Novak, Serguei (2013) On the accuracy of inference on heavy-tailed distributions. Theory of Probability and Its Applications, 58 (3). pp. 509-518. ISSN 0040-585X

Final accepted version (with author's formatting)

This version is available at: http://eprints.mdx.ac.uk/18124/

THEORY PROBAB. APPL.
Vol. 58, No. 3, pp. 1–662

ON THE ACCURACY OF INFERENCE ON HEAVY-TAILED DISTRIBUTIONS*

S. Y. NOVAK†

(*Translated by the author*)

Abstract. This paper suggests a simple method of deriving nonparametric lower bounds of the accuracy of statistical inference on heavy-tailed distributions. We present lower bounds of the mean squared error of the tail index, the tail constant, and extreme quantiles estimators. The results show that the normalizing sequences of robust estimators must depend in a specific way on the tail index and the tail constant.

Key words. lower bounds, heavy-tailed distribution

DOI. 10.1137/S0040585X97986710

1. Introduction. Heavy-tailed distributions naturally appear in finance, meteorology, hydrology, teletraffic engineering, etc. (see [10], [20]). In particular, it is widely observed that frequent financial data often exhibit heavy tails [5], [10], [15].

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

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

where $\alpha > 0$ and the (unknown) function L is slowly varying at infinity:

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

We denote by $\mathcal{H}$ the class of distributions with a heavy right tail.

The number α in (1) is called the *tail index*. It is the main characteristic describing the tail of a distribution. If $L(x) = C + o(1)$, then C is called the tail constant.

The problem of tail index estimation turned out to be a challenge; it has attracted the attention of researchers for decades (see [10], [18], [20], and references therein). Consistency and asymptotic normality have been established for a number of tail index estimators (see [10], [18]). However, the problem of establishing a lower bound of the mean squared error (MSE) of a tail index estimator remained open.

The famous Fréchet–Rao–Cramér inequality gives a lower bound of MSE of an estimator in a regular parametric case; lower bounds are known also for parametric families with certain irregularities [14], [21].

Note that $\mathcal{H}$ is a nonparametric class of distributions. It is typical of nonparametric estimation problems that estimators are functions of a tuning (nuisance) parameter; cf. (15). This makes estimation far from straightforward.

The first step towards establishing a lower bound of the accuracy of tail index estimation was made by Hall and Welsh [8], who proved the following result. Let $\mathcal{D}_A(\alpha_0, C_0, \varepsilon, b)$ be the class of distributions on $(0;\infty)$ with densities

$$f(x) = C\alpha x^{-\alpha-1}(1 + r(x)), \tag{2}$$

*Received by the editors December 20, 2009; revised March 3, 2013. This work was supported by a grant from the London Mathematical Society.
http://www.siam.org/journals/tvp/58-3/98671.html
†Department of Design Engineering and Mathematics Middlesex University, The Burroughs, London NW4 4BT, UK (S.Novak@mdx.ac.uk).

where $\sup_{x>0}|r(x)|x^{b\alpha} \leqq A$, $|\alpha-\alpha_0| \leqq \varepsilon$, $|C-C_0| \leqq \varepsilon$. Denote by $\widehat{\alpha}_n \equiv \widehat{\alpha}_n(X_1,\ldots,X_n)$ an arbitrary tail index estimator, where $X, X_1,\ldots,X_n$ are independent and identically distributed (i.i.d.) random variables, and let $\{z_n\}$ be a sequence of positive numbers. If

$$\lim_{n\to\infty}\sup_{F\in\mathcal{D}_A}\mathbf{P}_F(|\widehat{\alpha}_n-\alpha|>z_n)=0 \qquad (\forall A>0) \tag{3}$$

for some $\alpha_0>0$, $C_0>0$, $b>0$, $\varepsilon>0$, then

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

(to be precise, Hall and Welsh dealt with the r.v.'s $Y_i=1/X_i$, where X_i are distributed according to (2)). Beirlant, Bouquiaux, and Werker [1] established a similar result for a larger class $\mathcal{P}$ of distributions but required the estimators to be $O_{\mathbf{P}}(1)$ uniformly over $\mathcal{P}$.

A related result was established by Pfanzagl [19]. Let $\mathcal{D}_b$ be the class of distributions with densities (2) such that $\sup_{x>0}|r(x)|x^{\alpha b}<\infty$, $\alpha>0$. Denote

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

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

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

Donoho and Liu [3] present a lower bound of 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 is stated without a proof.

Drees [4] 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 nonaffine estimators.

Hall and Welsh [8] showed also that $z_n \gg (\log n)n^{-b/(2b+1)}$ if $\widehat{\alpha}_n-\alpha$ is replaced with $\widehat{C}_n-C$ in (3), where C is the tail constant and $\widehat{C}_n$ is an arbitrary tail constant estimator.

Among nonparametric families of heavy-tailed distributions considered in the literature, we should mention the class

$$\mathcal{H}_{a,b,c,d}=\Big\{\mathbf{P}\colon\ \mathbf{P}(X\geqq x)=cx^{-1/a}\Big(1+dx^{-b/a}(1+o(1))\Big)\Big\} \tag{4}$$

of distributions on $(1;\infty)$, where $a=1/\alpha>0$, $b>0$, $c\equiv c(a)>0$, $d\equiv d(a,b)\neq 0$ (see [17], [18]).

A few other classes and the comparison of Hill's and the ratio estimators of the tail index can be found in [18]. The parametric Pareto family

$$\mathbf{P}(X\geqq x)=x^{-1/a} \qquad (x\geqq 1,\ a>0) \tag{5}$$

can be considered a "limiting point" of $\mathcal{H}_{a,b,1,d}$, as the second index, b, tends to infinity. Note that

$$\mathbf{E}\left(\frac{a_n^*}{a} - 1\right)^2 = \frac{1}{n}, \qquad \left(\frac{a_n^*}{a} - 1\right)\sqrt{n} \Rightarrow \mathcal{N}(0;1) \quad (n \to \infty) \tag{6}$$

in the case of i.i.d. observations over the Pareto distribution (5), where $a_n^* = n^{-1}\sum_{i=1}^n \log X_i$.

This paper suggests a simple approach to establishing minimax lower bounds to the MSE. The method is based on Lemma 1, which is presented in section 3.

In the next section we apply the approach to the problems of statistical inference on heavy-tailed distributions and derive nonparametric lower bounds of the accuracy of tail index, tail constant, and extreme quantiles estimation. Lower bounds of the MSE of tail index and extreme quantile estimators seem to be established for the first time.

The bounds are higher for smaller α indicating that the estimation problem becomes harder for distributions with heavier tails. The results reveal that normalizing sequences of estimators depend in a specific way on the tail index and the tail constant. Proofs are presented in section 3.

2. Lower bounds. The traditional method of establishing nonparametric lower bounds advocates choosing "as many functions as possible which are distant from one another by no less than (a small quantity) $\delta > 0$" [12], [13], [9].

Another approach is based on constructing two "close" distribution functions (d.f.'s) F_0, F_1 [6], [8], [22].

We show that the latter approach is capable of producing minimax lower bounds to the MSEs of estimators of the tail index, the tail constant, and extreme quantiles.

It is well observed that the accuracy of estimation in the case of a nonparametric class of distributions is typically worse than in the regular parametric case. The accuracy of estimation depends on the degree of "richness" of a class of possible distributions: the richer the class, the harder the problem of choosing between close alternatives, and hence the lower the accuracy of estimation. For instance, in the case of the nonparametric density estimation problem, a researcher may deal with a class of distributions obeying a certain smoothing condition; the smoother the densities, the better the accuracy of estimation [12], while the rate of decay of the MSE can be arbitrarily poor if no restrictions are specified [2].

Similarly, the class $\mathcal{H}$ of heavy-tailed distributions appears too "rich" for meaningful inference. In what follows we deal with the nonparametric class

$$\mathcal{H}(b) = \Big\{P \in \mathcal{H}\colon \sup_{x > K_*(P)} |c_{F}^{-1} x^{\alpha_F} P(X \geqq x) - 1| x^{b\alpha_F} < \infty\Big\} \tag{7}$$

of distributions on $(0;\infty)$, where $b > 0$, $K_*(P)$ denotes the left end-point of a distribution, $\alpha_F \equiv \alpha_P$ is the tail index, and $c_F \equiv c_P$ is the tail constant. If $P \in \mathcal{H}(b)$, then

$$P(X \geqq x) = c_F x^{-\alpha_F}\left(1 + O(x^{-b\alpha_F})\right) \qquad (x \to \infty).$$

The problem of tail index estimation is equivalent to that of estimating α from a sample of i.i.d. nonnegative r.v.'s with the distribution

$$F(y) \equiv \mathbf{P}(Y < y) = y^{\alpha}\ell(y) \qquad (y > 0), \tag{8}$$

where function ℓ slowly varies at the origin. Denote by $\mathcal{F}$ the class of distributions obeying (8). Then $\alpha \equiv \alpha_F$ is a functional of F:

$$\alpha_F = \lim_{y \downarrow 0} \frac{\log F(y)}{\log y}. \tag{9}$$

If $\ell(y)$ tends to a constant (say, c_F) as $y \downarrow 0$, then the tail constant c_F is also a functional of F:

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

Note that $\mathcal{L}(Y) \in \mathcal{F}$ if and only if $\mathcal{L}(1/Y) \in \mathcal{H}$. The tradition of dealing with this equivalent problem stems from [7].

A counterpart to $\mathcal{H}(b)$ is the following nonparametric class of distribution functions on $(0;\infty)$:

$$\mathcal{F}(b) = \Big\{ F \in \mathcal{F}\colon \sup_{y < K^*(F)} |c_F^{-1} y^{-\alpha_F} F(y) - 1| y^{-b\alpha_F} < \infty \Big\}, \tag{10}$$

where b is a positive number and $K^*(F)$ is the right end-point of F. A distribution function (d.f.) $F \in \mathcal{F}(b)$ obeys

$$F(y) = c_F y^{\alpha_F} \left(1 + O(y^{b\alpha_F})\right) \qquad (y \to 0).$$

A counterpart to the Pareto family is $\{F_a\}_{a>0}$, where $F_a(y) = y^{1/a}$, $0 < y \leqq 1$. More general is the parametric family $\{F_{\alpha,c},\ \alpha > 0,\ 0 < c \leqq 1\}$, where

$$F_{\alpha,c}(y) = \left(\frac{y}{c}\right)^{\alpha} \qquad (0 < y \leqq c). \tag{11}$$

If $Y_1, \ldots, Y_n$ is a sample of independent r.v.'s distributed according to (11), then the maximum likelihood estimator

$$\widehat{a}_n^* \equiv \frac{1}{\widehat{\alpha}_n^*} = \log\Big(\max_{i \leqq n} Y_i\Big) - n^{-1} \sum_{i=1}^n \log Y_i$$

has the MSE $\mathbf{E}\,(\widehat{a}_n^*/a - 1)^2 = n^{-1}$. Since the nonparametric class $\mathcal{F}(b)$ is much "richer," the rate of the accuracy of minimax estimation in $\mathcal{F}(b)$ is worse than n^{-1}.

Denote

$$r = b/(1+2b).$$

When we deal with a d.f. F_i, we put $\alpha_i = \alpha_{F_i}$, $a_i = 1/\alpha_i$; $\mathbf{E}_i$ means the expectation with respect to F_i.

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

$$\max_{i \in \{0;1\}} \alpha_{F_i}^{r/b} c_{F_i}^r\, \mathbf{E}_i^{1/2} \left(\frac{\widehat{\alpha}_n}{\alpha_{F_i}} - 1\right)^2 \geqq \left(\frac{8r}{ne}\right)^r \frac{t_n}{2}, \tag{12}$$

$$\max_{i \in \{0;1\}} a_{F_i}^{-r/b} c_{F_i}^r\, \mathbf{E}_i^{1/2} \left(\frac{\widehat{a}_n}{a_{F_i}} - 1\right)^2 \geqq \left(\frac{8r}{ne}\right)^r \frac{t_n}{2} \tag{13}$$

as $n > 4\max\{\alpha^2 c^{-2\alpha b}; c^\alpha \alpha^{-1/b}\}$, *where*

$$t_n = \left(1 - \left(\frac{4}{n}\right)^r \alpha^{-r/b} \max_{i\in\{0;1\}} (c^{r\alpha_{F_i}} \vee 1)\right) e^{-1/(4n)}.$$

According to (12), for any estimator $\widehat{\alpha}_n$ there exists a d.f. $F \in \mathcal{F}(b)$ such that

$$\mathbf{E}_F^{1/2}\left(\frac{\widehat{\alpha}_n}{\alpha_F} - 1\right)^2 \geqq \frac{1}{2}\left(\frac{8r}{ne}\right)^r \alpha_F^{-r/b} c_F^{-r}(1 + o(1)). \tag{14}$$

The smaller α_F is, the heavier is the tail and the higher is the lower bound of $\mathbf{E}_F^{1/2}(\widehat{\alpha}_n/\alpha_F - 1)^2$. Theorem 1 provides a background for the common opinion that inference on distributions with heavier tails is more difficult.

The important feature of the result is the dependence of the lower bound on α_F and c_F. One can say that the bounds are "nonuniform." Inequalities (12)–(14) indicate that the natural normalizing sequence for $\widehat{\alpha}_n/\alpha_F - 1$ is

$$n^{-r}\alpha_F^{-r/b} c_F^{-r}$$

(cf. (6), noting that $r \to 1/2$ and $r/b \to 0$ as $b \to \infty$). Moreover, a "uniform" lower bound would be meaningless: for any estimator $\widehat{\alpha}_n$

$$\sup_{F\in\mathcal{F}(b)} \mathbf{E}_F\left(\frac{\widehat{\alpha}_n}{\alpha_F} - 1\right)^2 n^r \to \infty \qquad (n \to \infty).$$

Apparently, b often equals 1 (as in the case of the Fréchet distribution) or 2 (as in the case of the Cauchy distribution); see [8], [10]. Hence the typical rates of estimation of the tail index are $n^{-1/3}$ and $n^{-2/5}$.

Denote by

$$\widehat{a}_n^{\mathrm{RE}}(x) = \frac{\sum_{i=1}^n \log(X_i/x)\mathbf{1}\{X_i \geqq x\}}{\sum_{i=1}^n \mathbf{1}\{X_i \geqq x\}}, \qquad \widehat{\alpha}_n^{\mathrm{RE}}(x) = \frac{1}{a_n^{\mathrm{RE}}(x)} \tag{15}$$

the ratio estimators (RE) of the tail index α and index $a = 1/\alpha$. The ratio estimator seems to be the only tail index estimator for which the asymptotics of the MSE is known:

$$\mathbf{E}\left(\frac{\widehat{a}_n^{\mathrm{RE}}(x)}{a} - 1\right) = v(x), \quad \mathbf{E}\left(\frac{\widehat{a}_n^{\mathrm{RE}}(x)}{a} - 1\right)^2 \sim \frac{1}{n\mathbf{P}(X \geqq x)} + v^2(x) \tag{16}$$

in the case of i.i.d. heavy-tailed r.v.'s, where

$$v(x) = a^{-1}\mathbf{E}\left\{\log\left(\frac{X}{x}\right) \,\Big|\, X \geqq x\right\} - 1 \tag{17}$$

(see [16], [18]). The ratio estimator was introduced by Goldie and Smith [11] (see [8], [11], [18] concerning estimators of the tail constant). For the ratio estimator $\widehat{a}_n^{\mathrm{RE}}(x_n)$ with threshold $x \equiv x_n \sim (n/8r\alpha^2 c^\alpha)^{r/\alpha b}$ we have

$$\max_{i\in\{0;1\}} a_{F_i}^{-r/b} c_{F_i}^r\, \mathbf{E}_{F_i}^{1/2}\left(\frac{\widehat{a}_n^{\mathrm{RE}}}{a_{F_i}} - 1\right)^2 \sim (8r)^{-r/(2b)} n^{-r} \qquad (n \to \infty), \tag{18}$$

where F_i is the d.f. of $\mathcal{L}_i(1/Y)$. The right-hand side of (18) differs from that of (13) only by the factor of $e^r/\sqrt{2r}$.

Let $\widehat{c}_n$ denote an arbitrary tail constant estimator. The following theorem presents a lower bound of the MSE of a tail constant estimator.

THEOREM 2. *For any $\alpha > 0$ and $c > 0$ there exist distribution functions $F_0, F_1 \in \mathcal{F}(b)$ such that $\alpha_{F_0} = \alpha$, $c_{F_0} = c^{-\alpha}$, and*

$$\max_{i\in\{0;1\}} t_{i,n}\alpha_{F_i}^{r/b} c_{F_i}^r \mathbf{E}_i^{1/2}\left(\frac{\widehat{c}_n}{c_{F_i}} - 1\right)^2 \geqq \frac{r}{2b}(\log n)\left(\frac{8r}{ne}\right)^r \tag{19}$$

for all large enough n, where $\max_{i\in\{0;1\}}|t_{i,n} - 1| \to 0$ as $n \to \infty$.

We now present a lower bound of the accuracy of extreme quantiles estimation. We call a quantile "extreme" if the level $q_n \to 0$ as n grows. Of course, there is an infinite variety of possible rates of decay of q_n. Theorem 3 presents lower bounds in the case $q_n \asymp n^{-1/(1+2b)}$. More specifically, we deal with quantile levels $q_n = vn^{-1/(1+2b)}$, where v is bounded away from 0.

We denote by

$$y_i \equiv y_i(q_n) = F_i^{-1}(q_n)$$

the quantile of level q_n. Equivalently, $1/y_i$ is the upper quantile of $\mathcal{L}_i(1/Y)$. In financial applications (see, e.g., [18] and references therein) the level as high as 0.05 can be considered extreme as the empirical quantile estimator of level $q \leqq 0.05$ appears unreliable.

THEOREM 3. *Let $\widehat{y}_n$ be an arbitrary estimator of the level q_n quantile. For any $\alpha > 0$, $c > 0$, and $v < (8\alpha^2 rc^{-2\alpha b})^{r/b}$ there exist distribution functions $F_0, F_1 \in \mathcal{F}(b)$ such that $\alpha_{F_0} = \alpha$, $c_{F_0} = c^{-\alpha}$,*

$$\max_{i\in\{0;1\}} k_{i,n}\alpha_i^{r/b} c_{F_i}^r |\log w_i|^{-1} \mathbf{E}_i^{1/2}\left(\frac{\widehat{y}_n}{y_i} - 1\right)^2 \geqq \frac{1}{2}\left(\frac{8r}{ne}\right)^r \tag{20}$$

for all large enough n, where $w_i = v^{1/\alpha_i}(8r\alpha_i^2 c_{F_i}^{2b})^{-r/(\alpha_i b)}$ and $\max_{i\in\{0;1\}}|k_{i,n} - 1| \to 0$ as $n \to \infty$. Inequality (20) holds if $\widehat{y}_n/y_i - 1$ in the left-hand side is replaced by $y_i/\widehat{y}_n - 1$.

The smaller v is, the lower is q_n, and hence the harder is the estimation problem. Inequality (20) supports this point: $|\log w_i|$ grows as v decreases, lifting the lower bound.

Note that $1/\widehat{y}_n$ is an estimator of the upper quantile of level q_n of $\mathcal{L}(1/Y)$.

3. Proofs. Given a family $\mathcal{P}$ of distributions, Lemma 1 refers to a general problem of estimating a functional a_P of an unknown distribution $P \in \mathcal{P}$ from a sample $X_1, \ldots, X_n$ of i.i.d. r.v.'s.

We assume that the functional a_P is an element of a metric space $(\mathcal{X}, d)$. An estimator $\widehat{a}$ of a_P is a measurable function of $X_1, \ldots, X_n$ taking values in a subspace $\{a_P \colon P \in \mathcal{P}\} \subset \mathcal{X}$.

Given two distributions $\{P_0, P_1\} \in \mathcal{P}$, we put

$$a_i = a_{P_i}, \quad 2\delta = d(a_0; a_1),$$

and $\mathbf{E}_i \equiv \mathbf{E}_{P_i}$ is the mathematical expectation with respect to P_i. Let R denote a loss function, and let $d_{\mathrm{H}}(P_0; P_1)$ denote the Hellinger distance.

LEMMA 1. *If function $R^{1/2}$ is convex, then for any estimator $\widehat{a}$*

$$\max_{i\in\{0;1\}} \mathbf{E}_i R(d(\widehat{a};a_i)) \geqq R(\delta)\left(1-d_{\mathrm{H}}^2\right)^{2n}. \tag{21}$$

In particular, for a quadratic loss function we have

$$\max_{i\in\{0;1\}} \mathbf{E}_i^{1/2} d^2(\widehat{a};a_i) \geqq \delta\left(1-d_{\mathrm{H}}^2\right)^{n}. \tag{22}$$

See [9], [13], [22], and references therein concerning the literature on minimax lower bounds.

Proof of Lemma 1. We may assume that $a_1 \neq a_0$. Denote

$$\mathbf{E}_* = \max_{i\in\{0;1\}} \mathbf{E}_i R(d(\widehat{a};a_i))$$

and recall that

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

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 $\mathcal{L}_i(X_1,\ldots,X_n)$. By the triangle inequality,

$$\delta \leqq \frac{d(a_0;\widehat{a})}{2} + \frac{d(\widehat{a};a_1)}{2}.$$

Since function $R^{1/2}$ is convex, $R^{1/2}(\delta) \leqq R^{1/2}(d(\widehat{a};a_0))/2 + R^{1/2}(d(a_1;\widehat{a}))/2$. Therefore,

$$2R^{1/2}(\delta)(1-d_{\mathrm{H}}^2)^n \leqq \int R^{1/2}(d(\widehat{a};a_0))\sqrt{f_{0,n}}\sqrt{f_{1,n}} + \int R^{1/2}(d(a_1;\widehat{a}))\sqrt{f_{0,n}}\sqrt{f_{1,n}}.$$

This and the Cauchy–Bunyakovskiy inequality entail

$$2R^{1/2}(\delta)(1-d_{\mathrm{H}}^2)^n \leqq \mathbf{E}_0^{1/2} R(d(\widehat{a};a_0)) + \mathbf{E}_1^{1/2} R(d(\widehat{a};a_1)) \leqq 2\,\mathbf{E}_*^{1/2},$$

yielding (21). The proof of Lemma 1 is complete.

Our approach to establishing lower bounds involves two distributions P_0 and P_1, where P_0 is a Pareto distribution and $P_1 \equiv P_{1,n}$ is a "disturbed" version of P_0.

We then apply Lemma 1 that provides a nonasymptotic lower bound of the accuracy of estimation when choosing between the two close alternatives.

Proof of Theorem 1. Given an arbitrary $\alpha > 0$ and $c > 0$, we deal with distribution functions F_0 and F_1, where

$$F_0(y) = \left(\frac{y}{c}\right)^{\alpha} \mathbf{1}\{0 < y \leqq c\},$$
$$F_1(y) = \left(\frac{h}{c}\right)^{-\gamma}\left(\frac{y}{c}\right)^{\alpha_1} \mathbf{1}\{0 < y \leqq h\} + \left(\frac{y}{c}\right)^{\alpha} \mathbf{1}\{h < y \leqq c\},$$

$\alpha_1 > \alpha$, $h \in (0;c)$. It is easy to see that $F_1 \leqq F_0$ and

$$\alpha_{F_0} = \alpha, \quad \alpha_{F_1} = \alpha_1, \quad c_{F_0} = c^{-\alpha}, \quad c_{F_1} = c^{-\alpha}h^{-\gamma}. \tag{23}$$

Denote $\alpha_0 = \alpha$, and let

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

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

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

we have

$$\sup_{0<y\leqq c} |1 - c_{F_1}^{-1} y^{-\alpha_1} F_1(y)| y^{-b\alpha_1} = \sup_{h<y\leqq c} \left(1 - y^{-\gamma} h^{\gamma}\right) y^{-b\alpha_1}. \tag{24}$$

The function on the right-hand side of (24) assumes its maximum at $y_0 = h(1 + \gamma/b\alpha_1)^{1/\gamma}$; it is bounded by $e^{1/e\alpha}/b\alpha$.

It is easy to check that

$$d_{\mathrm{H}}^2(F_0; F_1) = \left(\frac{h}{c}\right)^{\alpha} \left(1 - \frac{\sqrt{1+\gamma/\alpha}}{1+\gamma/(2\alpha)}\right) \sim \frac{\gamma^{1/r}}{8\alpha^2 c^{\alpha}}$$

as $\gamma \to 0$ and

$$d_{\mathrm{H}}^2(F_0; F_1) \leqq \frac{\gamma^{1/r}}{8\alpha^2 c^{\alpha}}. \tag{25}$$

Inequality (25) is typical for nonparametric estimation problems. A nonparametric class is usually so "rich" that one can find distributions $\{P_t, t \geqq 0\}$ such that $d_{\mathrm{H}}(P_0; P_t) \asymp |t|^{\nu}$, with $\nu > 1$, while in a regular parametric case $d_{\mathrm{H}}(P_0; P_t) \asymp |t|$.

According to Lemma 1,

$$\max_{i\in\{0;1\}} \mathbf{E}_i^{1/2} (\widehat{\alpha}_n - \alpha_{F_i})^2 \geqq \frac{\gamma}{2} (1 - d_{\mathrm{H}}^2)^n \geqq \frac{\gamma}{2} \left(1 - \frac{\gamma^{1/r}}{8\alpha^2 c^{\alpha}}\right)^n. \tag{26}$$

Maximizing the right-hand side of this inequality in γ, we obtain

$$\max_{i\in\{0;1\}} \mathbf{E}_i^{1/2} (\widehat{\alpha}_n - \alpha_i)^2 \geqq \frac{1}{2} \left(\frac{8r\alpha^2 c^{\alpha}}{n}\right)^r \left(1 + \frac{r}{n}\right)^{-n-r}.$$

and the optimal choice of γ is given by $\gamma = \gamma_n$, where

$$\gamma_n \equiv \gamma_n(\alpha, b, c) = \frac{(8\alpha^2 c^{\alpha})^r}{(1 + n/r)^r}. \tag{27}$$

Note that $h < c$ as $n \geqq 8\alpha^2 r c^{-2b\alpha}$. It is easy to check that $\alpha/\alpha_1 = 1 - \gamma/\alpha_1$, $\gamma = \gamma_n \leqq (4\alpha^2 c^{\alpha}/n)^r$, and $c^{\alpha r} \leqq c^{\alpha_1 r}$ as $c > 1$. Hence

$$\left(\frac{\alpha}{\alpha_1}\right)^{2r} \geqq 1 - \frac{2r\gamma}{\alpha_1} \geqq 1 - 2r\left(\frac{8r}{n}\right)^r \alpha_1^{2r-1} c^{\alpha r} \geqq 1 - \left(\frac{4}{n}\right)^r \alpha_1^{2r-1} (c^{r\alpha_1} \vee 1). \tag{28}$$

Taking into account (23), we derive (12). Similarly ($a_i := 1/\alpha_i$, $a := a_0$),

$$\max_{i\in\{0;1\}} \mathbf{E}_i^{1/2} (\widehat{a}_n - a_i)^2 \geqq \frac{a a_1}{2} \gamma \left(1 - \frac{\gamma^{1/r} a^2}{8 c^{1/a}}\right)^n = \frac{a_1}{2} a^{1-2r} c^{r/a} \left(\frac{8r}{n}\right)^r \left(1 + \frac{r}{n}\right)^{-n-r},$$

leading to (13). The proof is complete.

Concerning (18), if the distribution function of $1/X$ is given by F_i, $i \in \{0;1\}$, it is natural to choose

$$x_n = \frac{1}{h} = \gamma_n^{-a/b} \sim \left(\frac{n}{8r\alpha^2 c^\alpha}\right)^{r/(\alpha b)}.$$

Then $v(x_n) = 0$ by (17), and (18) follows from (16).

Proof of Theorem 2. Denote $c_i := c_{F_i}$. With F_0 and F_1 defined as above,

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

Using this inequality and (22), we derive

$$\max_{i\in\{0;1\}} \mathbf{E}_i^{1/2}(\widehat{c}_n - c_i)^2 \geqq c^{-\alpha}(2\alpha b)^{-1}\gamma|\log\gamma|\left(1 - \frac{\gamma^{1/r}}{8\alpha^2 c^\alpha}\right)^n.$$

With γ given by (27), we have

$$\max_{i\in\{0;1\}} \mathbf{E}_i^{1/2}(\widehat{c}_n - c_i)^2 \geqq rc^{\alpha(r-1)}(2\alpha^{r/b}b)^{-1}\left(\frac{8r}{ne}\right)^r \log\left(\frac{n}{8r\alpha^2 c^\alpha}\right)$$

for all large enough n, and thus (19) follows.

Proof of Theorem 3. Denote

$$\kappa = v^{1/\alpha}(8\alpha^2 c^\alpha r)^{-r/(\alpha b)}.$$

It is more convenient for us to deal with the equivalent problem of estimating quantiles of the level

$$q_n = \kappa^\alpha \gamma_n^{1/b} \sim vn^{-r/b},$$

where $\gamma = \gamma_n$ is given by (27).

Note that $q_n = (\kappa h)^\alpha$. With functions F_0, F_1 defined as above, it is easy to see that

$$y_0 = c\kappa h = cq_n^{1/\alpha} < h, \quad y_1 = (c\kappa)^{\alpha/\alpha_1}h, \qquad c\kappa < 1.$$

Using the fact that $e^x - 1 \geqq xe^{x/2}$ $(x \geqq 0)$, we derive

$$\begin{aligned} y_1 - y_0 &= cq_n^{1/\alpha_1}\left(\left(\frac{h}{c}\right)^{\gamma/\alpha_1} - q_n^{\gamma/(\alpha\alpha_1)}\right) \\ &= ((c\kappa)^{-\gamma/\alpha_1} - 1)c\kappa h \geqq \gamma^{1+1/(\alpha b)}(c\kappa)^{1-\gamma/(2\alpha_1)}\frac{|\log c\kappa|}{\alpha_1}. \end{aligned}$$

This and (22) entail

$$\max_{i\in\{0;1\}} \mathbf{E}_i^{1/2}(\widehat{y}_n - y_i)^2 \geqq \frac{|\log c\kappa|}{2\alpha_1}c\kappa\gamma^{1+1/(\alpha b)}\left(1 - \frac{\gamma^{1/r}}{8\alpha^2 c^\alpha}\right)^n. \tag{29}$$

With $\gamma = \gamma_n$,

$$\max_{i\in\{0;1\}} \mathbf{E}_i^{1/2}\left(\frac{\widehat{y}_n}{y_i} - 1\right)^2 \geqq \frac{1}{y_i}\,\frac{|\log c\kappa|}{2\alpha_1}\,c\kappa\gamma^{1+1/(\alpha b)}\left(1 - \frac{\gamma^{1/r}}{8\alpha^2 c^\alpha}\right)^n$$

$$= \frac{|\log c\kappa|}{2\alpha_1} (c\kappa)^{1-\alpha/\alpha_i} \gamma_n \left(1 + \frac{r}{n}\right)^{-n}.$$

Using (27) and (28), we arrive at

$$\max_{i\in\{0;1\}} k_{i,n}\alpha_i^{r/b} c_{F_i}^r \mathbf{E}_i^{1/2}\left(\frac{\widehat{y}_n}{y_i} - 1\right)^2 \geqq \frac{1}{2}|\log c\kappa|\left(\frac{8r}{ne}\right)^r \tag{30}$$

for all large enough n. Note that

$$c\kappa = v^{1/\alpha}(8\alpha^2 r c_{F_0}^{2b})^{-r/(\alpha b)}.$$

Hence (20) follows from (30).

Since

$$|1/y_1 - 1/y_0| = |y_1 - y_0|/(y_0 y_1) \geqq \frac{1}{\alpha_i}|\log c\kappa|(c\kappa)^{-1+\gamma/(2\alpha_1)}\gamma^{1-1/\alpha b},$$

we have

$$\max_{i\in\{0;1\}} \mathbf{E}_i^{1/2}\left(\frac{y_i}{\widehat{y}_n} - 1\right)^2 \geqq \frac{|\log c\kappa|}{2\alpha_1}(c\kappa)^{\gamma/(2\alpha_1)}\gamma\left(1 - \frac{\gamma^{1/r}}{8\alpha^2 c^\alpha}\right)^n,$$

yielding the following counterpart to (20): in the assumptions of Theorem 3

$$\max_{i\in\{0;1\}} k_{i,n}\alpha_i^{r/b} c_{F_i}^r \mathbf{E}_i^{1/2}\left(\frac{y_i}{\widehat{y}_n} - 1\right)^2 \geqq \frac{1}{2}|\log c\kappa|\left(\frac{8r}{ne}\right)^r,$$

and hence

$$\max_{i\in\{0;1\}} k_{i,n}\alpha_i^{r/b} c_{F_i}^r |\log w_i|^{-1}\mathbf{E}_i^{1/2}\left(\frac{y_i}{\widehat{y}_n} - 1\right)^2 \geqq \frac{1}{2}\left(\frac{8r}{ne}\right)^r \tag{31}$$

for all large enough n. This is a minimax lower bound of the MSE of an arbitrary estimator $1/\widehat{y}_n$ of the upper q_n-quantile of the distributions $\mathcal{L}_i(1/Y) \in \mathcal{H}(b)$. The proof is complete.

Acknowledgments. The author is grateful to the referees for many helpful remarks.

REFERENCES

[1] J. Beirlant, C. Bouquiaux, and B. J. M. Werker, *Semiparametric lower bounds for tail index estimation*, J. Statist. Plann. Inference, 136 (2006), pp. 705–729.
[2] L. Devroye, *Another proof of a slow convergence result of Birgé*, Statist. Probab. Lett., 23 (1995), pp. 63–67.
[3] D. L. Donoho and R. C. Liu, *Geometrizing rates of convergence.* II, III, Ann. Statist., 19 (1991), pp. 633–701.
[4] H. Drees, *Minimax risk bounds in extreme values theory*, Ann. Statist., 29 (2001), pp. 266–294.
[5] E. F. Fama and R. Roll, *Some properties of symmetric stable distributions*, J. Amer. Statist. Assoc., 63 (1968), pp. 817–836.
[6] R. H. Farrell, *On the best obtainable asymptotic rates of convergence in estimation of a density function at a point*, Ann. Math. Statist., 43 (1972), pp. 170–180.
[7] P. Hall, *On estimating the endpoint of a distribution*, Ann. Statist., 10 (1982), pp. 556–568.
[8] P. Hall and A. H. Welsh, *Best attainable rates of convergence for estimates of parameters of regular variation*, Ann. Statist., 12 (1984), pp. 1079–1084.

[9] C. Huber, *Lower bounds for function estimation*, in Festschrift for L. Le Cam, Springer, New York, 1997, pp. 245–258.
[10] P. Embrechts, C. Klüppelberg, and T. Mikosch, *Modelling Extremal Events for Insurance and Finance*, Springer-Verlag, Berlin, 1997.
[11] C. M. Goldie and R. L. Smith, *Slow variation with remainder: Theory and applications*, Quart. J. Math. Oxford, 38 (1987), pp. 45–71.
[12] I. A. Ibragimov and R. Z. Khasminskii, *Estimation of distribution density*, Zap. Nauch. Sem. LOMI, 98 (1980), pp. 61–85.
[13] I. A. Ibragimov and R. Z. Khasminskii, *Statistical Estimation*, Springer, Berlin, 1981.
[14] A. J. Lenstra *Cramér–Rao revisited*, Bernoulli, 11 (2005), pp. 263–282.
[15] B. B. Mandelbrot, *New methods in statistical economics*, J. Political Economy, 71 (1963), pp. 421–440.
[16] S. Yu. Novak and S. A. Utev, *On the asymptotic distribution of the ratio of sums of random variables*, Siberian Math. J., 31 (1990), pp. 781–788.
[17] S. Yu. Novak, *On the distribution of the ratio of sums of random variables*, Theory Probab. Appl., 41 (1996), pp. 479–503.
[18] S. Yu. Novak, *Advances in extreme value theory with applications to finance*, in New Business and Finance Research Developments, C. Keene, ed., Nova Science, New York, 2009, pp. 199–251.
[19] J. Pfanzagl, *On local uniformity for estimators and confidence limits*, J. Statist. Plann. Inference, 84 (2000), pp. 27–53.
[20] S. I. Resnick, *Heavy tail modeling and teletraffic data*, Ann. Statist., 25 (1997), pp. 1805–1869.
[21] A. E. Shemyakin, *On information inequalities in the parametric estimation*, Theory Probab. Appl., 37 (1992), pp. 89–91.
[22] A. B. Tsybakov, *Introduction to Nonparametric Estimation*, Springer, New York, 2009.