Smooth tail-index estimation

Müller, S ; Rufibach, K

DOI: https://doi.org/10.1080/00949650802142667

Originally published at:
Müller, S; Rufibach, K (2009). Smooth tail-index estimation. Journal of Statistical Computation and Simulation, 79(9):1155-1167.
DOI: https://doi.org/10.1080/00949650802142667

Smooth tail index estimation

Samuel Müller
School of Mathematics and Statistics F07
University of Sydney, NSW 2006, Australia
E-mail: mueller@maths.usyd.edu.au
Phone: +61 (0)2 9036 7878
Fax: +61 (0)2 9351 4534

and

Kaspar Rufibach
Biostatistics Unit, Institute of Social and Preventive Medicine
University of Zurich
Hirschengraben 84
8001 Zurich, Switzerland
Email: kaspar.rufibach@ifspm.uzh.ch
Phone: +41 (0)44 634 46 43
Fax: +41 (0)44 634 49 86

Abstract

Both parametric distribution functions appearing in extreme value theory - the generalized extreme value distribution and the generalized Pareto distribution - have log-concave densities if the extreme value index $\gamma \in [-1, 0]$. Replacing the order statistics in tail index estimators by their corresponding quantiles from the distribution function that is based on the estimated log-concave density $\widehat{f}_n$ leads to novel smooth quantile and tail index estimators. These new estimators aim at estimating the tail index especially in small samples. Acting as a smoother of the empirical distribution function, the log–concave distribution function estimator reduces estimation variability to a much greater extent than it introduces bias. As a consequence, Monte Carlo simulations demonstrate that the smoothed version of the estimators are well superior to their non-smoothed counterparts, in terms of mean squared error.

Keywords: "extreme value" theory; log-concave density estimation; negative Hill estimator; Pickands estimator; tail index estimation; small–sample performance

2000 Mathematics Subject Classifications: Primary 62G32, 62G07; Secondary 60G70

1 Introduction

It is a well–known fact that asymptotic results are in general at best approximately valid in small-sample problems, but that in the latter situation bias is often a serious issue. For example in extreme value theory the small-sample bias in the estimation of the tail index is severe. We refer to [7] for a study of a number of estimators. There are only a few more articles that focus on the small-sample performance of tail-index estimators. We are aware of [3, 23, 40], which consider tail-index estimation for heavy-tailed distributions. Here, we investigate the small-sample behavior of a new smooth tail-index estimator for thin-tailed distributions and distributions with finite endpoint. The main aim of this article is to introduce smoothed estimators that exploit log-concavity of the limiting density of the exceedances or of the largest order statistics, respectively. In Section 2 we present the connection between log-concavity and extreme-value theory whereas in Section 3 we show that replacing the empirical distribution function by the smooth estimator $\widehat{F}_n$ (the latter is based on the log-concave density estimator, see (3) for a proper definition) leads to novel tail-index estimators that substantially decrease mean-squared error in small-sample situations. We illustrate this finding in a simulation study in Section 4 for two settings (a) a generalized Pareto distribution and (b) a domain of attraction scenario. The paper concludes with some brief remarks in Section 5.

2 Log-concavity in extreme-value theory

2.1 Max-domain of attraction of distributions with log-concave densities

Let $\{X_n,\, n \geq 1\}$ be a sequence of independent random variables with common cumulative distribution function F, such that F belongs to the max-domain of attraction of G, denoted by $F \in \mathcal{D}(G)$, i.e. there exist constants $a_n > 0$, $b_n \in \mathbb{R}$ such that for $x \in \mathbb{R}$ and

$$\begin{aligned} G(x) &= \lim_{n\to\infty} \mathrm{P}\left(a_n^{-1}[\max(X_1,\dots,X_n)+b_n] \leq x\right) \\ &= \lim_{n\to\infty} F^n(a_n x - b_n) \end{aligned}$$

which is equivalent to

$$\sup_{x\in\mathbb{R}} |F^n(a_n x - b_n) - G(x)| \to 0 \text{ as } n \to \infty.$$

From [16] it is known that $F \in \mathcal{D}(G)$ if and only if $G \in \{G_\gamma : \gamma \in \mathbb{R}\}$, where

$$G_\gamma(x) = \exp\left(-(1+\gamma x)^{-1/\gamma}\right),\ 1+\gamma x > 0,$$

where G_γ is called the extreme value distribution with tail index $\gamma \in \mathbb{R}$, shift parameter 0, and scale parameter 1. Since

$$(1+\gamma x)^{-1/\gamma} \to \exp(-x) \text{ for } \gamma \to 0,$$

interpret $G_0(x)$ as $\exp(-e^{-x})$. The two most common settings in the analysis of extreme values are that we either have an observed sequence of independent and identically distributed maxima, $M_{n,1},\dots,M_{n,k}$, or upper order statistics, $X_{(n)} \geq X_{(n-1)} \geq \dots \geq X_{(n-K_n)}$, from an independent and identically distributed sample $X_1,\dots,X_n$ with $K_n = k_n < n$ or $K_n = \#\{X_i : X_i \geq u_n\}$ a random number. Here, u_n is some suitably chosen high threshold and $\#A$ denotes the number of elements in set A. We focus on upper order statistics and assume intermediate sequences for K_n, i.e. $K_n/n = k_n/n \to \infty$ if k_n is a sequence of real numbers, and $K_n/n \to 0$ in probability if K_n is random.

2.2 The Generalized Pareto Distribution and log-concavity

To fix notation, define for a general distribution function F the lower endpoint $\alpha(F) := \inf\{x \in \mathbb{R} :\ F(x) > 0\}$ and the upper endpoint $\omega(F) := \sup\{x \in \mathbb{R} :\ F(x) < 1\}$. The quantile function of F for $q \in [0,1]$ is

$$F^{-1}(q) := \inf\{x \in \mathbb{R} \,:\, F(x) \geq q\}.$$

Exceedances $X_{(n)} - u, \ldots, X_{(n-k_n+1)} - u$ of a high threshold $u = u_n$ or of an intermediate order statistic $u = X_{(n-k_n)}$ are typically modelled by the generalized Pareto distribution (GPD), established by Pickands, see [31]. For $\gamma \in \mathbb{R}$ and $\sigma \in (0, \infty)$ the density of the GPD is given by

$$w_{\gamma,\sigma}(x) \;\; := \;\; \sigma^{-1}(1 + \gamma x/\sigma)^{-(1+1/\gamma)}, \; x \in [0, -\sigma/\gamma], \tag{1}$$

where $w_{0,\sigma}$ is again defined via continuity: $w_{0,\sigma}(x) = \sigma^{-1}\exp(-x/\sigma)$ for $x \in [0, \infty)$.

Both parametric distribution functions appearing in extreme value theory – the generalized extreme value distribution (GEV) and the generalized Pareto distribution – have log-concave densities if the extreme value index $\gamma \in [-1, 0]$ and all distribution functions F with log-concave density belong to the max-domain of attraction of the GEV with $\gamma \in [-1, 0]$, see [30]. For any distribution function F with corresponding real-valued log-concave density function f, this latter f can be written as

$$f(x) \;\; = \;\; \exp\{\varphi(x)\}, \tag{2}$$

for a concave function $\varphi : \mathbb{R} \to [-\infty, \infty)$. We will denote the class of distribution functions F having a log-concave density on its support $[\alpha(F), \omega(F)]$ by $\mathcal{F}_{\frown}$.

2.3 The restriction $\gamma \in [-1, 0]$

We are aware that the restriction of log-concave densities and of $\gamma \in [-1, 0]$, respectively, is a drawback if there is not sufficient evidence to assume that this restriction holds. On the other hand, there are often good reasons to assume that some distribution function F has all its moments finite or that its support is finite, implying $\gamma \in [-1, 0]$. Estimating the finite endpoint $\omega(F)$ of a distribution F is linked to the problem of estimating $\gamma < 0$ and its theory is well developed, see for example [6, 14, 19, 20]. The Dutch data set of total life span of people who were born in the years 1877-1881, as analyzed in [1], is a real life example that yields an estimated tail index between $-1/2$ and 0, and thus a finite upper endpoint. Further data sets on survival times of 208 mice exposed to radiation and on men's 100m running times of the 1988 and 1992 Olympic Games are analyzed in [20]. By definition, the distribution of the distance of two points in a closed convex set has finite support and there are many open problems regarding its limit behavior. For the current state of research we refer to [27]. A further example that naturally leads to the restriction $\gamma < 0$ is the estimation of the efficient frontier in economics, see [15]. In practical applications, $\gamma = -1/2$ is often seen as natural lower bound, e.g. [14] or [25, p. 62] .

The restriction $\gamma \geq -1$ automatically ensures that densities which are unbounded in a left neighborhood of their right endpoint are excluded. Ongoing research shows (see [26]) that even if the sign of γ is known, the connection to the general framework of GPDs is not completely severed, especially if one does not aim at reducing bias. In practical applications, we suggest that in an extreme value context estimators which are valid for the entire range of $\gamma \geq -1$ or $\gamma \geq -1/2$, respectively, as introduced in Section 3, are used for a first guess of the range of γ. Once the restriction $\gamma \in [-1, 0]$ seems plausible, our smoothed estimators can be used to considerably reduce mean squared error.

3 Tail index estimation

The estimation of γ is besides the related high quantile estimation the most important problem in univariate extreme value theory and there exists a vast number of different approaches. For example the Hill estimator [21], the maximum likelihood estimator [4, 19, 37, 38, 39], the moment estimator [8], the (iterated) negative Hill estimator also known as Falk's estimator [12, 13, 29], the (generalized) Pickands estimator [9, 31, 36], estimators based on near extremes [28], the weighted least squares estimator [24], probability weighted moments [22], and many more. All these estimators are based on an intermediate sequence of upper order statistics and it is well known (see e.g. [18]) that a major drawback of such estimators is their discrete character. Using kernel-type estimators is one possibility to overcome this deficiency. We refer to [5] for the smoothed Hill estimator in case $\gamma > 0$ and to [18] for general $\gamma \in \mathbb{R}$. Our alternative is to take advantage of the distribution function $\widehat{F}_n$ based on the log-concave density estimator $\widehat{f}_n$, an approach valid if $\gamma \in [-1, 0]$.

3.1 Motivation of new estimators

For an i.i.d. sample $X_1, \ldots, X_n$ where X_i has a log-concave density function as introduced in (2), let $\mathbb{F}_n$ be the empirical distribution function and

$$\widehat{F}_n(x) \quad := \quad \int_{-\infty}^{x} \widehat{f}_n(t)\,\mathrm{d}t \tag{3}$$

be the smoothed distribution function based on the log-concave density estimator $\widehat{f}_n$. The latter is the maximizer of the "adjusted" criterion function

$$L(\varphi) = \int_{\mathbb{R}} \varphi(x) d\mathbb{F}_n(x) - \int_{\mathbb{R}} \exp\{\varphi(x)\} dx$$

over all concave functions $\varphi : \mathbb{R} \to [-\infty, \infty)$. The log–concave density estimator is then $\widehat{f}_n := \exp \widehat{\varphi}_n$. In [11, 32] existence, uniqueness and many properties of $\widehat{f}_n$ are derived, whereas computational aspects are treated in [10, 11]. Finally, limiting distributions for $\widehat{f}_n$ and its mode as well as local asymptotic minimax lower bounds for the estimation of the mode by means of $\widehat{f}_n$ are derived in [2].

Let us mention only one special feature of $\widehat{f}_n$: the estimator $\widehat{\varphi}_n$ of the log-density is a piecewise linear function with knots only at some of the observations points $X_1, \ldots, X_n$ and $\widehat{\varphi}_n = -\infty$ on $\mathbb{R} \setminus [X_1, X_n]$. In [11] the following theorem is proven.

Theorem 3.1. *Let $X_1, \ldots, X_n$ be an i.i.d. sample stemming from a distribution with log-concave density $f = \exp \varphi$ Furthermore, for all $A \leq x < y \leq B$, an exponent $\beta \in [1, 2]$ and a constant $L > 0$ we assume that $|\varphi(x) - \varphi(y)| \leq L|x - y|$ if $\beta = 1$ and $|\varphi'(x) - \varphi'(y)| \leq L|x - y|^{\beta - 1}$ if $\beta > 1$. For φ' either $\varphi'(\cdot -)$ or $\varphi'(\cdot +)$, we finally stipulate that $\varphi'(x) - \varphi'(y) \geq C(y - x)$ for $C > 0$ and all $x, y \in [A, B]$. Then, as $n \to \infty$,*

$$\max_{t \in T_n} |\mathbb{F}_n - \widehat{F}_n|(t) \quad = \quad o_{\mathrm{p}}\left(n^{-1/2}\right),$$

where $T_n \to T := [A, B]$. Furthermore, for every n, $\widehat{F}_n(X_{(1)}) = 0$ and $\widehat{F}_n(X_{(n)}) = 1$.

This theorem implies that $\widehat{F}_n$ is essentially equivalent to $\mathbb{F}_n$, but as the integral of a piecewise exponential function very smooth. These properties turn out to be highly convenient in the estimation of the extreme value tail index γ. The smoothness of $\widehat{F}_n$ reduces the variance not only considerably in the estimation of γ but even for the estimation of quantiles of the generalized Pareto distribution, as is shown in Section 4.

Many well-known tail index estimators are based on a selection of log-spacings of the sample, see also [36]. The key idea is now simply to replace the order statistics (or quantiles of the empirical distribution function) $X_{(i)} = \mathbb{F}_n^{-1}(i/n),\ i = 1, \ldots, n$ in these log-spacings by quantiles received via $\widehat{F}_n$. This yields modified versions of the uniformly minimum variance unbiased estimator from Falk [12, 13] for the case of a known endpoint, the negative Hill estimator as defined in [13], Pickands' estimator [31] for the case of a unknown endpoint and finally for the minimal-variance generalized Pickands' estimator given in Theorem 4.1 of [36]. We will denote these new estimators as "smoothed estimators".

We choose the first two estimators because of their outstanding performance for $\gamma < 0$ and $\gamma < -1/2$, respectively. On the other hand it is well known that Pickands' estimator is not efficient and in addition it is able to estimate any $\gamma \in \mathbb{R}$, thus the comparison of the original to the smoothed version is not fair. However, Pickands' estimator serves as the building block for much more efficient generalized Pickands' estimators that are more general linear combinations of log-spacings of order statistics [36], and it is therefore appealing to assess the gain of smoothing for Pickands' estimator. Finally, we also included Segers' generalized Pickands' estimator in our simulation study, cf. [36].

We already pointed out in the introduction that there are many approaches to estimate the tail index. However, our primary goal in this paper is to show that replacing order statistics by quantiles of $\widehat{F}_n$ in small samples is a promising approach to reduce variance of the estimates, without introducing too much bias. In [36, Figure 1], among others the generalized Pickands' estimator is compared to the popular maximum likelihood estimator introduced by [37] and the moment estimator given in [8], for values of $\gamma \in [-1, 2]$. Estimated relative efficiencies relative to our smoothed tail index estimators can be computed by combining these results with our simulation.

The absolute performance of various smoothed and unsmoothed tail index estimators will be evaluated in future work.

3.2 Global and tail behavior

Extreme value theory is, as the name suggests, tail focused. Hence, the behavior of the conditional distribution $F_{X|X>u}$, where $u \to \omega(F)$, dominates the limit results. On the other hand, the log-concavity of $f = F'$ is a strong assumption on the entire shape of the distribution function F. If this strong assumption holds, then the smoothing of the tail index estimators should be based on $\widehat{F}_n$. Since tail index estimators only use information of the upper tail of $\mathbb{F}_n$ it would be sufficient that only the upper tail of F had a log-concave density. Therefore, we investigate two settings in our simulation study in Section 4: (a) $X_1, \ldots, X_n$ iid with log-concave densities. Here the data is smoothed by $\widehat{F}_n$ and (b) $X_1, \ldots, X_n$ iid in the domain of attraction of G_γ for some $\gamma \in [-1, 0]$). In this more general situation, F does not necessarily belong to $\mathcal{F}_{\bar{\cap}}$. Before the smooth tail index estimators are computed, the range of log-concavity of f has to be determined; e.g. using the bump-hunting method developped in [32] or by imposing some assumptions on F. Here, the data is smoothed by $\widehat{F}_m$, where $\widehat{F}_m$ is based on the $m > K_n$ largest order statistics and m is such that $F_{X|X>X_{n-m}} \in \mathcal{F}_{\bar{\cap}}$.

3.3 Smooth tail index estimators

Suppose we are given a sample $X_1, \ldots, X_n$ from a GPD wherefrom we know that $\gamma \in [-1, 0]$ and with empirical distribution function $\mathbb{F}_n$. Denote the order statistics by $X_{(1)}, \ldots, X_{(n)}$. For such a fixed sample, define for $k = 4, \ldots, n$ and $H \in \{\mathbb{F}_n, \widehat{F}_n\}$:

$$\widehat{\gamma}^k_{\text{Pick}}(H) = \frac{1}{\log 2} \log\Big(\frac{H^{-1}((n - r_k(H) + 1)/n) - H^{-1}((n - 2r_k(H) + 1)/n)}{H^{-1}((n - 2r_k(H) + 1)/n) - H^{-1}((n - 4r_k(H) + 1)/n)}\Big),$$

where

$$r_k(H) = \begin{cases} \lfloor k/4 \rfloor & \text{if } H = \mathbb{F}_n, \\ k/4 & \text{if } H = \widehat{F}_n, \end{cases}$$

with $\lfloor m \rfloor := \max\{n \in \mathbb{N}_0 : n \le m\}$. This construction not only exploits the superiority of the quantile estimates based on the smooth function $\widehat{F}_n$, but also avoids "rounding bias". Using the inverse of a continuous distribution function, quantiles do not coincide for four consecutive k's (order statistics), as it is the case for Pickands' original estimate.

To generalize the estimators in [12, 13], no discrimination regarding continuity of H is necessary. For $H \in \{\mathbb{F}_n, \widehat{F}_n\}$ let

$$\begin{aligned} \widehat{\gamma}^k_{\text{Falk}}(H) &= \frac{1}{k-1}\sum_{j=2}^{k}\log\Big(\frac{X_{(n)} - H^{-1}((n-j+1)/n)}{X_{(n)} - H^{-1}((n-k)/n)}\Big), \ k = 3, \dots, n-1 \\ \widehat{\gamma}^k_{\text{MVUE}}(H) &= \frac{1}{k}\sum_{j=1}^{k}\log\Big(\frac{\omega(F) - H^{-1}((n-j+1)/n)}{\omega(F) - H^{-1}((n-k)/n)}\Big), \ k = 2, \dots, n-1 \end{aligned}$$

where F is the true distribution function of the X_i's. Note that $\widehat{\gamma}^k_{\text{MVUE}}(H)$ is only consistent if $\gamma \in [-1, 0)$. The estimator introduced in [36] is

$$\widehat{\gamma}^k_{\text{Segers}}(H) = \sum_{j=1}^{k}\Big(\lambda(j/k) - \lambda((j-1)/k)\Big)\log\Big(H^{-1}((n - \lfloor cj \rfloor)/n) - H^{-1}((n-j)/n)\Big)$$

for $k = 1, \dots, n-1$. The weights are chosen to minimize the asymptotic variance of $\widehat{\gamma}^k_{\text{Segers}}(\mathbb{F}_n)$, see Theorem 4.1. in [36]. In our simple comparison, we use these same weights, although it is clear that they are not optimal anymore for $\widehat{\gamma}^k_{\text{Segers}}(\widehat{F}_n)$. However, our aim is to analyze the effect of smoothing the order statistics in the original estimators.

The chosen terminology reminds of the fact that when choosing $H = \mathbb{F}_n$, the above estimators boil down to Pickands', Falk's, Falk's MVUE and Segers' estimator, as discussed at the beginning of this section.

The new, smooth tail index estimators are now simply $\widehat{\gamma}^k_{\text{Pick}}(\widehat{F}_n), \widehat{\gamma}^k_{\text{Falk}}(\widehat{F}_n), \widehat{\gamma}^k_{\text{MVUE}}(\widehat{F}_n)$ and $\widehat{\gamma}^k_{\text{Segers}}(\widehat{F}_n)$. Figure 1 displays Hill plots for two GPD pseudo-random samples, i.e. plots of the estimators versus the number of order statistics k, for the smoothed and unsmoothed versions for $n = 64$ and $\gamma \in \{-0.1, -0.75\}$.

The smoothed estimators behave much more stable as a function of k and it is especially noteworthy that for the two generated data sets the smoothed estimators $\widehat{\gamma}^k_{\text{Pick}}(\widehat{F}_n)$, $\widehat{\gamma}^k_{\text{Falk}}(\widehat{F}_n)$, $\widehat{\gamma}^k_{\text{MVUE}}(\widehat{F}_n) \in [-1, 0]$ for every k. By construction, $\widehat{\gamma}_{\text{Pick}}, \widehat{\gamma}_{\text{Segers}} : \mathbb{R}^n \to \mathbb{R}$ and $\widehat{\gamma}_{\text{Falk}}, \widehat{\gamma}_{\text{MVUE}} : \mathbb{R}^n \to [-\infty, 0)$ what implies that non permissible estimates outside the interval $[-1, 0]$ potentially occur. However, due to consistency of $\mathbb{F}_n$ and $\widehat{F}_n$, this is asymptotically negligible. If in practice it is known that $\gamma \in [-1, 0]$ but a smoothed estimator yields a value outside $[-1, 0]$, a truncation to its closest boundary value is recommendable.

It is beyond the scope and not the primary goal of this article to discuss the asymptotic behavior of the smoothed tail index estimators.

3.4 Further shape constraints

Straightforward computation yields the following lemma.

Lemma 3.1. *The generalized Pareto density $w_{\gamma,\sigma}$ as defined in (1) has the following qualitative properties that do not depend on the value of the scale parameter σ:*

Since e.g. a density estimator $\widetilde{f}_n$ in the class of convex decreasing densities is available, see [17], the latter lemma raises the possibility to define smooth estimators of the tail index for other ranges of γ. However, especially in the latter case, slight modifications may be necessary, since $\widetilde{F}_n(X_{(n)}) \neq 1$. Furthermore, maximum likelihood estimators under other constraints may not be as smooth as $\widehat{F}_n$, since e.g. the estimator of a convex decreasing

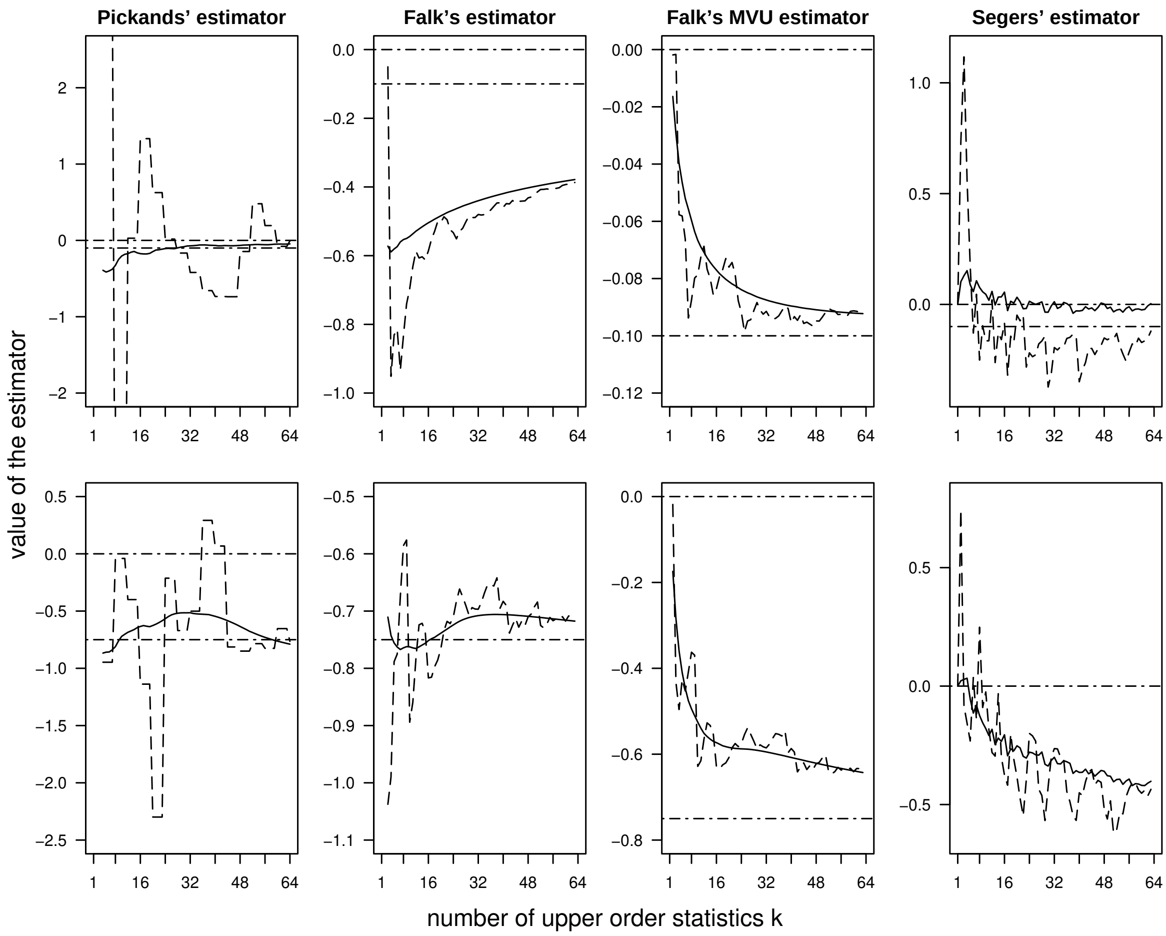

Figure 1: Hill plots for $n = 64$ and $\gamma = -0.1$ (plots in the upper row) and $\gamma = -0.75$ (plots in the lower row), smoothed (–) and original (- -) versions, horizontal lines (- · -) at zero and γ.

property	*parameter range*
convex non-decreasing	$\gamma \leq -1$
concave non-increasing	$\gamma \in [-1, -1/2]$
convex non-increasing	$\gamma \geq -1/2$
log-concave	$\gamma \in [-1, 0]$
log-convex	$\gamma \in (-\infty, -1] \cup [0, \infty)$

Table 1: Form of $w_{\gamma,\sigma}$ as determined by the tail index γ.

density is piecewise linear, whereas for log–concave densities this form appears for the estimator of the log–density.

3.5 Computational details

These new smoothed estimators are made available in the R-package `smoothtail`, see [35]. This latter package depends on the package `logcondens` [34], which offers two algorithms for the (weighted) estimation of an arbitrary log-concave density from an i.i.d. sample of observations. Both these packages are available from CRAN.

4 Simulations

4.1 Estimation of quantiles

The computation of the non-smoothed estimators $\widehat{\gamma}^k_{\text{Falk}}(\mathbb{F}_n)$, $\widehat{\gamma}^k_{\text{Pick}}(\mathbb{F}_n)$, $\widehat{\gamma}^k_{\text{MVUE}}(\mathbb{F}_n)$ and $\widehat{\gamma}^k_{\text{Segers}}(\mathbb{F}_n)$ heavily relies on the order statistics $X_{(i)}, i = 1, \ldots, n$. But these simply estimate the quantiles $W^{-1}_{\gamma,\sigma}(i/n)$ of the distribution whereof we want to estimate the tail index. Therefore, the accuracy of the tail index estimators is closely connected to the ability of estimating these quantiles $W^{-1}_{\gamma,\sigma}(i/n)$. To illustrate the superiority of log-concave quantile estimation over simply taking order statistics, we computed the relative efficiency of these two estimators. Since results were similar over an extended range of σ's, we concentrate on the case $\sigma = 1$.

To fix notation, define the log-concave estimate $\widehat{X}_{(i)} = \widehat{F}_n^{-1}(i/n)$ of an order statistic $X_{(i)}$, for $i = 1, \ldots, n$. Let $q_{(i)}$ denote either $\widehat{X}_{(i)}$ or $X_{(i)}$, then its estimated variance and bias with respect to the i/n-quantile of a GPD $W_{\gamma,1}$ for a fixed $\gamma \in [-1, 0]$ given simulated $q_{(i),j}$ that are based on M generated samples $X_{(1),j}, \ldots, X_{(n),j}$, $j = 1, \ldots, M$ of size n drawn from $W_{\gamma,1}$ is defined as follows:

$$\begin{aligned}
\widehat{\text{Var}}(q_{(i)}, M) &:= (M-1)^{-1} \sum_{j=1}^{M} \Big(q_{(i),j} - (1/M) \sum_{j=1}^{M} q_{(i),j} \Big)^2, \\
\widehat{\text{Bias}}(q_{(i)}, \gamma, M) &:= M^{-1} \Big(\sum_{j=1}^{M} q_{(i),j} \Big) - W^{-1}_{\gamma,1}(i/n).
\end{aligned}$$

The relative efficiency $\rho_{\gamma,n,M}(k)$ of log-concave quantile estimation to quantile estimation based on order statistics is then

$$\rho_{\gamma,n,M}(k) = \frac{[\widehat{\text{Bias}}(\widehat{X}_{(k)}, \gamma, M)]^2 + \widehat{\text{Var}}(\widehat{X}_{(k)}, M)}{[\widehat{\text{Bias}}(X_{(k)}, \gamma, M)]^2 + \widehat{\text{Var}}(X_{(k)}, M)}.$$

Figure 2 details $\rho_{\gamma,32,1000}(k)$ for $\gamma \in \{-1, -0.75, -0.5, -0.25, 0\}$ as a function of k. Relative efficiencies smaller than 1 are in favor of the log-concave quantile estimation and indicate its superiority. The use of the log-concave density estimator for the estimation of quantiles substantially reduces the variance of the estimation, due to its smoothing property detailed in Theorem 3.1. This transfers to a reduced MSE, uniformly in γ and k, as is detailed in Figure 2.

4.2 Smoothed versus unsmoothed tail index estimators

To assess the effect of smoothing the tail index estimators, we perform a simulation study for two settings. In Setting 1, we draw $M = 1000$ samples of size $n_1 = 64$ from a GPD with $\sigma = 1$ and extreme value tail index $\gamma \in \{-1, -0.75, -0.5, -1/3, -0.25, -0.1\}$. For every k, the log-concave density is estimated based on the full sample $X_{(1)}, \ldots, X_{(n)}$. Setting 2

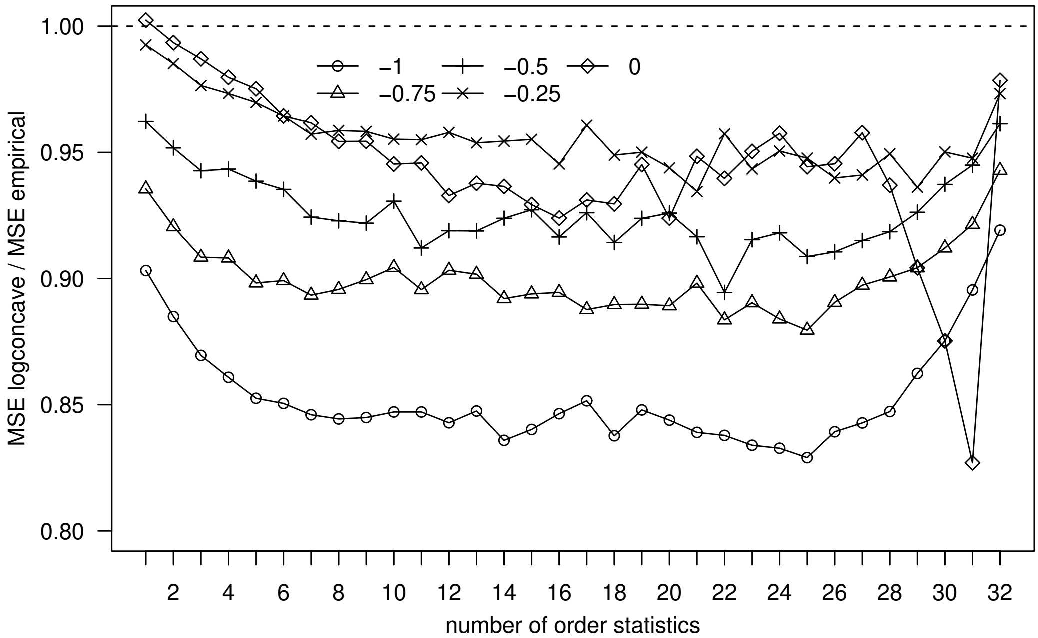

Figure 2: Relative MSE in quantile estimation for $n = 32$.

consists of $M = 1000$ samples from a $\beta(\theta_1, \theta_2)$-distribution having density

$$f_{\theta_1,\theta_2}(x) = \frac{\Gamma(\theta_1 + \theta_2)}{\Gamma(\theta_1)\Gamma(\theta_2)} x^{\theta_1 - 1}(1-x)^{\theta_2 - 1}, \quad x \in [0,1], \ \theta_1, \theta_2 > 0. \tag{4}$$

The upper tail of the β-distribution is dominated by θ_2, since for $x \uparrow 1$ one has $f_{\theta_1,\theta_2}(x) = c(\theta_1, \theta_2) w_{-\theta_2^{-1}, \theta_2^{-1}}(x)[1 + o(1)]$. Thus, $\beta(\theta_1, \theta_2) \in \mathcal{D}(G_{-1/\theta_2})$. We fix $\theta_1 = 1/2$ and thus only the upper tail of $f_{0.5,\theta_2}$ is log-concave. For the simulations, we choose $\theta_2 = -\gamma^{-1} \in \{1, 4/3, 2, 3, 4, 10\}$ and $n_2 = 128$. The latter n_2 deliberately equals $2n_1$, in order to underline the difference between the two settings. Here, the log-concave density estimator is based on the largest $n_2/2 = 64$ order statistics. In both settings we present results for a single sample size only because the results for $n_1, n_2/2 \in \{32, 128, 256, 512\}$ are very similar.

Setting 1 represents the "ideal" of observing pure "peak over threshold" data

$$X_1, \ldots, X_n \,\text{iid}\, \mathcal{L}(X) = \text{GPD} \in \mathcal{F}_{\bar{\cap}}.$$

Alternatively, Setting 2 stands for the more general situation

$$X_1, \ldots, X_n \,\text{iid}\, \mathcal{L}(X) \in \mathcal{D}(G_\gamma; \gamma \in [-1, 0]).$$

In the latter case, the well known problem of the tradeoff between bias and variance dominates the optimal choice of k, whereas in Setting 1 k can be chosen as large as possible since we are considering the "perfect" model.

4.3 Simulation results

We compute relative efficiencies for the estimation of the tail index as for quantile estimation. For Setting 1, results are displayed in Figure 3. If one knows that $\gamma \in [-1, 0]$, then clearly using the smoothed estimator is most worthwhile for Pickands' estimator. Also Falk's estimators benefit from smoothing, with highest gain in terms of MSE for small k's. Surprisingly, Segers' estimator benefits even from smoothing regarding variance, but only at the prize of increased bias. Only for the γ's close to 0 smoothing also slightly improves MSE for this estimator.

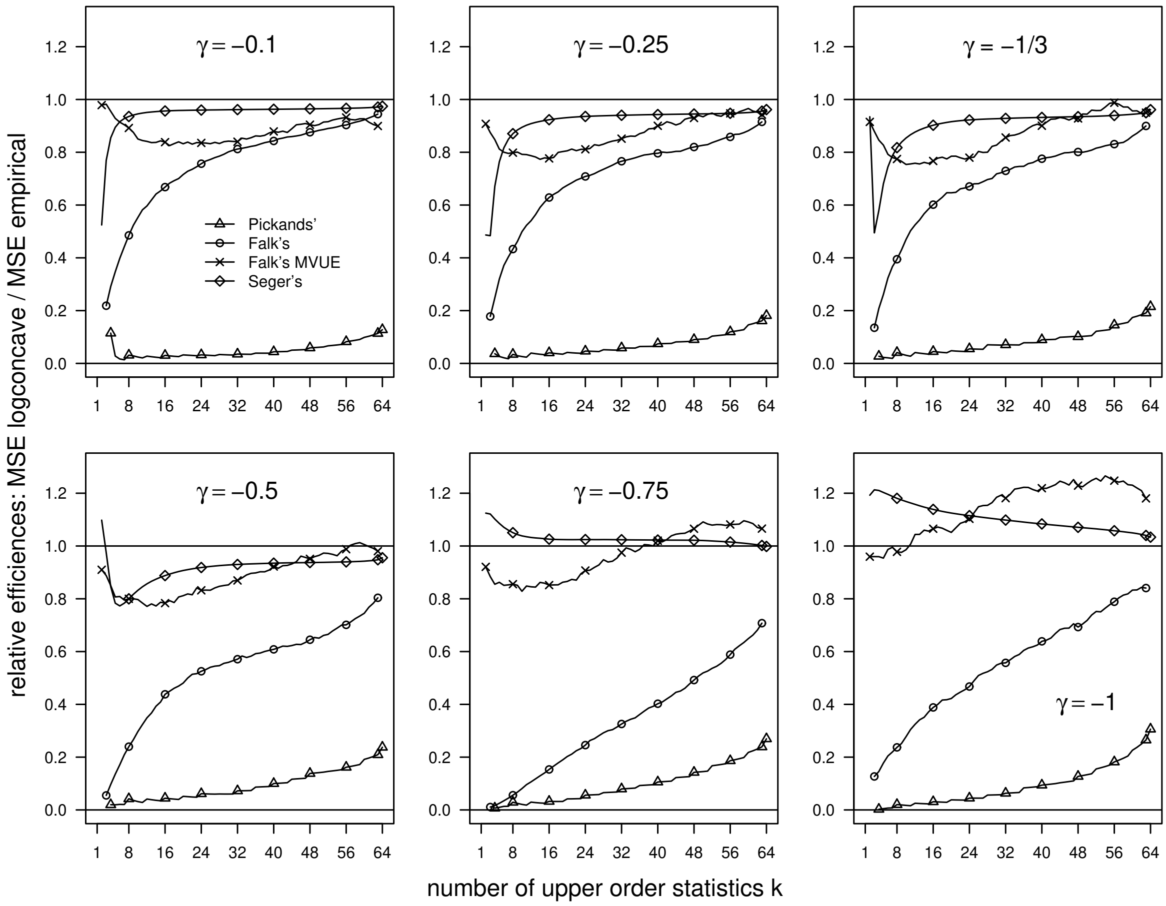

Figure 3: Relative MSE for tail index estimation for peaks over threshold data, $n = 64$.

Figure 4 sheds light on the bias-variance trade-off in Setting 2 for $\theta_2 = 3$; results for other choices of θ_2 were absolutely similar and therefore omitted. The variance in estimation of the tail index is dramatically reduced for the smoothed Pickands' estimate. The plot of the bias against k confirms that all estimators are biased as expected, since the data is generated by (4) and not by the GPD with $\gamma \in [-1, 0]$. Especially for Pickands' estimator, the bias is a much smoother function of k for the smoothed estimators than for the original estimator. Figure 5 shows the computed relative efficiencies for the estimation of the tail index for Setting 2. The results for $\widehat{\gamma}_{\text{Pick}}$ and $\widehat{\gamma}_{\text{Falk}}$ are similar to those in Setting 1, yielding the most substantial improvement for Pickands' estimator. On the other hand the efficiency of the smoothed and original $\widehat{\gamma}_{\text{MVUE}}$ is almost 1, independent of k and the

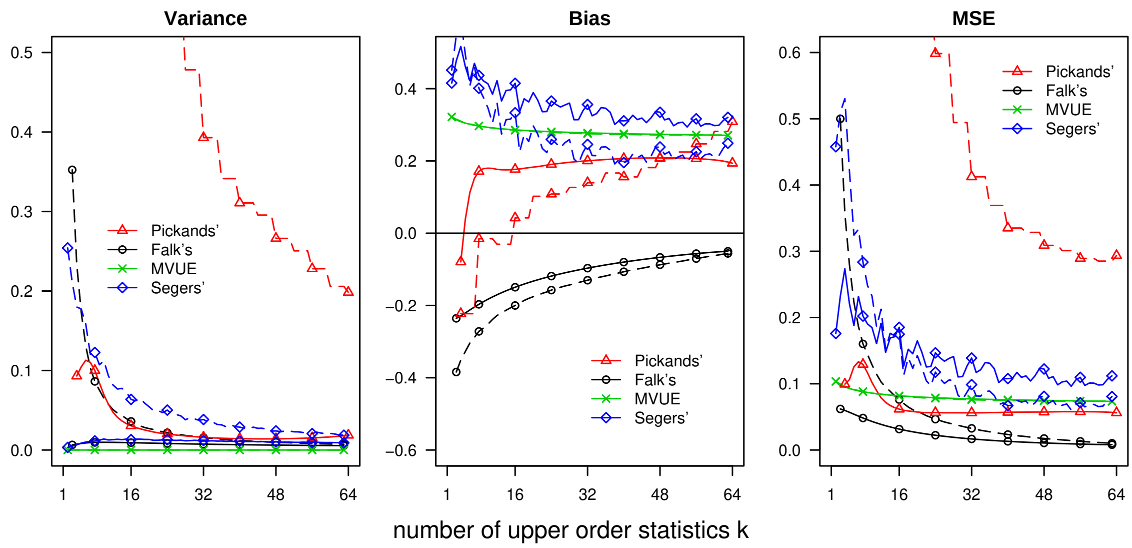

Figure 4: Domain of attraction scenario: $n = 128$, $\theta_1 = 1/2$, $\theta_2 = 3$, smoothed (–) and original (- -) versions.

choice of θ_2. As for Segers' estimator, smoothing yields a decreased MSE for all but the small k's.

5 Conclusions

In this article we showed that for the class of distributions $F \in \mathcal{D}(G_\gamma)$ having log-concave densities or log-concave conditional densities given that the observations exceed some threshold u, respectively, the smoothing of the empirical distribution function by the corresponding log-concave density estimator leads to improved quantile estimation uniformly in γ and the quantile to be estimated, and to more efficient tail index estimators. Not surprisingly, the reduction of the mean squared error of the smoothed tail index estimator is most substantial for Pickands' estimator, given the known poor efficiency of the latter. The estimator $\widehat{F}_n$ smoothes the empirical distribution function based on the global property of log-concavity of the density of the underlying distribution function F. We showed that if such a global property is present, then it can be exploited for the estimation of a tail property such as the extreme value index γ. Of course, the price to be paid is assuming that γ is restricted to the narrow interval $[-1, 0]$ and smoothing makes only sense if there is sufficient reason for that assumption. At present, we suggest to investigate Hill plots for tail index estimators that are valid for $\gamma \geq -1$ to visually check the restriction of $\gamma \in [-1, 0]$. It is ongoing (see [32]) and future research to find tests for assessing the log-concavity of densities. We want to stress the fact that the Hill plots in Section 3 and the results in our simulation study are not based on an optimal choice of the setting parameters and are reproducible for a wide range of parameter values for the $\beta(\theta_1, \theta_2)$-distribution unless the upper tail of the density is log-concave. Our simulation study can

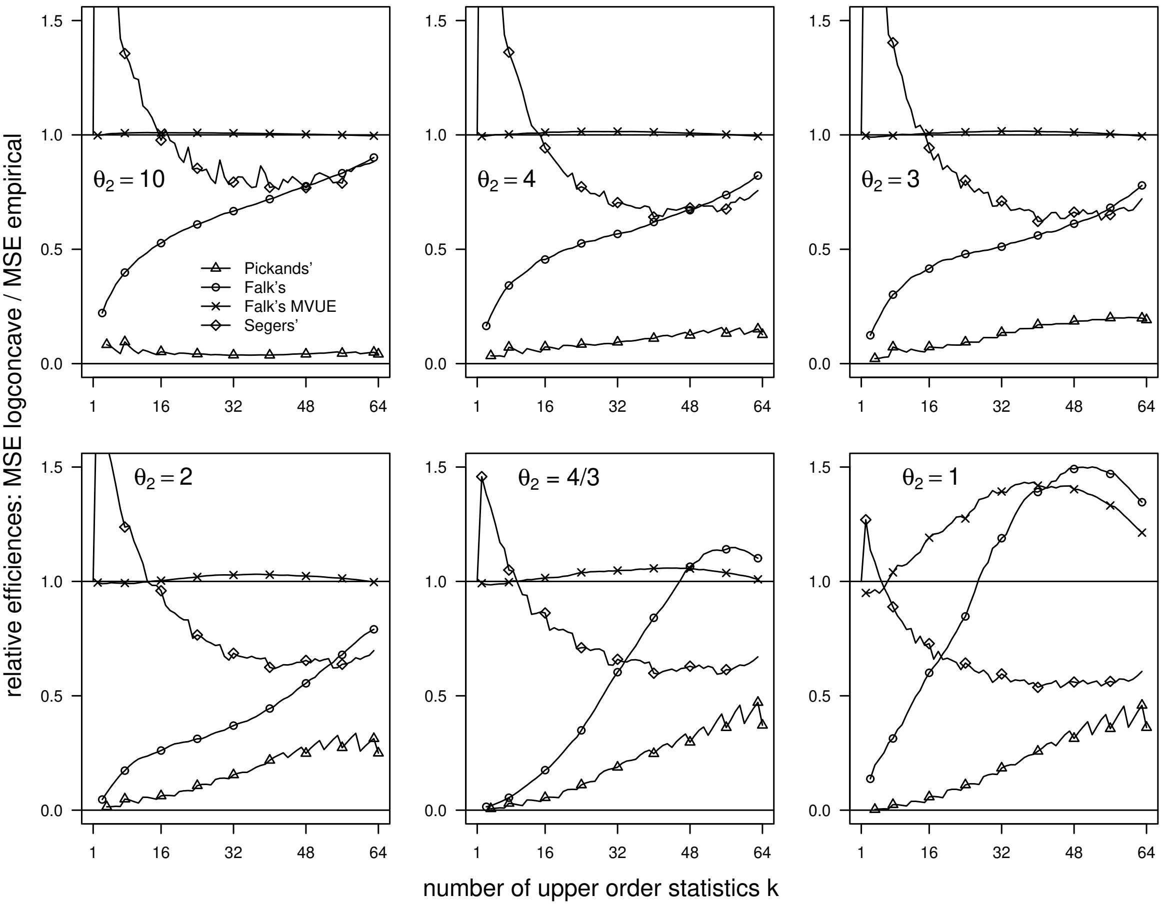

Figure 5: Relative MSE in domain of attraction scenario, $n = 128$ and $\theta_1 = 1/2$.

be reproduced using the R-package `smoothtail`, see [35].

The results in this article raise many new open questions and problems. We only mention two. First of all, it is challenging to prove asymptotic normality for the smoothed estimators. Then, fitting the generalized Pareto density to $\widehat{f}_n$ or $\log w_{\gamma,\sigma}$ to $\widehat{\varphi}_n$, respectively, opens the door to construct novel estimators for the tail index that are worthwhile being investigated in the future.

6 Acknowledgments

This research was initiated while both authors were working at the Institute of Mathematical Statistics and Actuarial Science at the University of Bern. K. Rufibach was partially supported by the Swiss National Science foundation.

References

[1] K. Aarssen and L. de Haan, *On the maximal life span of humans*, Math. Popul. Stud. 4 (1994), pp. 259-281.

[2] F. Balabdaoui, K. Rufibach, and J.A. Wellner, *Limit distribution theory for maximum likelihood estimation of a log-concave density*, Ann. Statist., to appear. Available at http://arxiv.org/abs/0708.3400v2.

[3] V. Brazauskas and R. Serfling, *Small sample performance of robust estimators of tail parameters for Pareto and exponential models*, J. Statist. Comput. Simulation 70 (2001), pp. 1-19.

[4] S.G. Coles and M.J. Dixon, *Likelihood-based inference for extreme value models*, Extremes 2 (1999), pp. 5-23.

[5] S. Csörgő, P. Deheuvels, and D.M. Mason, *Kernel estimates of the tail index of a distribution*, Ann. Statist. 13 (1985), pp. 1050-1077.

[6] S. Csörgő and D.M. Mason, *Simple estimators of the endpoint of a distribution*, in *Extreme Value Theory*, J. Hüsler and R.-D. Reiss, eds., Lecture Notes in Statistics, New York, 1989, pp. 132-147.

[7] M.M. Dacorogna, U.A. Müller, O.V. Pictet, and C.R. de Vries, *The distribution of extremal foreign exchange rate returns in extremely large data sets*, Discussion Paper 95-70, Tinbergen Institute, Netherlands, 1995.

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

[9] H. Drees, *Refined Pickands estimator of the extreme value index*, Ann. Statist. 23 (1995), pp. 2059-2080.

[10] L. Dümbgen, A. Hüsler, and K. Rufibach, *Active set and EM algorithms for log-concave densities based on complete and censored data*, preprint (2007). Available at http://arxiv.org/abs/0707.4643.

[11] L. Dümbgen, K. Rufibach, *Maximum likelihood estimation of a log-concave density and its distribution function: basic properties and uniform consistency*, Bernoulli, to appear (2007). Available at http://arxiv.org/abs/0709.0334.

[12] M. Falk, *Extreme quantile estimation in δ-neighborhoods of generalized Pareto distributions*, Statist. Probab. Lett. 20 (1994), pp. 9-21.

[13] —, *Some best parameter estimates for distributions with finite endpoint*, Statistics 27 (1995), pp. 115-125.

[14] A. Ferreira, L. de Haan, and L. Peng, *On optimising the estimation of high quantiles of a probability distribution*, Statistics 37 (2003), pp. 401-434.

[15] M.J. Farrell, *The measurement of productive efficiency*, J. Roy. Statist. Soc. Ser. A 120 (1957), pp. 253-281.

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

[17] P. Groeneboom, G. Jongbloed, and J.A. Wellner, *Estimation of a convex function: characterization and asymptotic theory*, Ann. Statist. 29 (2001), pp. 1653-1698.

[18] P. Groeneboom, H.P. Lopuhaä, and P.P de Wolf, *Kernel-type estimators for the extreme value index*, Ann. Statist. 31 (2003), pp. 1956-1995.

[19] P. Hall, *On estimating the endpoint of a distribution*, Ann. Statist. 10 (1982), pp. 556-568.

[20] P. Hall and J.Z. Wang, *Estimating the end-point of a probability distribution using minimum-distance methods*, Bernoulli 5 (1999), pp. 177-189.

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

[22] J.R.M. Hosking, J.R. Wallis, and E.F. Wood, *Estimation of the generalized extreme-value distribution by the method of probability-weighted moments*, Technometrics 27 (1985), pp. 251-261.

[23] R. Huisman, K.G. Koedijk, C.J.M. Kool, and F. Palm, *Tail-index estimates in small samples*, J. Bus. Econom. Statist. 19 (2001), pp. 208-216.

[24] J. Hüsler, D. Li, and S. Müller, *Weighted least squares estimation of the extreme value index*, Statist. Probab. Lett. 76 (2006), pp. 920-930.

[25] S. Kotz and S. Nadarajah, S., *Extreme Value Distributions. Theory and Applications*, Imperial College Press, London, 2000.

[26] D. Li and L. Peng, *Still fit generalized Pareto distributions?*, preprint (2008). Available at http://www.math.gatech.edu/ peng/GPD.pdf.

[27] M. Mayer and I. Molchanov, *Limit theorems for the diameter of a random sample in the unit ball*, Extremes 10 (2007), pp. 151-174.

[28] S. Müller, *Tail estimation based on numbers of near m-extremes*, Methodol. Comput. Appl. Probab. 5 (2003), pp. 197-210.

[29] S. Müller and J. Hüsler, *Iterative estimation of the extreme value index*, Methodol. Comput. Appl. Probab. 7 (2005), pp. 139-148.

[30] S. Müller and K. Rufibach, *On the max-domain of attraction of distributions with log-concave densities*, preprint (2008), to appear in Statist. Probab. Lett. Available at doi:10.1016/j.spl.2007.12.008.

[31] J. Pickands, *Statistical inference using extreme order statistics*, Ann. Statist. 3 (1975), pp. 119-131.

[32] K. Rufibach, *Log-concave density estimation and bump hunting for i.i.d. observations*, Ph.D. diss., Universities of Bern and Göttingen, 2006. Available at http://www.stub.unibe.ch/download/eldiss/06rufibach_k.pdf.

[33] K. Rufibach, *Computing maximum likelihood estimators of a log-concave density function*, J. Statist. Comput. Simul. 77 (2007), pp. 561-574.

[34] K. Rufibach and L. Dümbgen, *logcondens (Estimate a log-concave probability density from iid observations), R package version 1.3*, 2006; software available at http://www.r-project.org.

[35] K. Rufibach and S. Müller, *smoothtail (Smooth estimation of GPD shape parameter), R package version 1.1*, 2006; software available at http://www.r-project.org.

[36] J. Segers, *Generalized Pickands estimators for the extreme value index*, J. Statist. Plann. Inference 128 (2005), pp. 381-396.

[37] R.L. Smith, *Maximum likelihood estimation in a class of nonregular cases*, Biometrika 72 (1985), pp. 67-90.

[38] —, *Estimating tails of probability distributions*, Ann. Statist. 15 (1987), pp. 1174-1207.

[39] R.L. Smith and I. Weissman, *Maximum likelihood estimation of the lower tail of a probability distribution*, J. Roy. Statist. Soc. Ser. B 47 (1985), pp. 285-298.

[40] N. Wagner and T.A. Marsh, *Tail index estimation in small samples. Simulation results for independent and ARCH-type financial return models*, Statist. Papers 45 (2004), pp. 545-561.