

ON THE ESTIMATION OF EXTREME TAIL PROBABILITIES

BY PETER HALL AND ISHAY WEISSMAN

Australian National University and Technion

Applications of extreme value theory to problems of statistical inference typically involve estimating tail probabilities well beyond the range of the data, without the benefit of a concise mathematical model for the sampling distribution. The available model is generally only an asymptotic one. That is, an approximation to probabilities of extreme deviation is supposed, which is assumed to become increasingly accurate as one moves further from the range of the data, but whose concise accuracy is unknown. Quantification of the level of accuracy is essential for optimal estimation of tail probabilities. In the present paper we suggest a practical device, based on a nonstandard application of the bootstrap, for determining empirically the accuracy of the approximation and thereby constructing appropriate estimators.

1. Introduction. In many applications of extreme value theory there is no precise model for the tail of the distribution from which data are drawn, yet we wish to make inference about the distribution at a point that is well beyond the extent of the data. For example, we may wish to determine the probability that the most extreme wave height observed in the next 100 years exceeds a given value x , where x is considerably in excess of the largest height observed since data were first collected 50 years ago. We usually do not have a concise model for the distribution of wave height, but may be prepared to assume that the chance $\bar{F}(x) = 1 - F(x)$ that wave height in a given observation interval exceeds x , is approximated by an estimable quantity \bar{F}_θ depending on an unknown parameter θ . This model is typically believed to be asymptotically correct for large deviations of the distribution F , in the sense that $\bar{F}(x)/\bar{F}_\theta(x) \rightarrow 1$ as $x \rightarrow \infty$, although it would not be valid in any exact sense. Given an estimator $\hat{\theta}(t)$ of θ , depending on a tuning parameter t , we wish to select t so as to optimize the approximation of $\bar{F}(x)$ by $\bar{F}_{\hat{\theta}(t)}(x)$.

The fact that the true \bar{F} is unknown, even up to unknown parameters, means that the bias of $\bar{F}_{\hat{\theta}(t)}(x)$ is particularly difficult to quantify. This makes it quite awkward to implement the procedure outlined above. In the present paper we suggest a device based on the bootstrap for overcoming this obstacle. Our approach is founded on “pulling back” the problem to a location within the range of the data (assumed to be a sample of n independent and identically distributed random variables), by reducing the values of n and x to m and y , respectively, such that y is an order of magnitude less than the largest observed data value. After the problem has been “pulled back” to the range of

Received August 1994; revised October 1996.

AMS 1991 subject classifications. Primary 60G70, 62G05; secondary 62G09.

Key words and phrases. Bootstrap, extreme value, Hill's estimator, order statistic, Pareto approximation, regular variation, smoothing.

the data, it is feasible to approximate \bar{F} by the empirical distribution function, \hat{F} . The value of the optimal tuning parameter is relatively easy to estimate in this setting, and may be calibrated so that it is appropriate for the original context.

Our approach has points of similarity to accelerated life testing, although in our application the idea is to study circumstances that are less, rather than more, extreme. In accelerated life testing, data under normal dose or stress levels are difficult to obtain, and so experiments are run under higher levels. The results are then recalibrated to lower, “normal” circumstances.

The main idea, of transforming the scale of an extreme-value problem so that it is no longer so extreme, has a range of applications. We discuss some of them here. Our work is based on relatively ad hoc techniques used to solve a multivariate version of the present problem, although the complexity of that case makes it inappropriate for a concise theoretical investigation such as that described in this paper. It concerns the accuracy of navigation systems on a certain wide-bodied passenger and freight airliner, where different forms of information (e.g., from the aircraft’s inertial navigation system) are analyzed “independently” by three on-board computers to determine position on the earth’s surface. Each computer produces an approximation to the aircraft’s latitude and longitude. These three values, or estimates, of the coordinate pair may be used in a number of different ways—in the particular case under investigation, they were averaged using two different methods, one suggested by the aircraft manufacturer and the other devised by the company flying the aircraft. Data consisting of several hundred independent observations of the three coordinate pairs of estimates, and the true coordinates, were provided. The problem was to estimate the probability that a given type of coordinate average was in error by a given number of nautical miles. That specified distance was significantly in excess of the maximum observed discrepancy between true and estimated coordinate pairs.

Other multivariate problems, for example involving statements about the joint distribution of a set of extremes from a univariate distribution, may also be addressed using the approach described in this paper. A variant of the method may be employed for dependent data in cases where an appropriate model is available for the type of dependence (e.g., an autoregression of specified order). Applications to dependent data in less well-specified cases, for example those where the so-called block bootstrap would typically be employed for more standard, nonextreme inference, are more problematical.

Our method may be used to tackle a variety of forecasting problems. In particular, it may be employed to construct a prediction interval for the next record value, or a confidence interval for an extreme probability. The purpose of the present paper is to demonstrate its feasibility in a relatively simple “laboratory” case, where concise theoretical treatment is feasible. In particular, we address point estimation of an extreme quantile or probability.

The idea of using a sample of smaller size than the original has been noted before, by for example Beran and Srivastava (1985) and Athreya (1987) to estimate distributions that do not enjoy normal limits, and by Taylor (1989),

Faraway and Jhun (1990) and Hall (1990) in the context of statistical smoothing. The latter application is closer in spirit to that in the present paper.

We discuss only first-order properties of the estimator, with the aim of showing that first-order optimality is achieved by the bootstrap. In the context studied here, implementation of the bootstrap requires choice of a bivariate tuning parameter, $t = (m_1, m_2)$; see (3.3). The cases addressed by Taylor (1989), Faraway and Jhun (1990) and Hall (1990) also require tuning parameters. We do not discuss optimal selection of such quantities, although this topic could be addressed. It requires second-order asymptotic theory, which is unlikely to provide a good account of properties in real samples unless sample size is particularly large. An account of numerical selection of tuning parameters is beyond the scope of this paper, not least because in practice it is often quite ad hoc, based on trial and error. Typically, the final choice of t is based on simulating from distributions whose tail behavior is suggested by pilot estimators of θ .

Hall and Welsh (1985) described nonbootstrap approaches to extreme-value inference in Pareto-type distributions. Our method is actually less complex than the one suggested by Hall and Welsh, which requires choice of three (rather than our two) tuning parameters [the quantities s , t_1 and t_2 in Hall and Welsh's notation—see Section 5 of Hall and Welsh (1985)]. At the same time, our approach enjoys significantly more general application.

We employ mean squared error as the criterion of optimality, although it will be clear that other approaches could easily be adopted, requiring only minor modification of our procedure. Thus, we wish to select t so that

$$(1.1) \quad D_1(t; n, x) = E\{\bar{F}_{\hat{\theta}(t)}(x) - \bar{F}(x)\}^2$$

is minimized. In the event that we choose to estimate the $(1 - p)$ th quantile of F for a given value of p , that is, the solution x of $\bar{F}(x) = p$, we would select t so as to minimize

$$(1.2) \quad D_2(t; n, p) = D_1\{t, n, \bar{F}^{-1}(p)\} = E[\bar{F}_{\hat{\theta}(t)}\{\bar{F}^{-1}(p)\} - p]^2.$$

Theory describing properties of D_1 and D_2 is developed in Section 2. To estimate those quantities we replace \bar{F} in (1.1) and (1.2) by $\widehat{\bar{F}}$, and $\bar{F}_{\hat{\theta}(t)}$ by its bootstrap version, calculated for resamples of size $m \ll n$ drawn from the original n -sample. Let this approximant be $\bar{F}_{\hat{\theta}^*(t)}$. Then our bootstrap estimators of D_1 and D_2 are, respectively,

$$(1.3) \quad \hat{D}_1(t; m, y) = E'\{\bar{F}_{\hat{\theta}^*(t)}(y) - \widehat{\bar{F}}(y)\}^2,$$

$$(1.4) \quad \hat{D}_2(t; m, q) = E'\left[\bar{F}_{\hat{\theta}^*(t)}\{\widehat{\bar{F}}^{-1}(q)\} - q\right]^2,$$

where E' denotes expectation conditional on the original data. Details about the bootstrap are presented in Section 3. All proofs are given only in outline, and are collected together in Section 4.

At least two critical issues must be settled before methodology for the bootstrap may be developed. First, the manner in which (n, x) is related to (m, y) must be determined. We contend that in cases where the asymptotic model \bar{F}_θ for \bar{F} is of Pareto type, for example, regularly varying of the form $\bar{F}_{(\theta_1, \theta_2)} = \theta_1 x^{-\theta_2}$ where $\theta_1, \theta_2 > 0$, the most important thing is to ensure that the ratio $(\log x)/(\log n)$ is preserved, at least approximately, by the transformation that takes (n, x) to (m, y) . Second, the method of calibrating our estimate of t so that it applies in the back-transformed context, where (m, y) is replaced by (n, x) , must be addressed. This we do in Section 3.

In the present paper we concentrate on the Pareto-type model introduced above, although it will be clear that other possibilities could also be considered. We relabel the model as

$$(1.5) \quad \bar{F}_{\alpha, c} = cx^{-\alpha},$$

and estimate α and c as in Hill (1975). The estimators proposed there are based on the k largest order statistics; k plays the role of t . Specifically, if $X_{(1)} \leq \dots \leq X_{(n)}$ denote the order statistics of a sample $\mathcal{X} = \{X_1, \dots, X_n\}$ then

$$\hat{\alpha} = \hat{\alpha}(k) = \left(k^{-1} \sum_{i=1}^k \log X_{(n-i+1)} - \log X_{(n-k)} \right)^{-1},$$

$$\hat{c} = \hat{c}(k) = (k/n)(X_{(n-k)})^{\hat{\alpha}}.$$

Optimal choice of k requires the tuning parameter to increase with increasing sample size, at a rate depending on the difference $\bar{F} - \bar{F}_{\alpha, c}$ between the true distribution tail and that of the model [Hall (1982), Smith and Weissman (1987)].

Related work on tail estimation using Pareto-type models includes that of Davis and Resnick (1984) and Smith (1987). Anderson (1978, 1984) discussed conditions under which the relative error of estimates in such models converges to zero. Regular variation in very general terms has been described in detail by Bingham, Goldie and Teugels (1987).

By way of notation, for positive sequences a_n and b_n we shall say that " a_n is of size b_n " if the ratio a_n/b_n is bounded away from zero and infinity as n increases. If a_n should be a random variable then by the phrase " a_n is of size b_n in probability" we mean that a_n/b_n and b_n/a_n are both $O_p(1)$.

2. Asymptotic properties of D_1 and D_2 . Since

$$D_2(k; n, p) = D_1\{k; n, \bar{F}^{-1}(p)\}$$

then it suffices to treat D_1 . For that purpose we assume that the true distribution tail \bar{F} is regularly varying with remainder

$$(2.1) \quad \bar{F}(x) = cx^{-\alpha} \{1 + dx^{-\beta} + o(x^{-\beta})\}$$

as $x \rightarrow \infty$, where $\alpha, \beta > 0$, $c > 0$ and $-\infty < d < \infty$. (An example of this kind is the Fréchet distribution, for which $F(x) = \exp(-cx^{-\alpha})$, $\beta = \alpha$ and $d = -c/2$.)

We ask that $x = x(n)$ increase with sample size at such a rate that

$$(2.2) \quad (\log x)/(\log n) \text{ is bounded away from zero and infinity as } n \rightarrow \infty.$$

Our asymptotic, quasi-parametric approximation to \bar{F} is $\bar{F}_{\alpha, c}$, given by (1.5), and represents the first-order term on the right-hand side of (2.1).

We adopt the convention that, in the very unlikely event that $\bar{F}_{\hat{\alpha}, \hat{c}}(x)$ takes a value greater than 1, we redefine $\bar{F}_{\hat{\alpha}, \hat{c}}(x) = 1$. This point is hardly of any practical importance, but it does ensure that the random variable whose expected value is taken in (1.1) is always less than 1 in absolute value, and so there can be no argument over the finiteness of D_1 . We also suppose that the distribution of X is bounded below by a positive constant. This assumption rules out the possibility that $\log X_{(i)}$ is not well defined for some i , and avoids our having to qualify results by tedious caveats. Theorem 2.1 describes properties of D_1 up to an order of accuracy which, in most circumstances, enables us to determine first-order properties of the optimal k ; see the subsequent discussion.

THEOREM 2.1. *Under these assumptions, and (2.1) and (2.2),*

$$(2.3) \quad \begin{aligned} & E\{\bar{F}_{\hat{\alpha}, \hat{c}}(x) - \bar{F}(x)\}^2 \bar{F}_{\alpha, c}(x)^{-2} \\ &= D_1(k; n, x) \bar{F}_{\alpha, c}(x)^{-2} \\ &= E[\delta_1(k/n) - \delta_2(x) + k^{-1/2} Y_1 \\ &\quad + \alpha f(k/n, x) \{k^{-1/2} Y_2 - a(k/n)^{\beta/\alpha}\}]^2 \\ &\quad + o\{[k^{-1} + (k/n)^{2\beta/\alpha}]\{1 + f(k/n, x)^2\} \\ &\quad + \{k^{-1} + (k/n)^{2\beta/\alpha}\}^{1/2} |\delta_2(x)|\} \end{aligned}$$

uniformly in $1 \leq k \leq n^{1-\varepsilon}$ for any $\varepsilon > 0$, where Y_1, Y_2 are independent random variables with zero mean and unit variance; the quantities $\delta_1(k/n)$ and $\delta_2(x)$ depend only on the indicated arguments and satisfy $\delta_1(k/n) = dc^{-\beta/\alpha}(k/n)^{\beta/\alpha} + o\{(k/n)^{\beta/\alpha}\}$ and $\delta_2(x) = dx^{-\beta} + o(x^{-\beta})$ as $n, x \rightarrow \infty$, respectively; $a = dc^{-\beta/\alpha}\beta(\alpha + \beta)^{-1}$; and $f(k/n, x) = \log\{x/\bar{F}_{\alpha, c}^{-1}(k/n)\}$.

The theorem may be employed to derive asymptotic properties of the value k_0 of k that minimizes $D_1(k; n, x)$, and also of the value of $D_1(k_0; n, x)$, as we show in the corollary below. Of course, both quantities can depend significantly on x , even to first order. We shall consider only relatively large x 's, satisfying

$$(2.4) \quad xn^{-\{1/(\alpha+2\beta)\}-\varepsilon} \text{ is bounded away from zero for some } \varepsilon > 0.$$

The expansion at (2.3) is sufficiently detailed to yield accurate approximations to k_0 and $D_1(k_0; n, x)$ in the case of smaller x 's, but there we can find little motivation for going into detail. Bear in mind that under condition (2.1), and for each fixed l , the order statistics $X_{(n)}, \dots, X_{(n-l)}$ are all of size $n^{1/\alpha}$ in probability, and $n^{1/\alpha}$ is of larger order than $n^{\{1/(\alpha+2\beta)\}+\varepsilon}$ for some $\varepsilon > 0$. Therefore, if x should be of smaller order than $n^{\{1/(\alpha+2\beta)\}+\varepsilon}$ for each $\varepsilon > 0$ then x is well within the confines of the sample, and so $\bar{F}(x)$ may be estimated relatively

accurately without even considering the asymptotic approximation $\bar{F}_{\alpha,c}(x)$. Indeed, the empirical distribution function is consistent in this setting, in the strong, relative-error sense implied by $E\{\bar{F}(x) - \widehat{\bar{F}}(x)\}^2 / \bar{F}(x)^2 \rightarrow 0$.

COROLLARY 2.1. *Assume the conditions of Theorem 2.1, with (2.2) strengthened by asking that the ratio $(\log x)/(\log n)$ have a finite, positive limit l as $n \rightarrow \infty$. Suppose too that (2.4) holds and that $d \neq 0$. Then (i) result (2.3) holds uniformly in $|k - k_0| \leq \lambda k_0$ for each $\lambda > 0$; and (ii) $k_0 \sim C_1 n^{2\beta/(\alpha+2\beta)}$ and*

$$D_1(k_0; n, x) \sim C_2 n^{-2\beta/(\alpha+2\beta)} (\log n)^2 x^{-2\alpha},$$

where $C_1 = \{\alpha/(2\alpha^2\beta)\}^{\alpha/(\alpha+2\beta)}$ and

$$(2.5) \quad C_2 = \frac{1}{2} c^2 \alpha^2 \beta^{-1} (\alpha + 2\beta)^{-1} \{l(\alpha + 2\beta) - 1\}^2 (2\alpha^2 \beta / \alpha)^{\alpha/(\alpha+2\beta)}.$$

Note that by (2.4) it is always true that $l > (\alpha + 2\beta)^{-1} + \varepsilon$ for some $\varepsilon > 0$. In particular, the term $l(\alpha + 2\beta) - 1$ appearing in (2.5) is guaranteed to be strictly positive.

Our final result in this section addresses the problem of empirical choice of the optimal value of k . Again we focus on the context of Corollary 2.1, where (2.4) holds, since other cases are not so clearly motivated. We shall prove that an empirical bandwidth selection rule that estimates k_0 with sufficient accuracy achieves the levels of approximation discussed above. That is, it minimizes $D_1(k; n, x)$ up to terms of order $o(\eta)$, where $\eta \equiv n^{-2\beta/(\alpha+2\beta)} (\log n)^2 x^{-2\alpha}$.

Let k_0 denote the value of k that minimizes $D_1(k; n, x)$, and let \hat{k} be a function of the data \mathcal{X} , satisfying

$$(2.6) \quad P(|\hat{k} - k_0|/k_0 > \varepsilon) = o(\eta)$$

for each $\varepsilon > 0$. Define $D_{10}(n, x)$ to be that version of $E\{\bar{F}_{\hat{\alpha}, \hat{c}}(x) - \bar{F}(x)\}^2$ where k (in the construction of $\hat{\alpha}$ and \hat{c}) is replaced by \hat{k} before the expectation is taken. In this context, recall our convention that we replace $\bar{F}_{\hat{\alpha}, \hat{c}}(x)$ by 1 in the unlikely event that $\bar{F}_{\hat{\alpha}, \hat{c}}(x) > 1$.

THEOREM 2.2. *Assume the conditions of Corollary 2.1 and also that (2.6) holds. Then*

$$D_{10}(n, x) - D_1(k_0; n, x) = o(\eta)$$

as $n \rightarrow \infty$.

3. Bootstrap methods and calibration.

3.1. The bootstrap. We assume throughout that the true distribution function F satisfies (2.1), and that the asymptotic, quasi-parametric approximation $\bar{F}_{\alpha,c}$ with which we are working is that defined at (1.5). For the sake of brevity we concentrate on the criterion D_1 , although D_2 may be treated similarly.

First we describe the bootstrap algorithm. Let m , the resample size, be an order of magnitude smaller than n , and denote the resample by $\mathcal{X}^* =$

$\{X_1^*, \dots, X_m^*\}$. It is assumed to be drawn randomly, with replacement, from the full sample $\mathcal{X} = \{X_1, \dots, X_n\}$. Write $X_{(1)}^* \leq \dots \leq X_{(m)}^*$ for the order statistics of the resample, and put

$$\hat{\alpha}^* = \hat{\alpha}^*(k) = \left(k^{-1} \sum_{i=1}^k \log X_{(m-i+1)}^* - \log X_{(m-k)}^* \right)^{-1},$$

$$\hat{c}^* = \hat{c}^*(k) = (k/m)(X_{(m-k)}^*)^{\hat{\alpha}^*}.$$

Define \hat{D}_1 and \hat{D}_2 as at (1.3) and (1.4), with \hat{F} denoting the empirical distribution function of \mathcal{X} , E' representing expectation conditional on \mathcal{X} , k replacing the tuning parameter t , and the vector $(\hat{\alpha}^*(k), \hat{c}^*(k))$ denoting $\hat{\theta}^*(t)$.

Our main result in this subsection, Theorem 3.1, describes estimation of the optimal value of k after the pair (n, x) has been transformed to (m, y) . To simplify our mathematical argument we assume that we search for an empirical optimum only locally, in the vicinity of the true optimum; lengthier arguments can cope with more global optimization problems. Thus, given arbitrary but fixed constants $0 < a_1 < 1 < a_2 < \infty$, and values of m and y , we let $k_0(m, y)$ denote the value of k that minimizes $D_1(k; m, y)$, and write $\hat{k}(m, y)$ for the value of k in the interval $(a_1 k_0, a_2 k_0)$ that minimizes $\hat{D}_1(k; m, y)$. Thus, $\hat{k}(m, y)$ denotes a bootstrap estimator of $k_0(m, y)$. We are particularly interested in the performance of $\hat{k}(m, y)$ as an estimator of $k_0(m, y)$. In a slight abuse of notation we shall in Section 3.2 use the notation $\hat{k}(n, x)$ not for the version of $\hat{k}(m, y)$ evaluated at $(m, y) = (n, x)$, where the definition is anyway not practical, but rather for the calibrated version of $\hat{k}(m, y)$.

Our theorem is valid only when the order of magnitude of y is sufficiently small. To determine an appropriate definition of “sufficiently small,” note that the empirical distribution function tail, $\hat{F}(y)$, is unbiased for $\bar{F}(y)$ and has standard deviation of order $n^{-1/2}y^{-\alpha/2}$ as n and y increase. Bearing in mind that we wish this error not to be of larger order than the remainder $o(y^{-\alpha-\beta})$ in our asymptotic approximation to $\bar{F}(y)$ [see (2.1)], we ask that $n^{-1/2}y^{-\alpha/2}$ be an order of magnitude smaller than $y^{-\alpha-\beta}$, or equivalently that $y = o(n^{1/(\alpha+2\beta)})$. This consideration motivates the assumption

$$(3.1) \quad y = O(n^{\{1/(\alpha+2\beta)\}-\varepsilon})$$

for some $\varepsilon > 0$.

THEOREM 3.1. *Assume (3.1), and the conditions of Corollary 2.1, but imposed on (m, y) rather than (n, x) . Then for each $\varepsilon, \lambda > 0$,*

$$P\{|\hat{k}(m, y) - k_0(m, y)|/k_0(m, y) > \varepsilon\} = O(n^{-\lambda}).$$

The assumption that the conditions of Corollary 2.1 hold for (m, y) demands that $ym^{-\{1/(\alpha+2\beta)\}-\varepsilon_1}$ be bounded away from zero and infinity for some $\varepsilon_1 > 0$ [compare (2.4)], which in company with (3.1) implies that $m = O(n^{1-\varepsilon_2})$ for some $\varepsilon_2 > 0$.

3.2. *Calibration.* We begin by describing a method for calibration. Our aim is to convert the estimator $\hat{k}(m, y)$ into a quantity $\hat{k} = \hat{k}(n, x)$ that satisfies the crucial condition (2.6), and so also satisfies Theorem 2.2.

In order to appreciate the context of our approach, recall from Theorem 2.1 that the nature of the solution to the problem of minimizing $D_1(k; n, x)$ with respect to k depends critically on the limit l of $(\log x)/(\log n)$. Provided that $(\log y)/(\log m) = (\log x)/(\log n)$, l is preserved by the transformation $(n, x) \mapsto (m, y)$, whence by Corollary 2.1 and Theorem 3.1,

$$\log \hat{k}(m, y) = \log C_1 + \gamma \log m + \xi(m, y),$$

where C_1 is exactly as in Corollary 2.1, $\gamma = 2\beta/(\alpha + 2\beta)$ and

$$(3.2) \quad P\{|\xi(m, y)| > \varepsilon\} = O(n^{-\lambda})$$

for all $\varepsilon, \lambda > 0$. Linear regression of $\log \hat{k}(m, y)$ on $\log m$ for just two values of m may be used to produce estimators \hat{C}_1 and $\hat{\gamma}$ of C_1 and γ . We take

$$\hat{k}(n, x) = \hat{C}_1 n^{\hat{\gamma}}$$

as our estimator of the unknown $k_0 = k_0(n, x)$.

For example, if

$$\frac{\log y_1}{\log m_1} = \frac{\log y_2}{\log m_2} = \frac{\log x}{\log n},$$

$m_1 = n^{\xi_1}$ and $m_2 = m_1^{\xi_2}$, where $0 < \xi_1, \xi_2 < 1$ and each $\xi_j = \xi_j(n)$ is bounded away from 0 and 1, then by (3.2),

$$P(|\hat{C}_1 - C_1| > \varepsilon) + P\{|\hat{\gamma} - \gamma| > \varepsilon(\log n)^{-1}\} = O(n^{-\lambda})$$

for all $\varepsilon, \lambda > 0$. It follows that $P(|\hat{k} - k_0|/k_0 > \varepsilon) = O(n^{-\lambda})$ for all $\varepsilon, \lambda > 0$, which is stronger than the crucial condition (2.6). Therefore, \hat{k} satisfies Theorem 2.2.

4. Proofs. We shall derive Theorem 2.1 and Corollary 2.1 in moderate detail. Proofs of the other results will be only sketched, since the techniques employed in all of them bear a filial resemblance to those used to derive Theorem 2.1 and Corollary 2.1.

PROOF OF THEOREM 2.1.

Step (i). Formula for D_1 . In this step we produce a tractable expression for $D_1(k; n, x)$, by Taylor-expanding the functional $\bar{F}_{\hat{\alpha}, \hat{c}}$. Let C_1, C_2, \dots denote positive constants depending only on their arguments, which will be indicated. Put $\xi = \alpha^{-1}$ and $\hat{\xi} = \hat{\alpha}^{-1}$, and observe that if $|\hat{\xi} - \xi| \leq \frac{1}{2}\xi$ then $|\hat{\alpha} - \alpha + \alpha^2(\hat{\xi} - \xi)| \leq C_1(\xi)(\hat{\xi} - \xi)^2$. Therefore, if $|\hat{\xi} - \xi| \leq \min\{\frac{1}{2}\xi, (\log x)^{-1}\}$ then

$$(4.1) \quad \begin{aligned} x^{-\hat{\alpha}} &= x^{-\alpha} \exp\{-(\hat{\alpha} - \alpha) \log x\} \\ &= x^{-\alpha} \{1 + \alpha^2(\hat{\xi} - \xi) \log x + R_1(x)\}, \end{aligned}$$

where $|R_1(x)| \leq C_2(\xi)(\hat{\xi} - \xi)^2(\log x)^2$.

Put $\delta_3 = \{\bar{F}^{-1}(k/n) - \bar{F}_{\alpha,c}^{-1}(k/n)\}/\bar{F}_{\alpha,c}^{-1}(k/n)$ and

$$\Delta = \{X_{(n-k)} - \bar{F}^{-1}(k/n)\}/\bar{F}^{-1}(k/n),$$

and observe that

$$\begin{aligned} \hat{c} &= (k/n)(X_{(n-k)})^{\hat{\alpha}} \\ &= c \exp[\alpha \log(1 + \delta_3) + \alpha \log(1 + \Delta) \\ &\quad + (\hat{\alpha} - \alpha)\{\log \bar{F}_{\alpha,c}^{-1}(k/n) + \log(1 + \delta_3) + \log(1 + \Delta)\}]. \end{aligned}$$

Therefore, if simultaneously

$$(4.2) \quad |\hat{\xi} - \xi| \leq \min\{\frac{1}{2}\xi, (\log x)^{-1}\}, |\delta_3| \leq \frac{1}{2} \quad \text{and} \quad |\Delta| \leq \frac{1}{2}$$

then

$$\hat{c} = c\{1 + \alpha\delta_3 + \alpha\Delta - \alpha^2(\hat{\xi} - \xi) \log \bar{F}_{\alpha,c}^{-1}(k/n) + R_2\},$$

where, for $l = 2$,

$$(4.3) \quad |R_l| \leq C_3(\alpha, c)\{\delta_3^2 + \Delta^2 + (\hat{\xi} - \xi)^2(\log n)^2\}.$$

It follows from this result and (4.1) that if (4.2) holds and $x \geq x_0$ then $\bar{F}_{\hat{\alpha}, \hat{c}} \leq 1$ and

$$\begin{aligned} \bar{F}_{\hat{\alpha}, \hat{c}}(x) &= \hat{c}x^{-\hat{\alpha}} \\ &= \bar{F}_{\alpha,c}(x)[1 + \alpha\delta_1 + \alpha\Delta + \alpha^2(\hat{\xi} - \xi) \log\{x/\bar{F}_{\alpha,c}^{-1}(k/n)\} + R_3], \end{aligned}$$

where R_3 satisfies (4.3). Equivalently, defining $\delta_2 = \{\bar{F}(x) - \bar{F}_{\alpha,c}(x)\}/\bar{F}_{\alpha,c}(x)$ we have

$$\bar{F}_{\hat{\alpha}, \hat{c}}(x) - \bar{F}(x) = \bar{F}_{\alpha,c}(x)[\alpha\delta_1 - \delta_2 + \alpha\Delta + \alpha^2(\hat{\xi} - \xi) \log\{x/\bar{F}_{\alpha,c}^{-1}(k/n)\} + R_3].$$

Hence,

$$(4.4) \quad \begin{aligned} &E\{\bar{F}_{\hat{\alpha}, \hat{c}}(x) - \bar{F}(x)\}^2 \\ &= \bar{F}_{\alpha,c}(x)^2 E[\alpha\delta_3 - \delta_2 + \alpha\Delta + \alpha^2(\hat{\xi} - \xi) \log\{x/\bar{F}_{\alpha,c}^{-1}(k/n)\}]^2 + r_1, \end{aligned}$$

where $r_1 = r_1(x)$ and, provided $|\delta_1| \leq \frac{1}{2}$,

$$(4.5) \quad \begin{aligned} |r_1| &\leq C_4(\alpha, c)\bar{F}_{\alpha,c}(x)^2 \\ &\times \left(|\delta_3|^3 + E|\Delta|^3 + E|\hat{\xi} - \xi|^3\{(\log n)^3 + (\log x)^3\} \right. \\ &\quad \left. + |\delta_2| [|\delta_3|^2 + E(\Delta^2) + E(\hat{\xi} - \xi)^2\{(\log n)^2 + (\log x)^2\}] \right. \\ &\quad \left. + P[|\hat{\xi} - \xi| > \min\{\frac{1}{2}\xi, (\log x)^{-1}\}] + P(|\Delta| > \frac{1}{2}) \right). \end{aligned}$$

Step (ii). Approximation to $\hat{\xi}$ and Δ . In this step we introduce representations for extreme order statistics $X_{(i)}$, and thereby obtain tractable formulas for the quantities $\hat{\xi}$ and Δ in (4.4) and (4.5). First, note that in view of (2.1), $\bar{F}^{-1}(y) = (c/y)^\xi \{1 + \xi dc^{-\xi\beta} y^{\xi\beta} + o(y^{\xi\beta})\}$ as $y \rightarrow 0$, and so

$$(4.6) \quad \log \bar{F}^{-1}(e^{-z}) = \xi(\log c + z) + \xi dc^{-\xi\beta} e^{-\xi\beta z} + r_2(z),$$

where $|r_2(z)| = o(e^{-\xi\beta z})$ as $z \rightarrow \infty$. Next, observe that we may write

$$(4.7) \quad X_{(i)} = \bar{F}^{-1}\{\exp(-S_i)\}, \quad 1 \leq i \leq n,$$

where

$$S_i = \sum_{j=1}^i Z_j(n-j+1)^{-1} = \zeta_i + T_i,$$

the Z_j 's are independent exponential random variables with unit mean,

$$\zeta_{n-i+1} = \sum_{j=i}^n j^{-1} = \log(n/i) + O(n^{-1})$$

uniformly in $1 \leq i \leq n/2$, and $T_i = \sum_{j=1}^i (Z_j - 1)(n-j+1)^{-1}$. Therefore, with $U_i \equiv S_{n-i+1} - \log(n/i)$ we have

$$P\left(\sup_{1 \leq i \leq n/2} |U_i| > n^{\varepsilon-(1/2)}\right) = O(n^{-\lambda})$$

for all $\varepsilon, \lambda > 0$. Hence, $\exp(-\xi\beta S_{n-i+1}) = (i/n)^{\xi\beta} (1 + V_i)$ where

$$P\left(\sup_{1 \leq i \leq n/2} |V_i| > n^{\varepsilon-(1/2)}\right) = O(n^{-\lambda}).$$

Consequently, by (4.6) and (4.7), $\hat{\xi} - \xi = A(k) + B(k) + R_4(k)$, where

$$\begin{aligned} A(k) &= \xi \left\{ k^{-1} \sum_{i=1}^k \sum_{j=1}^{n-i+1} (Z_j - 1)(n-j+1)^{-1} - \sum_{j=1}^{n-k} (Z_j - 1)(n-j+1)^{-1} \right\} \\ &= \xi k^{-1} \sum_{j=n-k+1}^n (Z_j - 1), \end{aligned}$$

$$B(k) = \xi dc^{-\xi\beta} \left\{ k^{-1} \sum_{i=1}^k (i/n)^{\xi\beta} - (k/n)^{\xi\beta} \right\} \sim -\xi dc^{-\xi\beta} \beta(\alpha + \beta)^{-1} (k/n)^{\xi\beta}$$

and, for each $\varepsilon_1, \varepsilon_2, \lambda > 0$, and for $l = 4$,

$$(4.8) \quad P\left[\sup_{1 \leq k \leq n^{1-\varepsilon_1}} |R_l(k)| \{k^{-1/2} + (k/n)^{\xi\beta}\}^{-1} > \varepsilon_2\right] = O(n^{-\lambda}).$$

Since $B(k) = -\xi a(k/n)^{\xi\beta} + o\{(k/n)^{\xi\beta}\}$ where $a = dc^{-\xi\beta} \beta(\alpha + \beta)^{-1}$ then

$$(4.9) \quad \hat{\xi} - \xi = A(k) - \xi a(k/n)^{\xi\beta} + R_5(k),$$

where R_5 satisfies (4.8).

Similarly,

$$\begin{aligned} X_{(n-k)} &= \bar{F}^{-1}\{\exp(-S_{n-k})\} \\ &= c^\xi(n/k)^\xi \{1 + \xi T_{n-k} + \xi d c^{-\xi\beta}(k/n)^{\xi\beta} + R_6(k)\}, \end{aligned}$$

where R_6 satisfies (4.8). Hence, since $\bar{F}^{-1}(k/n) = c^\xi(n/k)^\xi [1 + \xi d c^{-\xi\beta}(k/n)^{\xi\beta} + o\{(k/n)^{\xi\beta}\}]$ then

$$(4.10) \quad \Delta = \xi T_{n-k} + R_7(k),$$

where R_7 satisfies (4.8).

Step (iii). Conclusion. It is straightforward to prove that for all $\varepsilon_1, \varepsilon_2, \lambda > 0$,

$$\sup_{1 \leq k \leq n^{1-\varepsilon_1}} (E|\hat{\xi} - \xi|^\lambda + E|\Delta|^\lambda) = O(n^{\varepsilon_2}).$$

Therefore, if $\mathcal{E} = \mathcal{E}(k)$ denotes an event satisfying $\sup_k P(\mathcal{E}) = O(n^{-\lambda})$ for all $\lambda > 0$, then for all $\varepsilon_1, \varepsilon_2, \lambda_1, \lambda_2 > 0$,

$$\sup_{1 \leq k \leq n^{1-\varepsilon_1}} [E\{|\hat{\xi} - \xi|^{\lambda_1} I(\mathcal{E})\} + E\{|\Delta|^{\lambda_2} I(\mathcal{E})\}] = O(n^{-\lambda_2}),$$

where I denotes the indicator function. Observe that $\delta_3 = \xi d c^{-\xi\beta}(k/n)^{\xi\beta} + o\{(k/n)^{\xi\beta}\}$ and $|f(k/n, x)| = O(\log n)$. We put $\delta_1 = \alpha\delta_3$. Combining these results with (4.4), (4.5) and (4.8)–(4.10), we deduce that for any $\varepsilon_1 > 0$ we have uniformly in $1 \leq k \leq n^{1-\varepsilon_1}$,

$$\begin{aligned} &E\{\bar{F}_{\hat{\alpha}, \hat{c}}(x) - \bar{F}(x)\}^2 \\ (4.11) \quad &= \bar{F}_{\alpha, c}(x)^2 \left\{ E[\delta_1 - \delta_2 + T_{n-k} + \alpha^2\{A(k) - \xi a(k/n)^{\xi\beta}\}f(k/n, x)]^2 \right. \\ &\quad \left. + o[\{k^{-1} + (k/n)^{2\xi\beta}\}\{1 + f(k/n, x)^2\} \right. \\ &\quad \left. + \{k^{-1} + (k/n)^{2\xi\beta}\}^{1/2} |\delta_2| \right\}. \end{aligned}$$

Theorem 2.1 follows from (4.11), on replacing T_{n-k} and $A(k)$ by $k^{-1/2}Y_1$ and $k^{-1/2}\xi Y_2$, respectively.

PROOF OF COROLLARY 2.1.

Step (i). Preliminaries. Expanding the right-hand side of (2.3) we obtain,

$$\begin{aligned} &E\{\bar{F}_{\hat{\alpha}, \hat{c}}(x) - \bar{F}(x)\}^2 \bar{F}_{\alpha, c}(x)^{-2} \\ (4.12) \quad &= \alpha^2 f(k/n, x)^2 \{k^{-1} + a^2(k/n)^{2\beta/\alpha}\} + k^{-1} \\ &\quad + 2\alpha a f(k/n, x)(k/n)^{\beta/\alpha} \{\delta_2(x) - \delta_1(k/n)\} \\ &\quad + \text{terms that are either negligible or do not depend on } k. \end{aligned}$$

Here, “negligible” means of the same order as the “ o ” remainder term in (2.3). Let c_1, c_2, \dots denote positive constants depending only on c, d, α and β . Note that the quantity $k^{-1} + a^2(k/n)^{2\beta/\alpha}$ appearing on the right-hand side of (4.12) is minimized by taking $k = k_1 = c_1 n^{2\beta/(\alpha+2\beta)}$, say. (In fact, $c_1 = C_1$, the latter

defined in Corollary 2.1.) When $k = k_1$, the function $f(k/n, x)$ is rendered equal to zero by taking $x = c_2 x_1$, say, where $x_1 = n^{1/(\alpha+2\beta)}$. In this notation, the condition that $xn^{-\{1/(\alpha+2\beta)\}-\varepsilon}$ be bounded away from zero is equivalent to asking that $x/(x_1 n^\varepsilon)$ be so bounded.

To explain why the term $\delta_1(k/n)^2$ is not included in (4.12), we assume without essential loss of generality that $k \delta_1(k/n)^2$ has a limit, L say, either finite or infinite. If $L = 0$ then $\delta_1(k/n)^2$ is negligible relative to an earlier term, k^{-1} , in (4.12), and so may be dropped. If $L = \infty$ then k is of strictly larger order than $n^{2\beta/(\alpha+2\beta)}$, in which case it follows from step (iii) below that

$$(4.13) \quad p(k) \equiv \alpha^2 f(k/n, x)^2 \{k^{-1} + a^2(k/n)^{2\beta/\alpha}\} + k^{-1}$$

is of size at least $(k/n)^{2\beta/\alpha}(\log n)^2$, which in turn is of strictly larger size than $\delta_1(k/n)^2$. Therefore, again $\delta_1(k/n)^2$ may be dropped from (4.12). Finally, if $0 < L < \infty$ then, also by step (iii), $p(k)$ is of size $n^{-2\beta/(\alpha+2\beta)}(\log n)^2$, which is of strictly larger size than $\delta_1(k/n)^2$, implying once more that the latter may be dropped from (4.12).

In the steps that follow, particularly Step (v), the remark that k is “in the vicinity of k_0 ” means that $|k - k_0| < \lambda k_0$ for some $\lambda > 0$, sufficiently large.

Step (ii). Bound for early terms in (4.12). Here we derive lower and upper bounds to $p(k)$. Now, k^{-1} is of smaller order than $r = n^{-2\beta/(\alpha+2\beta)}(\log n)^2$ if and only if $(n/k)^{1/\alpha}$ is of smaller order than $x_1(\log n)^{2/\alpha}$. Furthermore, in the event that $x/\{(n/k)^{1/\alpha} n^\varepsilon\}$ is bounded away from zero for some $\varepsilon > 0$, the ratio $|f(k/n, x)|/\log n$ is also so bounded. Additionally, if

$$x/(x_1 n^{2\varepsilon}) = \{(n/k)^{1/\alpha}/(x_1 n^\varepsilon)\} [x/\{(n/k)^{1/\alpha} n^\varepsilon\}]$$

is bounded away from zero for some $\varepsilon > 0$ then so too is

$$\{(n/k)^{1/\alpha}/(x_1 n^\varepsilon)\} + [x/\{(n/k)^{1/\alpha} n^\varepsilon\}],$$

and hence, in view of the points made in the previous two sentences, $k^{-1}r^{-1} + |f(k/n, x)|(\log n)^{-1}$ is also bounded away from zero. When $k^{-1}r^{-1} \geq c_3$ we clearly have $p(k) \geq c_3 r$; and when $|f(k/n, x)|(\log n)^{-1} \geq c_3$ then, since the minimum value of $k^{-1} + a^2(k/n)^{2\beta/\alpha}$ is $c_4 k_1^{-1}$ and is attained at $k = k_1$, then $f(k/n, x)^2 \{k^{-1} + a^2(k/n)^{2\beta/\alpha}\}$ is bounded below by a constant multiple of r . Hence, when $x/(x_1 n^\varepsilon)$ is bounded away from zero the minimum over k of the quantity at (4.13) is bounded below by a constant multiple of r . Taking $k = k_1$ and noting that $(\log x)/(\log n)$ is bounded, we see that the minimum is also bounded above by a constant multiple of r . Therefore,

$$(4.14) \quad \inf_k p(k) \asymp r.$$

Step (iii). Asymptotic formula for $\operatorname{argmin} p(k)$. Here we show that the value k'_0 of k which produces a minimum of p satisfies $k'_0 \sim k_1$. In view of (4.14), and since $p(k) \geq k^{-1}$, the smallest k at which p can be minimized is $c_5 r^{-1} = c_5 c_1^{-1} k_1 (\log n)^{-2}$, for some $c_5 > 0$. Should k be of larger order than k_1 then, since $x/\{(n/k)^{1/\alpha} n^\varepsilon\}$ would be bounded away from zero for some $\varepsilon > 0$, $|f(k/n, x)|/\log n$ would also be bounded away from zero. Hence, $p(k)$

would be of order at least equal to that of $(\log n)^2(k/n)^{2\beta/\alpha}$, which in turn would be of larger order than r . Therefore, in view of (4.14), such a value of k cannot minimize p . It follows that the value of k'_0 must lie in the interval $\mathcal{J} = [c_5r^{-1}, c_6k_1]$ for some $c_6 > 0$. With $g(n, x) = \log x - \alpha^{-1} \log(n/k_1)$ we have $f(k/n, x) = g(n, x) + o(\log n)$ uniformly in $k \in \mathcal{J}$, and also, since $x/(x_1n^\varepsilon)$ is bounded away from zero for some $\varepsilon > 0$, $|g(n, x)|/\log n$ is bounded away from zero. The latter ratio is clearly bounded away from infinity. Therefore,

$$(4.15) \quad \begin{aligned} p(k) &= \{1 + o(1)\} \alpha^2 g(n, x)^2 \{k^{-1} + a^2(k/n)^{2\beta/\alpha}\} + k^{-1} \\ &= \{1 + o(1)\} \alpha^2 g(n, x)^2 \{k^{-1} + a^2(k/n)^{2\beta/\alpha}\} \end{aligned}$$

uniformly in $k \in \mathcal{J}$. The value of k that minimizes the last line is of course asymptotic to k_1 , and therefore $k'_0 \sim k_1$.

Define $p_1(k) = 2\alpha\alpha f(k/n, x)(k/n)^{\beta/\alpha} \delta_1(k/n)$ and

$$p_2(k) = 2\alpha\alpha f(k/n, x)(k/n)^{\beta/\alpha} \delta_2(x).$$

Step (iv). Negligibility of $p_1(k)$. Here we show that

$$(4.16) \quad \begin{aligned} \inf_k \{p(k) - p_1(k)\} &\sim \inf_k p(k), \\ \operatorname{argmin}\{p(k) - p_1(k)\} &\sim \operatorname{argmin} p(k). \end{aligned}$$

Note that $p_1(k) = O\{(\log n)(k/n)^{2\beta/\alpha}\}$. From the deliberations leading to (4.14) we know that $k^{-1}r^{-1} + |f(k/n, x)|(\log n)^{-1}$ is bounded away from zero, and so

$$\begin{aligned} k \leq c_7r^{-1} &= O\{n^{2\beta/(\alpha+2\beta)} (\log n)^{-2}\} \\ &= o\{n^{2\beta/(\alpha+2\beta)} (\log n)^{-\alpha/(\alpha+2\beta)}\}. \end{aligned}$$

It follows that $k p_1(k) \rightarrow 0$, which, since $p(k) \geq k^{-1}$, implies that $p_1(k)$ is of smaller order than $p(k)$. That implies (4.16).

Step (v). Conclusion. We know from steps (iii) and (iv), particularly (4.15) and (4.16), that

$$\inf_k \{p(k) - p_1(k)\} \sim \alpha^2 g(n, x)^2 \{k_1^{-1} + a^2(k_1/n)^{2\beta/\alpha}\},$$

and that the value k'_0 of k which minimizes $p(k) - p_1(k)$ satisfies $k'_0 \sim k_1$. If

$$(\log n)(k_1/n)^{\beta/\alpha} x^{-\beta} = o\{(\log n)^2(k_1/n)^{2\beta/\alpha}\},$$

or equivalently,

$$(4.17) \quad x^{-1} = o\{n^{-1/(2\beta+\alpha)} (\log n)^{1/\beta}\},$$

then $p_2(k_1) = o(r)$ and so

$$(4.18) \quad \inf_k \{p(k) - p_1(k) + p_2(k)\} \sim \alpha^2 g(n, x)^2 \{k_1^{-1} + a^2(k_1/n)^{2\beta/\alpha}\},$$

and the value k_0''' of k that minimizes $p(k) - p_1(k) + p_2(k)$ satisfies $k_0''' \sim k_1$. Condition (4.17) is implied by (2.4). Formula (2.5) follows from (4.12) and (4.18).

PROOF OF THEOREM 2.2. In view of (2.6), we may choose a sequence ε_n of positive constants that decrease to zero so slowly that

$$P(|\hat{k} - k_0|/k_0 > \varepsilon_n) = o(\eta).$$

Put \tilde{k} equal to \hat{k} if $|\hat{k} - k_0|/k_0 \leq \varepsilon_n$, and otherwise equal to the smallest integer not less than $k_0(1 + \varepsilon_n)$. Let $D_{11}(n, x)$ denote the version of $D_{10}(n, x)$ in which \tilde{k} replaces \hat{k} . Since $D_{10}(n, x)$ and $D_{11}(n, x)$ both represent expected values of quantities whose absolute values are bounded by 1 then, by (2.6), $D_{10}(n, x) - D_{11}(n, x) = o(\eta)$. Therefore it suffices to prove the theorem for $D_{11}(n, x)$ instead of $D_{10}(n, x)$:

$$(4.19) \quad D_{11}(n, x) - D_1(k_0; n, x) = o(\eta).$$

As a prelude to proving (4.19) we note that the argument given during the proof of Corollary 2.1, starting from (4.12), shows that

$$(4.20) \quad D_1(k_0; n, x) = \bar{F}_{\alpha, c}(x)^2 [\alpha^2 g(x, n)^2 \{k_0^{-1} + a^2(k_0/n)^{2\beta/\alpha}\} + k_0^{-1} + 2a\alpha g(x, n)(k_0/n)^{\beta/\alpha} \delta_2(x) + \delta_2(x)^2] + o(\eta).$$

To derive (4.19) we rework the arguments in the proofs of Theorem 2.1 and Corollary 2.1, but obtaining the various remainder terms uniformly in $|k - k_0| \leq \varepsilon_n k_0$ rather than for a single k . This is quite straightforward, since quantities such as $A(k)$ and T_{n-k} admit invariance principles where the remainders, calculated in terms of the difference away from their counterparts with $k = k_0$, are appropriately small. In particular, with $W(k)$ denoting either $A(k)$ or T_{n-k} we have

$$E \left(\max_{k: |k-k_0| \leq \varepsilon_n k_0} |W(k) - W(k_0)|^\lambda \right) = o(k_0^{-\lambda/2})$$

for all $\lambda > 0$. Arguing thus we may derive the version of (4.20) in which the left-hand side is replaced by either

$$E \left[\max_{k: |k-k_0| \leq \varepsilon_n k_0} \{ \bar{F}_{\hat{\alpha}(k), \hat{c}(k)}(x) - \bar{F}(x) \}^2 \right]$$

or

$$E \left[\min_{k: |k-k_0| \leq \varepsilon_n k_0} \{ \bar{F}_{\hat{\alpha}(k), \hat{c}(k)}(x) - \bar{F}(x) \}^2 \right],$$

and the right-hand side is unchanged. It follows that (4.20) is valid if the left-hand side is changed to $D_{11}(n, x)$. Let the resulting formula be (4.20'). The desired result (4.19) follows on subtracting (4.20) from (4.20').

PROOF OF THEOREM 3.1. Observe that for all $\varepsilon \in (0, \frac{1}{2})$ and $\lambda > 0$,

$$P\left\{\sup_{x>0} |\widehat{F}(x) - \bar{F}(x)| \bar{F}(x)^{\varepsilon-(1/2)} > n^{\varepsilon-(1/2)}\right\} = O(n^{-\lambda}).$$

With this result in hand one may rework the proof of Theorem 2.1 in the context where $(\hat{\alpha}^*, \hat{c}^*, m)$ replaces $(\hat{\alpha}, \hat{c}, n)$ and probabilities and expectations are interpreted conditionally. Arguing in that manner, it may be proved that under the conditions of Theorem 3.1, and uniformly in $k \in (a_1 k_0, a_2 k_0)$ for any $0 < a_1 < 1 < a_2 < \infty$ [where $k_0 = k_0(m, y)$], we have the following expansion:

$$E'\{\bar{F}_{\hat{\alpha}^*, \hat{c}^*}(y) - \widehat{F}(y)\}^2 \bar{F}_{\alpha, c}(y)^{-2} = \delta_2(y)^2 + \alpha^2 g_1(m, y)^2 \{k^{-1} + \alpha^2 (k/m)^{2\beta/\alpha}\} \\ + b g_1(m, y) (k/m)^{\beta/\alpha} y^{-\beta/\alpha} + R(k, m, y),$$

where $g_1(y, m) = \log y - \alpha^{-1} \log\{m/k_0(m, y)\}$ and, for all $\varepsilon, \lambda > 0$,

$$(4.21) \quad P\left\{\max_{k \in (a_1 k_0, a_2 k_0)} |R(k, m, y)| > \varepsilon k_0(m, y)^{-1} (\log n)^2\right\} = O(n^{-\lambda}).$$

The argument in the proof of Corollary 2.1 may be employed to derive Theorem 3.1 from (4.21).

Acknowledgments. We are particularly grateful for a very helpful report from an Associate Editor, who pointed out that in the first version of this paper we had defined the function δ_2 incorrectly. The second author is grateful to the Centre for Mathematics and its Applications, Australian National University, for its hospitality during the summer of 1994 and to the Israel-US Binational Fund for partial support.

REFERENCES

- ANDERSON, C. W. (1978). Super-slowly varying functions in extreme value theory. *J. Roy. Statist. Soc. Ser. B* **40** 197–202.
- ANDERSON, C. W. (1984). Large deviations of extremes. In *Statistical Extremes and Applications* (J. Tiago de Oliveira, ed.) 325–340. Reidel, Dordrecht.
- ATHREYA, K. B. (1987). Bootstrap of the mean in the infinite variance case. *Ann. Statist.* **15** 724–731.
- BERAN, R. and SRIVASTAVA, M. S. (1985). Bootstrap tests and confidence regions for functions of a covariance matrix. *Ann. Statist.* **13** 95–115.
- BINGHAM, N. H., GOLDIE, C. M. and TEUGELS, J. L. (1987). *Regular Variation*. Cambridge Univ. Press.
- DAVIS, R. A. and RESNICK, S. I. (1984). Tail estimates motivated by extreme value theory. *Ann. Statist.* **12** 1467–1487.
- FARAWAY, J. J. and JHUN, M. (1990). Bootstrap choice of bandwidth for density estimation. *J. Amer. Statist. Assoc.* **85** 1119–1122.
- HALL, P. (1982). On some simple estimates of an exponent of regular variation. *J. Roy. Statist. Soc. Ser. B* **44** 37–42.
- HALL, P. (1990). Using the bootstrap to estimate mean squared error and select smoothing parameter in nonparametric problems. *J. Multivariate Anal.* **32** 177–203.
- HALL, P. and WELSH, A. H. (1985). Adaptive estimators of parameters of regular variation. *Ann. Statist.* **13** 331–341.

- HILL, B. M. (1975). A simple general approach to inference about the tail of a distribution. *Ann. Statist.* **3** 1163–1174.
- SMITH, R. L. (1987). Estimating the tails of probability distributions. *Ann. Statist.* **15** 1174–1207.
- SMITH, R. L. and WEISSMAN, I. (1987). Large deviations of tail estimators based on the Pareto approximation. *J. Appl. Probab.* **24** 619–630.
- TAYLOR, C. C. (1989). Bootstrap choice of the smoothing parameter in kernel density estimation. *Biometrika* **76** 705–712.

CENTRE FOR MATHEMATICS
AND ITS APPLICATIONS
AUSTRALIAN NATIONAL UNIVERSITY
CANBERRA ACT 0200
AUSTRALIA

TECHNION-ISRAEL INSTITUTE
OF TECHNOLOGY
HAIFA 32000
ISRAEL
E-MAIL: ieriw01@ie.technion.ac.il