Computational Statistics and Data Analysis 82 (2015) 223–237

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

journal homepage: www.elsevier.com/locate/csda

A new partially reduced-bias mean-of-order p class of extreme value index estimators

M. Ivette Gomes [a,*], M. Fátima Brilhante [b,c], Frederico Caeiro [d], Dinis Pestana [a]

[a] Centro de Estatística e Aplicações, Faculdade de Ciências, Universidade de Lisboa, Portugal

[b] Centro de Estatística e Aplicações, Universidade de Lisboa, Portugal

[c] Departamento de Matemática, Universidade dos Açores, Portugal

[d] Centro de Matemática e Aplicações, Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa, Portugal

ARTICLE INFO

Article history:
Received 28 April 2014
Received in revised form 20 August 2014
Accepted 29 August 2014
Available online 19 September 2014

Keywords:
Bias estimation
Heavy right-tails
Heuristic methods
Optimal levels
Semi-parametric estimation
Statistics of extremes

ABSTRACT

A class of partially reduced-bias estimators of a positive *extreme value index* (EVI), related to a mean-of-order-p class of EVI-estimators, is introduced and studied both asymptotically and for finite samples through a Monte-Carlo simulation study. A comparison between this class and a representative class of *minimum-variance reduced-bias* (MVRB) EVI-estimators is further considered. The MVRB EVI-estimators are related to a direct removal of the dominant component of the bias of a classical estimator of a positive EVI, the Hill estimator, attaining as well minimal asymptotic variance. Heuristic choices for the *tuning* parameters p and k, the number of top order statistics used in the estimation, are put forward, and applied to simulated and real data.

1. Introduction and preliminaries

Let $X_1, \ldots, X_n$ be independent, identically distributed (i.i.d.), or possibly weakly dependent and stationary random variables (r.v.'s) from an underlying cumulative distribution function (c.d.f.) F. Let us denote the associated ascending order statistics (o.s.) by $X_{1:n} \le \cdots \le X_{n:n}$. Let us also assume that there exist sequences of real constants $\{a_n > 0\}$ and $\{b_n \in \mathbb{R}\}$ such that the maximum, linearly normalized, i.e. $(X_{n:n} - b_n)/a_n$, has a non-degenerate limit. Then the limiting distribution is necessarily an *extreme value* (EV) distribution, denoted $\mathrm{EV}_\xi(\cdot)$, with the functional form

$$\mathrm{EV}_\xi(x) = \begin{cases} \exp(-(1+\xi x)^{-1/\xi}), & 1+\xi x > 0, \text{ if } \xi \neq 0, \\ \exp(-\exp(-x)), & x \in \mathbb{R}, \text{ if } \xi = 0. \end{cases} \quad (1)$$

The c.d.f. F is then said to belong to the max-domain of attraction of EV_ξ, and we write $F \in \mathcal{D}_{\mathrm{M}}(\mathrm{EV}_\xi)$. The parameter ξ is the *extreme value index* (EVI), the primary parameter of extreme events, with a low frequency, but with a usually high impact. The EVI measures the heaviness of the *right tail function* (RTF), $\overline{F}(x) := 1 - F(x)$, as $x \to \infty$, and the heavier the tail, the larger the EVI is. We shall work here with Pareto-type distributions, with a strictly positive EVI.

* Correspondence to: Departamento de Estatística e Investigação Operacional, Faculdade de Ciências de Lisboa, Bloco C6, Piso 4, 1749-016 Lisboa, Portugal.

E-mail addresses: ivette.gomes@fc.ul.pt (M.I. Gomes), fbrilhante@uac.pt (M.F. Brilhante), fac@fct.unl.pt (F. Caeiro), dinis.pestana@fc.ul.pt (D. Pestana).

http://dx.doi.org/10.1016/j.csda.2014.08.017
0167-9473/

1.1. First and second-order conditions for heavy tails

Power laws, such as the Pareto income distribution (Pareto, 1965) and the Zipf's law for city-size distribution (Zipf, 1941), have been observed a long time ago in many important phenomena in economics and biology and have recently seriously attracted scientists. In statistics of extremes, F is often said to be heavy-tailed whenever the RTF, $\overline{F}$, is a *regularly varying* (RV) function with a negative index of regular variation equal to $-1/\xi,\ \xi > 0$. Equivalently, with $F^{\leftarrow}(x) := \inf\{y : F(y) \ge x\}$ denoting the generalized inverse function of F, the *reciprocal tail quantile function* (RTQF), $U(t) := F^{\leftarrow}(1-1/t),\ t \ge 1$, is of regular variation with index ξ. For details on regular variation, see Bingham et al. (1987). With the notation RV_a for the class of RV functions with an index of regular variation a, i.e., positive measurable functions $g(\cdot)$ such that $\lim_{t\to\infty} g(tx)/g(t) = x^a$, for all $x > 0$,

$$F \in \mathcal{D}_{\mathrm{M}}\left(\mathrm{EV}_{\xi>0}\right) \iff \overline{F} \in \mathrm{RV}_{-1/\xi} \text{ (Gnedenko, 1943)}$$
$$\iff U \in \mathrm{RV}_{\xi} \text{ (de Haan, 1984)}. \tag{2}$$

The second-order parameter, ρ (≤ 0), rules the rate of convergence in any of the first-order conditions in (2), and it is the non-positive parameter appearing in the limiting relation

$$\lim_{t\to\infty} \frac{\ln U(tx) - \ln U(t) - \xi \ln x}{A(t)} = \eta_\rho(x) := \begin{cases} \dfrac{x^\rho - 1}{\rho}, & \text{if } \rho < 0, \\ \ln x, & \text{if } \rho = 0, \end{cases} \tag{3}$$

which we assume to hold for every $x > 0$, and where $|A|$ must then be of regular variation with index ρ (Geluk and de Haan, 1987). Note that for a model like the sin-Burr$_{\xi,r}$ model, with an RTQF, $U(t) = (t^{-r} - \sin t^{-r})^{-\xi/r},\ t \ge 1$, the second-order condition (3) does not hold. We shall further assume everywhere in the article that $\rho < 0$, excluding thus models with $\rho = 0$, like the log-gamma. Whenever dealing with bias reduction, we shall further assume that we are working in the more strict Hall–Welsh class of Pareto-type models (Hall and Welsh, 1985), with an RTF,

$$\overline{F}(x) = Cx^{-1/\xi}\left(1 + D_1 x^{\rho/\xi} + o\left(x^{\rho/\xi}\right)\right), \quad \text{as } x \to \infty,\ \xi > 0, \tag{4}$$

for $C > 0, D_1 \ne 0, \rho < 0$. Regarding the RTQF, we can then say that there exist $c > 0$ and $d_1 \ne 0$ such that $U(t) = ct^\xi\left(1 + d_1 t^\rho + o\left(t^\rho\right)\right)$, as $t \to \infty$. Therefore, condition (3) holds and we may choose there $A(t) = \alpha t^\rho$, for an adequate α, which we reparameterize as,

$$A(t) = \xi\beta t^\rho, \quad \rho < 0. \tag{5}$$

1.2. The class of EVI-estimators under play

For Pareto-type models, the most commonly used EVI-estimators are the Hill estimators (Hill, 1975), which are the averages of the log-excesses, V_{ik}, i.e.

$$\mathrm{H}(k) := \frac{1}{k}\sum_{i=1}^{k} V_{ik}, \qquad V_{ik} := \ln X_{n-i+1:n} - \ln X_{n-k:n}, \quad 1 \le i \le k < n. \tag{6}$$

But since we can write

$$\mathrm{H}(k) = \sum_{i=1}^{k} \ln\left(\frac{X_{n-i+1:n}}{X_{n-k:n}}\right)^{1/k} = \ln\left(\prod_{i=1}^{k}\frac{X_{n-i+1:n}}{X_{n-k:n}}\right)^{1/k}, \quad 1 \le k \le n,$$

the Hill estimator can be thought as the logarithm of the geometric mean (or mean-of-order-0) of $\underline{\mathbb{U}} := \{U_{ik} := X_{n-i+1:n}/X_{n-k:n},\ 1 \le i \le k < n\}$. More generally, Brilhante et al. (2013) considered as basic statistics the mean-of-order-p (MOP) of $\underline{\mathbb{U}}$, with $p \ge 0$, i.e., the class of statistics

$$A_p(k) = \begin{cases} \left(\dfrac{1}{k}\displaystyle\sum_{i=1}^{k} U_{ik}^p\right)^{1/p}, & \text{if } p > 0, \\ \left(\displaystyle\prod_{i=1}^{k} U_{ik}\right)^{1/k}, & \text{if } p = 0, \end{cases}$$

and the class of MOP EVI-estimators,

$$\mathrm{H}_p(k) \equiv \mathrm{MOP}_p(k) := \begin{cases} \left(1 - A_p^{-p}(k)\right)/p, & \text{if } 0 < p < 1/\xi, \\ \ln A_0(k) = \mathrm{H}(k), & \text{if } p = 0, \end{cases} \tag{7}$$

with $\mathrm{H}_0(k) \equiv \mathrm{H}(k)$, given in (6). We now state the following result, proved for $p = 0$ in de Haan and Peng (1998) and for $0 < p < 1/(2\xi)$ in Brilhante et al. (2013).