Bernoulli **20**(2), 2014, 979–989
DOI: 10.3150/13-BEJ512

Lower bounds to the accuracy of inference on heavy tails

S.Y. NOVAK

Middlesex University, The Burroughs, London NW44BT, UK. E-mail: S.Novak@mdx.ac.uk

The paper suggests a simple method of deriving minimax lower bounds to the accuracy of statistical inference on heavy tails. A well-known result by Hall and Welsh (*Ann. Statist.* **12** (1984) 1079–1084) states that if $\hat{\alpha}_n$ is an estimator of the tail index α_P and $\{z_n\}$ is a sequence of positive numbers such that $\sup_{P\in\mathcal{D}_r}\mathbb{P}(|\hat{\alpha}_n-\alpha_P|\geq z_n)\to 0$, where $\mathcal{D}_r$ is a certain class of heavy-tailed distributions, then $z_n\gg n^{-r}$. The paper presents a non-asymptotic lower bound to the probabilities $\mathbb{P}(|\hat{\alpha}_n-\alpha_P|\geq z_n)$. We also establish non-uniform lower bounds to the accuracy of tail constant and extreme quantiles estimation. The results reveal that normalising sequences of robust estimators should depend in a specific way on the tail index and the tail constant.

Keywords: heavy-tailed distribution; lower bounds

1. Introduction

A growing number of publications is devoted to the problem of statistical inference on heavy-tailed distributions. Such distributions naturally appear in finance, meteorology, hydrology, teletraffic engineering, etc. [4, 13]. In particular, it is widely accepted that frequent financial data (e.g., daily and hourly log-returns of share prices, stock indexes and currency exchange rates) often exhibits heavy tails [4, 5, 10, 11], while less frequent financial data is typically light-tailed. The heaviness of a tail of the distribution appears to be responsible for extreme movements of stock indexes and share prices. The tail index indicates how heavy the tail is; extreme quantiles are used as measures of financial risk [4, 11]. The need to evaluate the tail index and extreme quantiles stimulated research on methods of statistical inference on heavy-tailed data.

The distribution of a random variable (r.v.) X is said to have a *heavy right tail* if

$$\mathbb{P}(X\geq x)=L(x)x^{-\alpha}\qquad(\alpha>0),\tag{1}$$

where the (unknown) function L is slowly varying at infinity:

$$\lim_{x\to\infty}L(xt)/L(x)=1\qquad(\forall t>0).$$

We denote by $\mathcal{H}$ the non-parametric class of distributions obeying (1).

The tail index α is the main characteristic describing the tail of a distribution. If $L(x)=c+\mathrm{o}(1)$, then c is called the tail constant.

Let $F(\cdot)=\mathbb{P}(X<\cdot)$ denote the distribution function (d.f.). Obviously, the tail index is a functional of the distribution function:

$$\alpha_F \equiv \alpha_P = -\lim_{x\to\infty} \frac{\ln \mathbb{P}(X \geq x)}{\ln x}. \tag{2}$$

If $L(x)$ tends to a constant (say, c_F) as $x\to\infty$, then the tail constant is also a functional of F:

$$c_F \equiv c_P = \lim_{x\to\infty} x^{\alpha_F}\mathbb{P}(X\geq x).$$

The statistical inference on a heavy-tailed distribution is straightforward if the class of unknown distributions is assumed to be a regular parametric family. The drawback of the parametric approach is that one usually cannot reliably check whether the unknown distribution belongs to a chosen parametric family.

A lot of attention during the past three decades has been given to the problem of reliable inference on heavy tails without parametric assumptions. The advantage of the non-parametric approach is that a class of unknown distributions, $\mathcal{P}$, is so large that the problem of testing the hypothesis that the unknown distribution belongs to $\mathcal{P}$ does not arise. The disadvantage of the non-parametric approach is that virtually no question concerning inference on heavy tails can be given a simple answer. In particular, the problem of establishing a lower bound to the accuracy of tail index estimation remained open for decades.

A lower bound to the accuracy of statistical inference sets a benchmark against which the accuracy of any particular estimator can be compared. When looking for an estimator $\hat{a}_n$ of a quantity of interest, a_P, where $P\in\mathcal{P}$ is the unknown distribution, $\mathcal{P}$ is the class of distributions and a_P is a functional of P, one often would like to choose an estimator that minimises a loss function uniformly in $\mathcal{P}$ (e.g., $\sup_{P\in\mathcal{P}}\mathbb{E}_P\ell(|\hat{a}_n-a_P|)$, where ℓ is a particular loss function). A lower bound to $\sup_{P\in\mathcal{P}}\mathbb{E}_P\ell(|\hat{a}_n-a_P|)$ follows if one can establish a lower bound to

$$\sup_{P\in\mathcal{P}}\mathbb{P}(|\hat{a}_n-a_P|\geq u) \qquad (u>0).$$

The first step towards establishing a lower bound to the accuracy of tail index estimation was made by Hall and Welsh [7], who proved the following result. Note that the class $\mathcal{H}$ of heavy-tailed distributions is too "rich" for meaningful inference, and one usually deals with a subclass of $\mathcal{H}$, imposing certain restrictions on the asymptotics of $L(\cdot)$. Hall and Welsh dealt with the class $\mathcal{D}_{b,A}\equiv\mathcal{D}_{b,A}(\alpha_0,c_0,\varepsilon)$ of distributions on $(0;\infty)$ with densities

$$f(x)=c\alpha x^{-\alpha-1}(1+u(x)), \tag{3}$$

where $\sup_{x>0}|u(x)|x^{b\alpha}\leq A,\ |\alpha-\alpha_0|\leq\varepsilon,\ |c-c_0|\leq\varepsilon$. Note that the range of possible values of the tail index is restricted to interval $[\alpha_0-\varepsilon;\alpha_0+\varepsilon]$. Let

$$\hat{\alpha}_n\equiv\hat{\alpha}_n(X_1,\dots,X_n)$$

be an arbitrary tail index estimator, where $X_1, \ldots, X_n$ are independent and identically distributed (i.i.d.) random variables, and let $\{z_n\}$ be a sequence of positive numbers. If

$$\lim_{n\to\infty} \sup_{F\in\mathcal{D}_{b,A}} \mathbb{P}_F(|\hat{\alpha}_n - \alpha_F| \geq z_n) = 0 \qquad (\forall A > 0), \tag{4}$$

then

$$z_n \gg n^{-b/(2b+1)} \qquad (n\to\infty)$$

(to be precise, Hall and Welsh [7] dealt with the random variables $Y_i = 1/X_i$, where X_i are distributed according to (3)).

Beirlant et al. [1] have a similar result for a larger class $\mathcal{P}$ of distributions but require the estimators are uniformly consistent in $\mathcal{P}$. Pfanzagl [12] has established a lower bound in terms of a modulus of continuity related to the total variation distance d_{TV}. Let $\mathcal{D}_b^+$ be the class of distributions with densities (3) such that $\sup_{x>0}|u(x)|x^{\alpha b} < \infty, \alpha > 0$, and set

$$s_n(\varepsilon, P_0) = \sup_{P\in\mathcal{P}_{n,\varepsilon}} |\alpha_P - \alpha_{P_0}|,$$

where α_P is the tail index of distribution P and $\mathcal{P}_{n,\varepsilon} = \{P \in \mathcal{D}_b^+ \colon\, d_{\mathrm{TV}}(P_0^n; P^n) \leq \varepsilon\}$ is a neighborhood of $P_0 \in \mathcal{D}_b^+$. Pfanzagl has showed that neither estimator can converge to α uniformly in $\mathcal{P}_{n,\varepsilon}$ with the rate better than $s_n(\varepsilon, P_0)$, and

$$\inf_{0<\varepsilon<1} \varepsilon^{-2b/(1+2b)} \liminf_{n\to\infty} n^{b/(1+2b)} s_n(\varepsilon, P_0) > 0.$$

Donoho and Liu [2] present a lower bound to the accuracy of tail index estimation in terms of a modulus of continuity $\Delta_A(n,\varepsilon)$. However, they do not calculate $\Delta_A(n,\varepsilon)$. The claim that a particular heavy-tailed distribution is stochastically dominant over all heavy-tailed distributions with the same tail index appears without proof. Assuming that the range of possible values of the tail index is restricted to an interval of fixed length, Drees [3] derives the asymptotic minimax risk for affine estimators of the tail index and indicates an approach to numerical computation of the asymptotic minimax risk for non-affine ones.

The paper presents a simple method of deriving minimax lower bounds to the accuracy of non-parametric inference on heavy-tailed distributions. The results are non-asymptotic, the constants in the bounds are shown explicitly, the range of possible values of the tail index is not restricted to an interval of fixed length. The information functional seems to be found for the first time, as well as the lower bound to the accuracy of extreme quantiles estimation.

The results indicate that the traditional minimax approach may require revising. The classical approach suggests looking for an estimator $\hat{a}_n$ that minimises, say,

$$\sup_{P\in\mathcal{P}} \mathbb{E}_P|\hat{a}_n - a_P|$$

(cf. [8, 9, 14]), while our results suggest looking for an estimator $\hat{a}_n^*$ that minimises

$$\sup_{P\in\mathcal{P}} g_P \mathbb{E}_P|\hat{a}_n^* - a_P|,$$

where g_P is the "information functional" (an analogue of Fisher's information). Theorems 1–4 reveal the information functionals and indicate that the normalising sequence of a robust estimator should depend in a specific way on the characteristics of the unknown distribution.

2. Results

In the sequel, we deal with the non-parametric class

$$\mathcal{H}(b) = \Big\{P \in \mathcal{H}\colon \sup_{x>K_*(P)} |c_F^{-1} x^{\alpha_F} P(X \ge x) - 1| x^{b\alpha_F} < \infty\Big\} \tag{5}$$

of distributions on $(0;\infty)$, where $b > 0$ and $K_*(P)$ is the left end-point of the distribution. If $\mathcal{L}(X) \in \mathcal{H}(b)$, then

$$\mathbb{P}(X \ge x) = c_F x^{-\alpha_F}(1 + \mathrm{O}(x^{-b\alpha_F})) \qquad (x \to \infty).$$

The class $\mathcal{H}(b)$ is larger than $\mathcal{D}_b^+$; the range of possible values of the tail index is not restricted to an interval of fixed length. Below, given a distribution function (d.f.) F_i, we put

$$a_{F_i} = 1/\alpha_{F_i}, \qquad r = b/(1 + 2b),$$

$\mathbb{E}_i$ means the mathematical expectation with respect to F_i and $\mathbb{P}_i$ is the corresponding distribution. We set $K \equiv K_{\alpha,b,c} = \alpha^{-2r} c^{-\alpha r} \mathrm{e}^{-1}(c^{\alpha b} \wedge \mathrm{e}^{-2b})$.

Theorem 1. *For any $\alpha > 0$, $c > 0$, any tail index estimator $\hat{\alpha}_n$ and any estimator $\hat{a}_n$ of index $a = 1/\alpha$ there exist d.f.s $F_0, F_1 \in \mathcal{H}(b)$ such that $\alpha_{F_0} = \alpha, c_{F_0} = c^{-\alpha}$, and*

$$\max_{i\in\{0;1\}} \mathbb{P}_i(|\hat{\alpha}_n/\alpha_{F_i} - 1|\alpha_{F_i}^{r/b} c_{F_i}^r n^r \ge v/2) \ge (1 - v^{1/r}/8n)^{2n}/4, \tag{6}$$

$$\max_{i\in\{0;1\}} \mathbb{P}_i(|\hat{a}_n/a_{F_i} - 1|a_{F_i}^{-r/b} c_{F_i}^r n^r \ge v/2) \ge (1 - v^{1/r}/8n)^{2n}/4 \tag{7}$$

as $n > 4\max\{\alpha^2 c^{-2\alpha b}; c^{2\alpha}\alpha^{-2/b}\}$ and $v \in [0; Kn^r]$.

Note that if $\max_{i\in\{0;1\}} \mathbb{P}_i(|\hat{\alpha}_n/\alpha_{F_i} - 1| \ge z_n) \to 0$ as $n \to \infty$, then for any $C > 0$ we have $z_n \ge Cn^{-r}$ for all large enough n, yielding $z_n \gg n^{-r}$. Thus, the Hall–Welsh result follows from (6).

Theorem 1 shows that the natural normalising sequence for $\hat{\alpha}_n/\alpha_F - 1$ is $n^{-r}\alpha_F^{-r/b} c_F^{-r}$. The information functional $g_F = \alpha_F^{r/b} c_F^r$ plays here the same role as Fisher's information function in the Fréchet–Rao–Cramér inequality.

Theorem 1 yields also minimax lower bounds to the moments of $|\hat{\alpha}_n/\alpha_{F_i} - 1|$. In particular, there holds

Corollary 2. *For any $\alpha > 0$, $c > 0$ there exist distribution functions $F_0, F_1 \in \mathcal{H}(b)$ such that $\alpha_{F_0} = \alpha, c_{F_0} = c^{-\alpha}$, and for any tail index estimator $\hat{\alpha}_n$*

$$\max_{i \in \{0;1\}} \alpha_{F_i}^{r/b} c_{F_i}^r \mathbb{E}_{F_i} |\hat{\alpha}_n/\alpha_{F_i} - 1| n^r \geq 4^r r\Gamma(r)/8 + \mathrm{o}(1). \tag{8}$$

The result holds if $\alpha_{F_i}^{r/b} c_{F_i}^r \mathbb{E}_{F_i} |\hat{\alpha}_n/\alpha_{F_i} - 1|$ in (8) is replaced with $a_{F_i}^{-r/b} c_{F_i}^r \mathbb{E}_{F_i} |\hat{a}_n/a_{F_i} - 1|$.

Let $\mathcal{H}_n(b) \subset \mathcal{H}(b)$ be a class of d.f.s such that $\inf_{F \in \mathcal{H}_n(b)} K_{\alpha_F, b, c_F} n^r \to \infty$ as $n \to \infty$. Then for any estimator $\hat{\alpha}_n$

$$\sup_{F \in \mathcal{H}_n(b)} \alpha_F^{r/b} c_F^r \mathbb{E}_F |\hat{\alpha}_n/\alpha_F - 1| n^r \geq 4^r r\Gamma(r)/8 + \mathrm{o}(1). \tag{8*}$$

A lower bound to $\mathbb{E}_F |\hat{\alpha}_n/\alpha_F - 1|$ seems to be established for the first time.

The presence of the information functional makes the bound non-uniform. Note that a uniform lower bound would be meaningless: as the range of possible values of α_F is not restricted to an interval of fixed length, it follows from (8*) that

$$\sup_{F \in \mathcal{H}_n(b)} \mathbb{E}_F |\hat{\alpha}_n/\alpha_F - 1| \to \infty \qquad (n \to \infty).$$

More generally, $\sup_{F \in \mathcal{H}_n(b)} \tilde{g}_F \mathbb{E}_F |\hat{\alpha}_n/\alpha_F - 1|$ may tend to ∞ as $n \to \infty$ if $\tilde{g}_F/g_F \neq \mathrm{const}$.

Let $\hat{c}_n$ be an arbitrary tail constant estimator. The next theorem presents a lower bound to the probabilities $\mathbb{P}_F(|\hat{c}_n - c_F| \geq x)$.

Theorem 3. *Let $\hat{c}_n$ be an arbitrary tail constant estimator. For any $\alpha \geq n^{-r/2}$ and $c > 0$ there exist distribution functions $F_0, F_1 \in \mathcal{H}(b)$ such that $\alpha_{F_0} = \alpha, c_{F_0} = c^{-\alpha}$, and for all large enough n, as $v \in [0; \alpha^{-2} c^{-\alpha} \ln n]$,*

$$\max_{i \in \{0;1\}} \mathbb{P}_i(|\hat{c}_n/c_{F_i} - 1| \alpha_{F_i}^{r/b} c_{F_i}^r \geq r v^r n^{-r} \ln(n/\ln n) t_n/2b) \geq (1 - v/8n)^{2n}/4, \tag{9}$$

where $t_n = \exp(-r(1-r) n^{-r/2} (\ln(n/\ln n))^{r+1}/b)$.

Similarly to (8) Theorem 3 yields lower bounds to the moments of $|\hat{c}_n/c_{F_i} - 1|$. In particular, (9) entails

$$\max_{i \in \{0;1\}} \alpha_{F_i}^{r/b} c_{F_i}^r \mathbb{E}_{F_i} |\hat{c}_n/c_{F_i} - 1| \geq (\ln n) n^{-r} r^2 4^{r-1} \Gamma(r)/(2b + \mathrm{o}(1)). \tag{9*}$$

According to Hall and Welsh [7],

$$z_n \gg (\ln n) n^{-b/(2b+1)}$$

if $\lim_{n \to \infty} \sup_{F \in \mathcal{D}_{b,A}} \mathbb{P}_F(|\hat{c}_n - c_F| \geq z_n) = 0$ $(\forall A > 0)$. This fact can be obtained as a consequence to Theorem 3: if $\max_{i \in \{0;1\}} \mathbb{P}_i(|\hat{c}_n - c_{F_i}| \geq z_n) \to 0$ as $n \to \infty$, then for any $C > 0$ we have $z_n \geq C n^{-r} \ln n$ for all large enough n, hence $z_n \gg n^{-r} \ln n$.

We now present a lower bound to the accuracy of estimating extreme upper quantiles. We call an upper quantile of level q "extreme" if $q \equiv q_n$ tends to 0 as n grows. In financial applications (see, e.g., [4, 11]), an upper quantile of the level as high as 0.05 can be considered extreme as the empirical quantile estimator appears unreliable. Of course, there is an infinite variety of possible rates of decay of q_n. Theorem 4 presents lower bounds in the case $q_n = sn^{-1/(1+2b)}$, where s is bounded away from 0 and ∞.

Set $\bar{F} = 1 - F$. We denote the upper quantile of level q_n by

$$x_{F,n} = \bar{F}^{-1}(q_n).$$

Let $\hat{x}_n$ be an arbitrary estimator of $x_{F,n}$. Denote $w_{F_i} \equiv w_{F_i}(\alpha_{F_i}, c_{F_i}, b, s, u) = |\ln(u\alpha_{F_i}^{2r} c_{F_i}^{2br}/s^b)|$.

Theorem 4. *For any $\alpha > 0$, $c > 0$ there exist distribution functions $F_0, F_1 \in \mathcal{H}(b)$ such that $\alpha_{F_0} = \alpha, c_{F_0} = c^{-\alpha}$, and for all large enough n and $u \in (s^b\alpha^{-2r}c^{2\alpha br}; Kn^r)$,*

$$\max_{i\in\{0;1\}} \mathbb{P}_i(|\hat{x}_n/x_{F_i,n} - 1|\alpha_{F_i}^{2(1-r)} c_{F_i}^r / w_{F_i} t_{i,n}^\star \geq un^{-r}/2b) \geq (1 - u^{1/r}/8n)^{2n}/4, \quad (10)$$

$$\max_{i\in\{0;1\}} \mathbb{P}_i(|x_{F_i,n}/\hat{x}_n - 1|\alpha_{F_i}^{2(1-r)} c_{F_i}^r / w_{F_i} t_{i,n}^\star \geq un^{-r}/2b) \geq (1 - u^{1/r}/8n)^{2n}/4, \quad (11)$$

where $\max_{i\in\{0;1\}} |t_{i,n}^\star - 1| \to 0$ *as* $n \to \infty$.

3. Proofs

Our approach to establishing lower bounds requires constructing two distribution functions F_0 and F_1, where F_0 is a Pareto d.f. and $F_1 \equiv F_{1,n}$ is a "disturbed" version of F_0. We then apply Lemma 5 that provides a non-asymptotic lower bound to the accuracy of estimation when choosing between two close alternatives.

The problem of estimating the tail index, the tail constant and $x_{F,n}$ from $X_1, \ldots, X_n$ is equivalent to the problem of estimating α_F, c_F and quantiles from a sample $Y_1, \ldots, Y_n$ of i.i.d. positive r.v.s with the distribution

$$F(y) \equiv \mathbb{P}(Y \leq y) = y^\alpha \ell(y) \qquad (y > 0), \quad (12)$$

where function ℓ slowly varies at the origin.

We denote by $\mathcal{F}$ the class of distributions obeying (12). Note that $\mathcal{L}(Y) \in \mathcal{F}$ if and only if $\mathcal{L}(1/Y) \in \mathcal{H}$. Obviously, a tail index estimator $\alpha_n(X_1, \ldots, X_n)$ can be considered an estimator $\alpha_n(1/Y_1, \ldots, 1/Y_n)$ of index α from the sample $Y_1 = 1/X_1, \ldots, Y_n = 1/X_n$, and vice versa. The tradition of dealing with this equivalent problem stems from [6]. We proceed with this equivalent formulation.

A counterpart to $\mathcal{H}(b)$ is the following non-parametric class of d.f.s on $(0; \infty)$:

$$\mathcal{F}(b) = \Big\{F \in \mathcal{F}\colon \sup_{y<K^*(F)} |c_F^{-1} y^{-\alpha_F} F(y) - 1| y^{-b\alpha_F} < \infty\Big\}, \quad (13)$$

where $b > 0$ and $K^*(F)$ is the right end-point of F. A d.f. $F \in \mathcal{F}(b)$ obeys

$$F(y) = c_F y^{\alpha_F}(1 + \mathrm{O}(y^{b\alpha_F})) \qquad (y \to 0),$$

where $\alpha_F = \lim_{y\downarrow 0}(\ln F(y))/\ln y$ and $c_F = \lim_{y\downarrow 0} y^{-\alpha_F} F(y)$.

Proof of Theorem 1. Let $h \in (0;c)$, and denote

$$\alpha_0 = \alpha, \qquad \alpha_1 = \alpha + \gamma, \qquad \gamma = h^{\alpha b}.$$

We will employ the distribution functions F_0 and F_1, where

$$\begin{aligned} F_0(y) &= (y/c)^\alpha \mathbb{1}\{0 < y \le c\}, \\ F_1(y) &= (h/c)^{-\gamma}(y/c)^{\alpha_1}\mathbb{1}\{0 < y \le h\} + (y/c)^\alpha \mathbb{1}\{h < y \le c\}. \end{aligned}$$

The counterparts to these distributions are

$$\begin{aligned} \mathbb{P}_0(X > x) &= (cx)^{-\alpha}\mathbb{1}\{x \ge 1/c\}, \\ \mathbb{P}_1(X > x) &= (cx)^{-\alpha}\mathbb{1}\{1/c \le x < 1/h\} + c^{-\alpha}h^{-\gamma}x^{-\alpha_1}\mathbb{1}\{x \ge 1/h\}. \end{aligned}$$

It is easy to see that $F_1 \le F_0$ and

$$\alpha_{F_0} = \alpha, \qquad \alpha_{F_1} = \alpha_1, \qquad c_{F_0} = c^{-\alpha}, \qquad c_{F_1} = c^{-\alpha}h^{-\gamma}. \tag{14}$$

Obviously, $F_0 \in \mathcal{F}(b)$. We now check that $F_1 \in \mathcal{F}(b)$.

Since

$$c_{F_1}^{-1} y^{-\alpha_1} F_1(y) = y^{-\gamma}h^\gamma \qquad (h < y \le c),$$

we have

$$\sup_{0<y\le c} |1 - c_{F_1}^{-1} y^{-\alpha_1} F_1(y)| y^{-b\alpha_1} = \sup_{h<y\le c} (1 - y^{-\gamma}h^\gamma) y^{-b\alpha_1}. \tag{15}$$

The right-hand side of (15) takes on its maximum at $y_0 = h(1+\gamma/b\alpha_1)^{1/\gamma}$; the supremum is bounded by $A := \mathrm{e}^{1/\mathrm{e}\alpha}/b\alpha$. Note that $\{F_0, F_1\} \subset \mathcal{D}_{b,A}$.

Let $d_H^2(P_0;P_1)$ denote the Hellinger distance. It is easy to check that

$$d_H^2(F_0;F_1) \le \gamma^{1/r}/8\alpha^2 c^\alpha. \tag{16}$$

According to Lemma 5 below,

$$\max_{i\in\{0;1\}} \mathbb{P}_i(|\hat{\alpha}_n - \alpha_{F_i}| \ge \gamma/2) \ge (1 - \gamma^{1/r}/8\alpha^2 c^\alpha)^{2n}/4. \tag{17}$$

Let $\gamma = \gamma_n$, where

$$\gamma_n \equiv \gamma_n(\alpha, c, v) = v(\alpha^2 c^\alpha/n)^r.$$

Note that $h<c$ as $n>\alpha^2c^{-2b\alpha}v^{1/r}$. From (17),

$$\max_{i\in\{0;1\}}\mathbb{P}_i(|\hat\alpha_n/\alpha_{F_i}-1|\alpha_{F_i}^{r/b}c_{F_i}^r n^r\ge vt_{n,i}/2)\ge(1-v^{1/r}/8n)^{2n}/4, \qquad (18)$$

where $t_{n,0}=1$ and $t_{n,1}=1/f(\gamma)$, $f(\gamma)=(1+\gamma/\alpha)^2\gamma^{\gamma/\alpha b}$. Note that $f(\gamma)\le 1$ as $\gamma\le \mathrm{e}^{-1-2b}$. Hence, $t_{n,1}\ge 1$ as $v\in[0;Kn^r]$ and (6) follows.

Let $\hat a_n$ be an arbitrary estimator of index $a=1/\alpha$. Denote $a=a_0$. Since $|a_0-a_1|=\gamma aa_1$, Lemma 5 yields

$$\max_{i\in\{0;1\}}\mathbb{P}_i(|\hat a_n-a_{F_i}|\ge\gamma aa_1/2)\ge(1-\gamma^{1/r}/8\alpha^2c^\alpha)^{2n}/4.$$

With $\gamma=\gamma_n$, the left-hand side of this inequality is

$$\max_{i\in\{0;1\}}\mathbb{P}_i(|\hat a_n-a_{F_i}|\ge vn^{-r}a^{1-2r}a_1/2c_{F_0}^r)=\max_{i\in\{0;1\}}\mathbb{P}_i(|\hat a_n/a_{F_i}-1|a_{F_i}^{-r/b}c_{F_i}^r n^r\ge vt^+_{n,i}/2),$$

where $t^+_{n,0}=(1+\gamma a)^r\ge 1$ and $t^+_{n,1}=(1+\gamma a)^{r/b}\gamma^{-r\gamma a/b}\ge 1$, leading to (7). □

Proof of Corollary 2. Note that

$$\mathbb{E}\xi=\int_0^\infty\mathbb{P}(\xi\ge x)\,\mathrm{d}x \qquad (19)$$

for any non-negative r.v. ξ. Since

$$\int_0^{z_n}(1-v^{1/r}/8n)^{2n}\,\mathrm{d}v=4^r r\Gamma(r)+\mathrm{o}(1) \qquad (n\to\infty)$$

as $z_n\to\infty$, $z_n=\mathrm{o}(n^r)$, (6) and (19) entail (8). □

Proof of Theorem 3. With F_0 and F_1 defined as above, we have

$$c_{F_1}-c_{F_0}=c^{-\alpha}(\gamma^{-\gamma/\alpha b}-1)\ge c^{-\alpha}\gamma|\ln\gamma|/\alpha b.$$

Using this inequality, (17) and Lemma 5, we derive

$$\max_{i\in\{0;1\}}\mathbb{P}_i(|\hat c_n-c_{F_i}|\ge c^{-\alpha}\gamma|\ln\gamma|/2\alpha b)\ge(1-\gamma^{1/r}/8\alpha^2c^\alpha)^{2n}/4.$$

Let $\gamma\equiv\gamma(n)=(v\alpha^2c^\alpha/n)^r$. Then

$$\max_{i\in\{0;1\}}\mathbb{P}_i(|\hat c_n-c_{F_i}|\ge c_{F_0}(v\alpha^2c^\alpha/n)^r r\ln(n/\ln n)/2\alpha b)\ge(1-v/8n)^{2n}/4.$$

Note that $\alpha^2c^\alpha/\alpha_1^2c_{F_1}^{-1}\ge t_n$ as $v\in[0;\alpha^{-2}c^{-\alpha}\ln n]$. The result follows. □

Proof of Theorem 4. Denote

$$x_i\equiv x_{F_i,n}, \qquad y_i=1/x_i.$$

Obviously, y_i is the quantile of $\mathcal{L}_i(1/X)$. We find convenient dealing with the equivalent problem of estimating quantiles of the distribution of a random variable $Y = 1/X$.

With functions F_0, F_1 defined as above, it is easy to see that

$$y_0 = c q_n^{1/\alpha} = c\kappa h, \qquad y_1 = c^{\alpha/\alpha_1} q_n^{1/\alpha_1} h^{\gamma/\alpha_1} = y_0 (c\kappa)^{-\gamma/\alpha_1}, \tag{20}$$

where we put $\kappa = q_n^{1/\alpha}/h$. Note that $y_1 = h(c\kappa)^{1-\gamma/\alpha_1}$. Hence $y_i < h$ if $c\kappa < 1$ ($i \in \{0, 1\}$).

Denote

$$\gamma \equiv \gamma_n(\alpha, b, c) = u(\alpha^2 c^\alpha / n)^r. \tag{21}$$

Then $\kappa = s^{1/\alpha}(\alpha^2 c^\alpha)^{-r/\alpha b} u^{-1/\alpha b}$ and

$$c\kappa = u^{-1/\alpha b} s^{1/\alpha} c^{2r} \alpha^{-2r/\alpha b} < 1 \tag{22}$$

by the assumption.

Using the facts that $\mathrm{e}^x - 1 \geq x\mathrm{e}^{x/2}$ and $1 - \mathrm{e}^{-x} \geq x\mathrm{e}^{-x/2}$ as $x \geq 0$, we derive

$$\begin{aligned} y_1 - y_0 &= y_0((c\kappa)^{-\gamma/\alpha_1} - 1) \\ &\geq \gamma^{1+1/\alpha b}(c\kappa)^{1-\gamma/2\alpha_1} |\ln c\kappa| / \alpha_1. \end{aligned}$$

Hence, $(y_1 - y_0)/y_0 \geq \gamma |\ln c\kappa|/\alpha_1$ and $(y_1 - y_0)/y_1 = 1 - (c\kappa)^{\gamma/\alpha_1} \geq \gamma |\ln c\kappa| (c\kappa)^{\gamma/2\alpha_1}/\alpha_1$. By Lemma 5,

$$\max_{i \in \{0;1\}} \mathbb{P}_i(|\hat{y}_n - y_i| \geq \gamma^{1+1/\alpha b}(c\kappa)^{1-\gamma/2\alpha_1} |\ln c\kappa|/2\alpha_1) \geq (1 - \gamma^{1/r}/8\alpha^2 c^\alpha)^{2n}/4$$

for any estimator $\hat{y}_n$. Thus,

$$\max_{i \in \{0;1\}} \mathbb{P}_i(|\hat{y}_n/y_i - 1| \geq \gamma |\ln(c\kappa)^{\alpha b}| t^\star_{n,i}/2b\alpha^2_{F_i}) \geq (1 - \gamma^{1/r}/8\alpha^2 c^\alpha)^{2n}/4,$$

where $t^\star_{n,0} = 1/(1 + \gamma/\alpha) = 1/(1 + u\alpha_{F_0}^{-r/b} c_{F_0}^{-r} n^{-r})$ and $t^\star_{n,1} = (1 + \gamma/\alpha)(c\kappa)^{\gamma/2\alpha}$. Taking into account (21) and (22), we derive

$$\max_{i \in \{0;1\}} \mathbb{P}_i(|\hat{y}_n/y_i - 1| \geq u\alpha_{F_i}^{2(r-1)} c_{F_i}^{-r} n^{-r} \ln(u\alpha^{2r}/s^b c^{2r\alpha b}) t^\star_{n,i}/2b) \geq (1 - u^{1/r}/8n)^{2n}/4.$$

This leads to (11).

Recall that $x_i = 1/y_i$. From (20),

$$|x_1 - x_0| = |y_1 - y_0|/y_0 y_1 \geq \gamma^{1-1/\alpha b}(c\kappa)^{-1+\gamma/2\alpha_1} |\ln c\kappa|/\alpha_1.$$

By Lemma 5,

$$\max_{i \in \{0;1\}} \mathbb{P}_i(|\hat{x}_n - x_i| \geq \gamma^{1-1/\alpha b}(c\kappa)^{-1+\gamma/2\alpha_1} |\ln c\kappa|/2\alpha_1) \geq (1 - \gamma^{1/r}/8\alpha^2 c^\alpha)^{2n}/4.$$

Hence,

$$\max_{i\in\{0;1\}} \mathbb{P}_i(|\hat{x}_n/x_i - 1| \geq u\alpha_{F_i}^{2(r-1)} c_{F_i}^{-r} n^{-r} |\ln(s^b c^{2r\alpha b}/\alpha^{2r} u)| \tilde{t}_{n,i}/2b) \geq (1 - u^{1/r}/8n)^{2n}/4,$$

where $\tilde{t}_{n,0} = (c\kappa)^{\gamma/2\alpha}/(1+\gamma/\alpha) = (u^{-1/\alpha b} s^{1/\alpha} c^{2r} \alpha^{-2r/\alpha b})^{\gamma/2\alpha}/(1 + u\alpha_{F_0}^{-r/b} c_{F_0}^{-r} n^{-r})$ and $\tilde{t}_{n,1} = 1$. The proof is complete. □

The next lemma presents a lower bound to the accuracy of choosing between two "close" alternatives.

Let $\mathcal{P}$ be an arbitrary class of distributions, and assume that the quantity of interest, a_P, is an element of a metric space $(\mathcal{X}, d)$. An estimator $\hat{a}$ of a_P is a measurable function of $X_1, \ldots, X_n$ taking values in a subspace $\{a_P\colon\ P \in \mathcal{P}\}$ of the metric space $(\mathcal{X}, d)$.

Examples of functionals a_P include (a) $a_{P_\theta} = \theta$, where $\mathcal{P} = \{P_\theta, \theta \in \Theta\}$ is a parametric family of distributions ($\Theta \subset \mathbb{R}^m$); (b) $a_P = f_P$, where f_P is the density of P with respect to a particular measure; (c) $a_P = P$. A minimax lower bound over $\mathcal{P}$ follows from a lower bound to $\max_{i\in\{0;1\}} \mathbb{P}_i(d(\hat{a}; a_{P_i}) \geq \delta)$, where $P_0, P_1 \in \mathcal{P}$.

Lemma 5. *Denote $2\delta = d(a_{P_1}; a_{P_0})$. Then*

$$\max_{i\in\{0;1\}} \mathbb{P}_i(d(\hat{a}; a_{P_i}) \geq \delta) \geq (1 - d_H^2)^{2n}/4, \tag{23}$$

where $d_H \equiv d_H(P_0; P_1)$ is the Hellinger distance.

There is considerable literature on techniques of deriving minimax lower bounds of this kind (cf. [8, 9, 14]). Classical results include Fano's and Assuad's lemmas. Inequality (23) is sharper than Lemma 1 in [8]. Another related result is Theorem 2.2 in [14].

Proof of Lemma 5. Recall that

$$d_H^2(P_0; P_1) = \frac{1}{2} \int (f_0^{1/2} - f_1^{1/2})^2 = 1 - \int \sqrt{f_0 f_1},$$

where f_i is a density of P_i with respect to a certain measure (e.g., $P_0 + P_1$).

Let $f_{i,n}$ denote the density of $\mathcal{L}_i(X_1, \ldots, X_n)$, and put $a_i = a_{P_i}$. By the triangle inequality, $2\delta \leq d(a_{P_0}; \hat{a}) + d(\hat{a}; a_{P_1})$. Therefore, $1 \leq \mathbb{1}_0 + \mathbb{1}_1$, where

$$\mathbb{1}_0 = \mathbb{1}\{d(a_0; \hat{a}) \geq \delta\}, \qquad \mathbb{1}_1 = \mathbb{1}\{d(\hat{a}; a_1) \geq \delta\}.$$

Using the definition of d_H and the Bunyakovskiy–Cauchy–Schwarz inequality, we derive

$$\begin{aligned}(1 - d_H^2)^n &= \int \sqrt{f_{0,n} f_{1,n}} \\ &\leq \int \sqrt{f_{0,n} f_{1,n}}\mathbb{1}_0 + \int \sqrt{f_{0,n} f_{1,n}}\mathbb{1}_1 \\ &\leq \mathbb{P}_0^{1/2}(d(\hat{a}; a_0) \geq \delta) + \mathbb{P}_1^{1/2}(d(\hat{a}; a_1) \geq \delta).\end{aligned}$$

Hence $(1 - d_H^2)^{2n} \leq 2(\mathbb{P}_0(d(\hat{a}; a_0) \geq \delta) + \mathbb{P}_1(d(\hat{a}; a_1) \geq \delta))$, leading to (23). □

Acknowledgements

The author is grateful to the Editor, the Associate Editor and two referees for many helpful remarks. Supported by a grant from the London Mathematical Society.

References

[1] Beirlant, J., Bouquiaux, C. and Werker, B.J.M. (2006). Semiparametric lower bounds for tail index estimation. *J. Statist. Plann. Inference* **136** 705–729. MR2181974

[2] Donoho, D.L. and Liu, R.C. (1991). Geometrizing rates of convergence. II, III. *Ann. Statist.* **19** 633–667, 668–701. MR1105839

[3] Drees, H. (2001). Minimax risk bounds in extreme value theory. *Ann. Statist.* **29** 266–294. MR1833966

[4] Embrechts, P., Klüppelberg, C. and Mikosch, T. (1997). *Modelling Extremal Events: For Insurance and Finance*. *Applications of Mathematics (New York)* **33**. Berlin: Springer. MR1458613

[5] Fama, E.F. and Roll, R. (1968). Some properties of symmetric stable distributions. *J. Amer. Statist. Assoc.* **63** 817–836. MR0233452

[6] Hall, P. (1982). On estimating the endpoint of a distribution. *Ann. Statist.* **10** 556–568. MR0653530

[7] Hall, P. and Welsh, A.H. (1984). Best attainable rates of convergence for estimates of parameters of regular variation. *Ann. Statist.* **12** 1079–1084. MR0751294

[8] Huber, C. (1997). Lower bounds for function estimation. In *Festschrift for Lucien Le Cam* 245–258. New York: Springer. MR1462949

[9] Ibragimov, I.A. and Has'minskiĭ, R.Z. (1981). *Statistical Estimation: Asymptotic Theory*. *Applications of Mathematics* **16**. New York: Springer. MR0620321

[10] Mandelbrot, B.B. (1963). New methods in statistical economics. *J. Political Economy* **71** 421–440.

[11] Novak, S.Y. (2011). *Extreme Value Methods with Applications to Finance*. *Monographs on Statistics and Applied Probability* **122**. Boca Raton, FL: CRC Press. MR2933280

[12] Pfanzagl, J. (2000). On local uniformity for estimators and confidence limits. *J. Statist. Plann. Inference* **84** 27–53. MR1747496

[13] Resnick, S.I. (1997). Heavy tail modeling and teletraffic data. *Ann. Statist.* **25** 1805–1869. MR1474072

[14] Tsybakov, A.B. (2009). *Introduction to Nonparametric Estimation*. *Springer Series in Statistics*. New York: Springer. MR2724359

Received June 2012 and revised January 2013