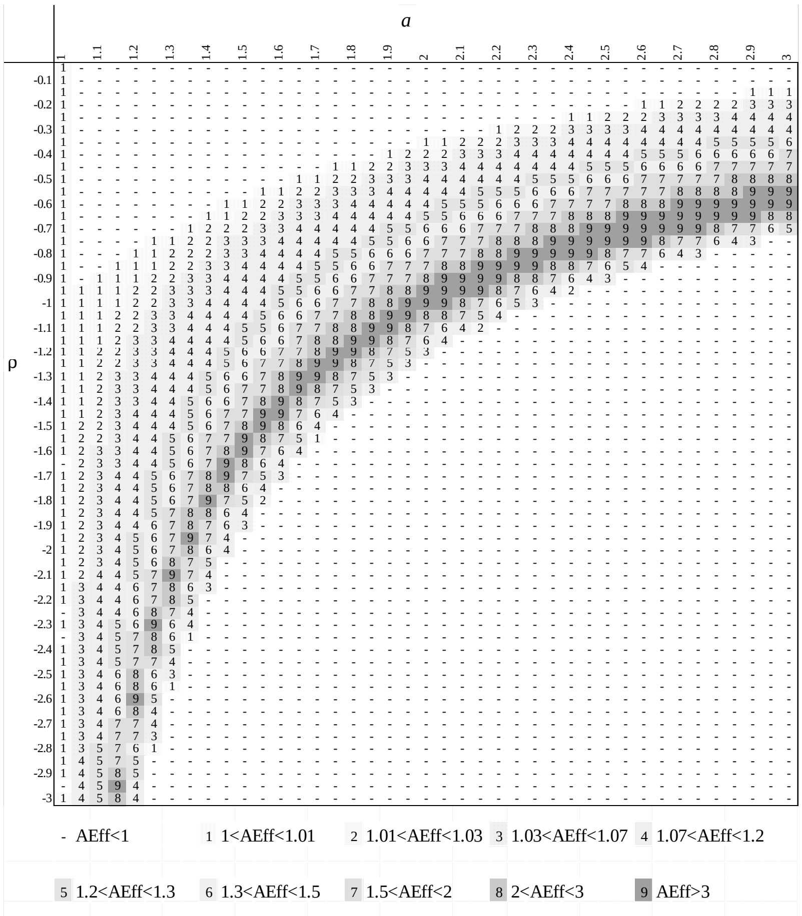

Fig. 2: Asymptotic efficiency of $\widehat{\gamma}_{n,0}^{(a)}$ relatively to $\widehat{\gamma}_{n,0}$, in the (a,ρ)-plane (independently of the tail index γ).

For a related work concerning the application of the Generalized Jackknife theory to the estimation of parameters of rare events, see Gomes *et al.* (2000).

3 – Further comments on the Generalized Jackknife statistics

The classical way of applying Jackknife's methodology is the following: given an estimator $\widehat{\gamma}_n(k)$, of a certain functional $\gamma = \gamma(F)$, apply Quenouille's re-sampling technique (Quenouille, 1956) and build the n estimators $\widehat{\gamma}_{n-1,i}(k)$, $1 \le i \le n$, with the same functional form of the original estimator $\widehat{\gamma}_n(k)$, but based on a sample of size $n-1$, obtained from the original sample, after remotion of the i-th element. Consider then the new estimator

$$\overline{\widehat{\gamma}}_n(k) := \frac{1}{n} \sum_{i=1}^{n} \widehat{\gamma}_{n-1,i}(k) \ .$$

The Quenouille estimator associated to $\widehat{\gamma}_n(k)$ is

$$\widehat{\gamma}_n^Q(k) := n\,\widehat{\gamma}_n(k) - (n-1)\,\overline{\widehat{\gamma}}_n(k) \ .$$

Notice however that $\overline{\widehat{\gamma}}_n(k)$ is the average of n original estimators based on samples of size $n-1$; thus, for an intermediate sequence $k = k_n$ such that $\sqrt{k}\,A(n/k) \to \lambda$, finite, as $n \to \infty$, $Bias_\infty[\overline{\widehat{\gamma}}_n(k)] = Bias_\infty[\widehat{\gamma}_{n-1}(k)] = b\,A\left(\frac{n-1}{k}\right)$.

If we here consider $(\widehat{\gamma}_n(k), \overline{\widehat{\gamma}}_n(k))$, we then have a quotient of asymptotic biases asymptotically equivalent to $(n/(n-1))^\rho$, and thus the associated Generalized Jackknife statistic is

$$\widehat{\gamma}_{n,\widehat{\rho}}^{G_Q}(k) := \frac{\widehat{\gamma}_n^{(1)}(k) - (n/(n-1))^{\widehat{\rho}}\ \overline{\widehat{\gamma}}_n(k)}{1 - (n/(n-1))^{\widehat{\rho}}} \ ,$$

where $\widehat{\rho}$ is a suitable estimator of ρ. But if we assume again, as for the building of $\widehat{\gamma}_n^G(k)$ in (2.9), $\rho = -1$, we obviously obtain a general estimator of the tail index γ,

$$\widehat{\gamma}_n^{G*}(k) := n\,\widehat{\gamma}_n(k) - (n-1)\,\overline{\widehat{\gamma}}_n(k) \equiv \widehat{\gamma}_n^Q(k) \ ,$$

which is exactly the *pure Jackknife statistic* of Quenouille–Tukey associated to the original estimator $\widehat{\gamma}_n(k)$. However, simulation results (not shown, due to their practical irrelevance) suggested that this estimator has a terribly high variance, and it is not at all competitive with the other estimators we have presented before.

The expected reduction in bias for $\widehat{\gamma}_n^G(k)$, based on theoretical developments, was not achieved in finite samples, as may be seen in next section. This lead us to turn back to the Generalized Jackknife estimator, presented in (2.7), based on the affine combination of an estimator at two different levels. We have seen that, if we took the weight (q), as the quotient between the biases of the two original estimators, we would get an unbiased estimator for γ. The estimator $\widehat{\gamma}_n^G(k)$ was developed taking this goal in mind and assuming the relations

(**i**) $Bias_\infty[\widehat{\gamma}_n(k)] = b\,A(n/k)$;

(**ii**) $A(n/\theta k) \sim \theta^{-\rho} A(n/k)$, as $n \to \infty$,

which led us to $q = \theta^\rho$. For the Fréchet model ($F(x) = \exp(-x^{-1/\gamma})$, $x > 0$) we have $A(t) \sim -\frac{\gamma}{2}\ln(1-\frac{1}{t})$, as $t \to \infty$, and if we compute $A(n/k)/A(n/\theta k)$ we find that the asymptotic relation (ii) is a rather poor approximation for n/k not too high. If in (2.7) we choose $q = \ln(1-k/n)/\ln(1-\theta k/n)$, which converges towards θ^{-1} as $k/n \to 0$, instead of $q = \theta^\rho$ ($\rho = -1$), we obtain an estimator with the asymptotic properties already described for $\widehat{\gamma}_{n,\theta}^{1/\theta}(k)$ in (2.8). For the reason therein exposed we choose $\theta = 1/2$ and define

$$\widehat{\gamma}_n^{G_F}(k) := \frac{\widehat{\gamma}_n(k) - \frac{\ln(1-k/n)}{\ln(1-k/2n)}\,\widehat{\gamma}_n(k/2)}{1 - \frac{\ln(1-k/n)}{\ln(1-k/2n)}}\,, \tag{3.1}$$

which is obviously specially devised for a Fréchet parent.

The estimators $\widehat{\gamma}_n^G(k)$ and $\widehat{\gamma}_n^{G_F}(k)$ are asymptotically undistinguishable, but despite of that, they have quite distinct exact properties. In the next section, we present simulated results for the particular case where $\widehat{\gamma}_n(k)$ is the Hill estimator; it may be observed that while $\widehat{\gamma}_n^G(k)$ has some bias, $\widehat{\gamma}_n^{G_F}(k)$ is almost unbiased and has a *MSE* at the optimal level much lower than that of $\widehat{\gamma}_n^G(k)$, and also lower than that of the original estimator.

With this better approximation for biases quotient we could really achieve a drastic reduction in *MSE*, although at the cost of involving a large number of observations in the sample. However, we cannot forget that this drastic reduction in *Bias* and *MSE* was achieved for one specific model, for which we know a good approximation for the bias of the Hill estimator, and hence for the quotient $A(n/k)/A(n/\theta k)$. But this quotient depends drastically on the model, and so from a practical point of view, it would be very useful to have a way of estimating that quotient or to have a better approximation for it, valid for a wide class of models.

Following this idea we are going to consider another approximation for $A(n/k)/A(n/\theta k)$, valid for a large class of models. Indeed we have found that a second term for that quocient may be added, assuming the validity of a second order regular variation condition for $A(\cdot)$ (Martins, 2000). For a sub-class of Hall's class where the Generalized Pareto and the Burr models ($\gamma > 0$) are included, we have

$$\frac{A(n/k)}{A(n/\theta k)} \sim \theta^\rho \left[1 - \left(\frac{k}{n}\right)^{-\rho} (\theta^{-\rho} - 1)\right], \qquad \frac{n}{k} \to \infty \ .$$

Proceeding in the same way as we did for $\widehat{\gamma}_n^G(k)$, i.e. working with $\rho = -1$ and $\theta = 1/2$, we get $q = 2 + k/n$, and the estimator

$$\widehat{\gamma}_n^{G_S}(k) := \frac{(2 + k/n)\, \widehat{\gamma}_n(k/2) - \widehat{\gamma}_n(k)}{1 + k/n} \ , \tag{3.2}$$

with the same asymptotic properties as $\widehat{\gamma}_n^G(k)$ and $\widehat{\gamma}_n^{G_F}(k)$. The improvement achieved in finite samples is illustrated in section 4 (Figures 4 and 5).

It thus seems that it pays to invest on finding better approximations for the bias, or to estimate the bias by means of bootstrap techniques (Hall(1990), Gomes(1994), Gomes and Oliveira (2000)). Such a technique, and its influence on the behaviour of a Generalized Jackknife statistic of the type of the one in (2.7) is however beyond the scope of this paper, and is being investigated.

4 – Some results for finite samples

The results presented in this section illustrate the behaviour of the proposed estimators in finite samples, when the original estimator is the Hill estimator, $\widehat{\gamma}_n(k) \equiv \widehat{\gamma}_n^H(k)$ in (1.8).

The simulations of mean values and *MSE*'s of the estimators are based on 5000 replicas. The relative efficiency of an estimator is defined as the square root of the quocient between the simulated *MSE* for the Hill and for that estimator, both computed at their optimal simulated levels. Those relative efficiencies are based on 20×5000 replicas.

4.1. Choice of a in $\widehat{\gamma}_n^{(a)}$

Given a sample, we may choose the value a_0 of a providing the greater stability of the estimates of $\widehat{\gamma}_n^{(a)}(k)$ around its mean value, through a minimum square technique similar to the one used in Gomes and Martins (2001). In Figure 3 we present sample paths (plots of estimates vs. k) for $\widehat{\gamma}_n^H(k)$ and $\widehat{\gamma}_n^{(a_0)}(k)$ and for Burr parents $\Big(1-F(x) = (1 + x^{-\rho/\gamma})^{1/\rho}$, $x>0$, $\gamma>0$ and $\rho<0\Big)$ with $\gamma = 1$ and different values of ρ.

Notice the higher stability of the sample path of these new estimators — small bias for a wide range of k values — and consequently less relevance for the choice of the optimal level.

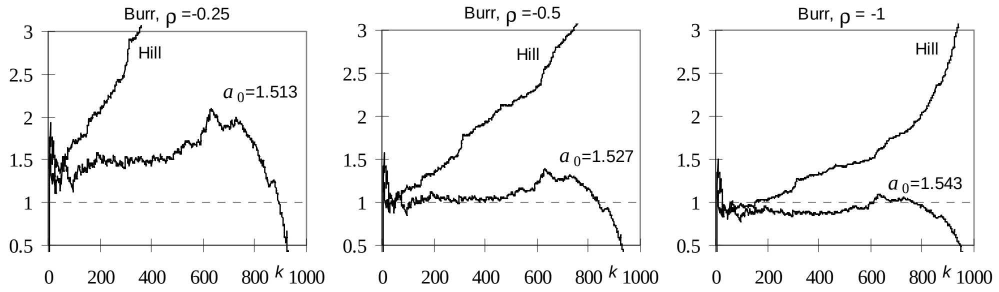

Fig. 3: Sample path of $\widehat{\gamma}_n^{(a_0)}(k)$, $1 \leq k \leq n-1$ for a sample of size $n = 1000$ from Burr parents.

4.2. Comparison between the Hill and the Jackknife Statistics

In Figures 4 and 5 we present simulated mean values and *MSE's* of the estimators $\widehat{\gamma}_n^H$, $\widehat{\gamma}_n^G$, $\widehat{\gamma}_n^{(a_0)}$, $\widehat{\gamma}_n^{G_F}$ and $\widehat{\gamma}_n^{G_S}$, for samples with size $n = 1000$, from the Fréchet model with $\gamma = 1$ $(\rho = -1)$ and the Burr model with $\gamma = 1$ and $\rho = -0.5$, respectively. Notice that for the Fréchet and the Burr models, the distribution of $\widehat{\gamma}_n^H(k)/\gamma$ is independent of γ, so the results presented for $\gamma = 1$, are also valid for $\mathrm{E}[\cdot]/\gamma$ and $MSE[\cdot]/\gamma^2$, whatever the value of $\gamma > 0$.

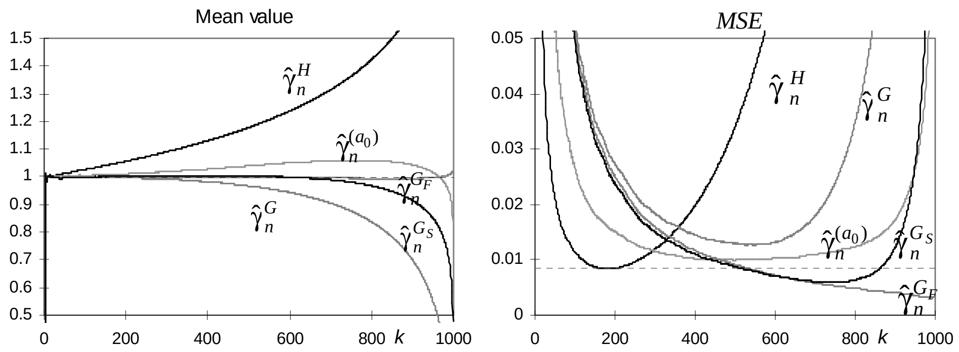

Fig. 4: Simulated mean values and *MSE* for Fréchet parents with $\gamma = 1$ $(\rho = -1)$.

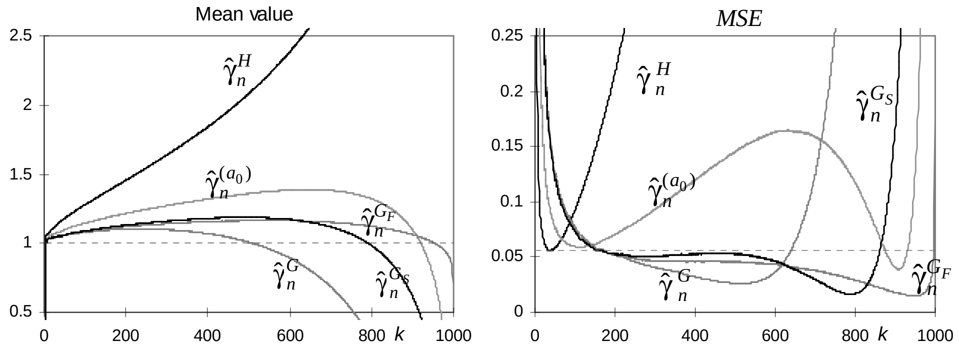

Fig. 5: Simulated mean values and *MSE* for Burr parents with $\gamma = 1$ and $\rho = -0.5$.

All the proposed Generalized Jackknife estimators have a smaller bias than the original estimator, here the Hill estimator. The reduction seems to be more effective when the weight q_n in (2.7) depends on the level k, as happens in $\widehat{\gamma}_n^{G_F}(k)$ and $\widehat{\gamma}_n^{G_S}(k)$. In what concerns the *MSE*, there is a reduction on its dependence on k, and it may be seen that we go below the optimal *MSE* of the Hill estimator.

Table 1: Efficiencies of $\gamma_n^{G_F}(k_0^{G_F}(n))$ relatively to $\gamma_n^H(k_0^H(n))$.

n	Fréchet	Burr $\rho=-0.25$	Burr $\rho=-0.5$	Burr $\rho=-1$	Burr $\rho=-2$
100	1.19	1.22	1.26	1.02	0.77
200	1.32	1.33	1.42	1.07	0.75
500	1.48	1.48	1.70	1.15	0.73
1000	1.62	1.60	1.96	1.22	0.71
2000	1.77	1.73	2.26	1.29	0.70
5000	2.02	1.90	2.78	1.38	0.68
10000	2.25	2.04	3.25	1.45	0.67
20000	2.51	2.20	3.81	1.53	0.66

Table 2: Efficiencies of $\gamma_n^{G_S}(k_0^{G_S}(n))$ relatively to $\gamma_n^H(k_0^H(n))$.

n	Fréchet	Burr $\rho=-0.25$	Burr $\rho=-0.5$	Burr $\rho=-1$	Burr $\rho=-2$
100	0.91	1.25	1.18	0.94	0.71
200	0.98	1.46	1.33	1.01	0.72
500	1.08	1.82	1.59	1.13	0.72
1000	1.18	2.18	1.84	1.22	0.72
2000	1.28	2.64	2.14	1.33	0.72
5000	1.45	3.42	2.63	1.49	0.72
10000	1.59	4.18	3.08	1.63	0.71
20000	1.75	5.11	3.60	1.79	0.70

At the optimal levels, we present in Tables 1 and 2, respectively, the efficiencies of $\widehat{\gamma}_n^{G_F}$ and $\widehat{\gamma}_n^{G_S}$ relatively to the Hill estimator, and for Fréchet and Burr parents, for several sample sizes n.

Except for very small ρ values, here illustrated with $\rho = -2$, there is a great reduction of the *MSE* of the Generalized Jackknife estimators at the respective optimal levels, relatively to the Hill estimator.

REFERENCES

[1] Dekkers, A.L.M.; Einmahl, J.H.J. and Haan, L. de – A moment estimator for the index of an extreme-value distribution, *Ann. Statist.,* 17 (1989), 1833–1855.

[2] Dekkers, A.L.M. and Haan, L. de – Optimal choice of sample fraction in extreme-value estimation, *J. of Multivariate Analysis,* 47 (1993), 173–195.

[3] Geluk, J. and Haan, L. de – *Regular Variation, Extensions and Tauberian Theorems*, CWI Tract 40, Center for Mathematics and Computer Science, Amsterdam, Netherlands, 1987.

[4] Gnedenko, B.V. – Sur la distribution limite du terme maximum d'une série aléatoire, *Ann. Math.,* 44 (1943), 423–453.

[5] Gomes, M.I. – *Metodologias Jackknife e Bootstrap em Estatística de Extremos*, Actas do II Congresso Anual da Sociedade Portuguesa de Estatística, 1994, 31–46.

[6] Gomes, M.I. and Martins, M.J. – Generalizations of the Hill estimator — asymptotic versus finite sample behaviour, *J. Statist. Plann. Inference,* 93 (2001), 161–180.

[7] Gomes, M.I.; Martins, M.J. and Neves, M. – Alternatives to a semi-parametric estimator of parameters of rare events – the Jackknife methodology, *Extremes,* 3(3) (2000), 207–229.

[8] Gomes, M.I. and Oliveira, O. – The bootstrap methodology in Statistical Extremes – choice of the optimal sample fraction, *Extremes,* 4(4) (2001), 331–358.

[9] Gray, H.L. and Schucany, W.R. – *The Generalized Jackknife Statistic*, Marcel Dekker, 1972.

[10] Haan, L. de – *On Regular Variation and its Application to the Weak Convergence of Sample Extremes*, Mathematical Centre Tract 32, Amsterdam, 1970.

[11] Haan, L. de and Stadtmüller, U. – Generalized regular variation of second order, *J. Austral. Math. Soc. (A),* 61 (1996), 381–395.

[12] Hall, P. – On some simple estimates of an exponent of regular variation, *J. R. Statist. Soc.,* 44(1) (1982), 37–42.

[13] Hall, P. – Using the bootstrap to estimate mean squared error and select smoothing parameter in nonparametric problems, *J. Multivariate Anal.,* 32 (1990), 177–203.

[14] Hill, B.M. – A simple general approach to inference about the tail of a distribution, *Ann. Statist.,* 3 (1975), 1163–1174.

[15] Martins, M.J. – *Estimação de Caudas Pesadas – Variantes ao Estimador de Hill*, Ph.D. Thesis, Universidade de Lisboa, 2000.

[16] Martins, M.J.; Gomes, M.I. and Neves, M. – Some results on the behaviour of Hill's estimator, *J. Statist. Comput. and Simulation,* 63 (1999), 283–297.

[17] Quenouille, B. – Notes on bias in estimation, *Biometrika,* 43 (1956), 353–360.

M. Ivette Gomes,
D.E.I.O. and C.E.A.U.L.,
Universidade de Lisboa – PORTUGAL

and

M. João Martins and Manuela Neves,
Departamento de Matemática, Instituto Superior de Agronomia,
Universidade Técnica de Lisboa – PORTUGAL