The Annals of Statistics
1987, Vol. 15, No. 3, 1174–1207

ESTIMATING TAILS OF PROBABILITY DISTRIBUTIONS

BY RICHARD L. SMITH

University of Surrey

We study the asymptotic properties of estimators of the tail of a distribution based on the excesses over a threshold. A key idea is the use of Pickands' generalised Pareto distribution and its fitting, in most cases, by the method of maximum likelihood. The results cover all three limiting types of extreme value theory. We propose a new estimator for an index of regular variation and show that it often performs better than Hill's estimator. We give new results for estimating the endpoint of a distribution, extending earlier work by Hall and by Smith and Weissman. Finally, we give detailed results for the domain of attraction of $\exp(-e^{-x})$ and show that, in most cases, our proposed estimator is more efficient than two others, one based on the exponential distribution and the other due to Davis and Resnick. We also touch briefly on the problem of large deviations from a statistical point of view. The results make extensive use of existing work on rates of convergence.

1. Introduction. Suppose we have a random sample from an unknown distribution function F, and we are interested in the upper tail of F, i.e., in $F(x)$ as $F(x) \to 1$. It seems reasonable that statistical procedures should be based only on the extreme order statistics in the sample, and numerous such procedures have been proposed. Two problems, in particular, have been studied in detail.

PROBLEM 1. *Estimation of an index of regular variation*. Suppose $\{1 - F(tx)\}/\{1 - F(x)\} \to t^{-\alpha}$ as $x \to \infty$ for each $t > 0$, and we wish to estimate α. This problem arises naturally in dealing with long-tailed (Pareto-type) distributions; applications include insurance claim distributions [Teugels (1984)] and a number of problems in mathematical economics and the social sciences [Mandelbrot (1982), Chapter 38]. A simple estimator of α was proposed by Hill (1975) and from a different perspective by Weissman (1978) and its properties have been much studied [Hall (1982a), Davis and Resnick (1984), Haeusler and Teugels (1985), Csörgő and Mason (1985) and Goldie and Smith (1987)]. For the case $1 - F(x) = Cx^{-\alpha}\{1 + O(x^{-\beta})\}$ as $x \to \infty$, Hall and Welsh (1984) showed that no estimator of α converges at a faster rate than $n^{-\beta/(\alpha+2\beta)}$ and Hall and Welsh (1985) proposed an estimator that achieves this rate of convergence. Other estimators have been proposed, e.g., de Haan and Resnick (1980) and Teugels (1981) proposed simple estimators based on order statistics and more recently Csörgő, Deheuvels and Mason (1985) have studied a very general estimator which includes Hill's estimator as a special case. Our approach will lead to yet another estimator of α. DuMouchel (1983) discussed similar procedures in the

Received October 1984; revised January 1987.
AMS 1980 *subject classifications*. Primary 62G05, 62F10; secondary 62G30, 62F12.
Key words and phrases. Extreme value theory, generalised Pareto distribution, maximum likelihood.

context of estimating a stable law. He argued that procedures of the above type, based on the extreme order statistics, may well be superior to fitting a stable law to the whole sample.

PROBLEM 2. *Estimating an endpoint.* The opposite problem to Problem 1 occurs with short-tailed distributions and is to estimate the endpoint of the distribution,

$$x_0 = \sup\{x\colon F(x) < 1\},$$

on the assumption that this is finite. In most applications, the problem is formulated in terms of the lower endpoint rather than the upper endpoint. Some general estimators have been proposed by Robson and Whitlock (1964), Weiss (1971), Cooke (1979, 1980), de Haan (1981) and Ariyawansa and Templeton (1983, 1986), amongst others. The last three papers were concerned especially with an application to nonlinear optimisation, also developed by Patel and Smith (1983). We shall consider an estimator equivalent to those of Hall (1982b) and Smith and Weissman (1985).

These two problems may be considered special cases of the general problem of estimating the tail of a distribution. In this paper we propose a solution to the general problem, which covers all three limit laws of classical extreme value theory. A distribution function F is said to be in the *domain of* (maximum) *attraction* of another distribution function H if $F^n(a_n x + b_n) \to H(x)$ for some $a_n > 0$, b_n. It is known [see, e.g., Galambos (1978) or Leadbetter, Lindgren and Rootzén (1983)] that H must be the same, up to location and scale, as one of the three cases $\Phi_\alpha(x) = \exp(-x^{-\alpha})$, $x > 0$, $\alpha > 0$, $\Psi_\alpha(x) = \exp(-(-x)^\alpha)$, $x < 0$, $\alpha > 0$, or $\Lambda(x) = \exp(-e^{-x})$, $-\infty < x < \infty$, and the domain of attraction of each of these is completely determined. Our approach starts with a result of Pickands (1975). Let $x_0 \le \infty$ denote the upper endpoint of F and define

$$F_u(y) = \frac{F(u+y) - F(u)}{1 - F(u)}, \qquad u < x_0, 0 < y < x_0 - u,$$

the conditional distribution function of $X - u$ given $X > u$. Define the *generalised Pareto distribution* (GPD) by

$$G(y;\sigma,k) = \begin{cases} 1 - (1 - ky/\sigma)^{1/k}, & k \neq 0, \sigma > 0, \\ 1 - \exp(-y/\sigma), & k = 0, \sigma > 0, \end{cases}$$

the range of y being $0 < y < \infty$ $(k \le 0)$ or $0 < y < \sigma/k$ $(k > 0)$. Pickands showed that the GPD is a good approximation of F_u, in the sense that

$$\lim_{u \to x_0} \sup_{0 < y < x_0 - u} |F_u(y) - G(y;\sigma(u),k)| = 0, \tag{1.1}$$

for some fixed k and function $\sigma(u)$, if and only if F is in the domain of attraction of one of the three limit laws. The constant k is $-\alpha^{-1}$ if $H = \Phi_\alpha$, $+\alpha^{-1}$ if $H = \Psi_\alpha$ and 0 if $H = \Lambda$.

Pickands used this result to derive an estimator of the upper tail of F which is consistent for any of the three domains of attraction. His estimates for k and

$\sigma(u)$, however, are based on simple functions of the order statistics and are not asymptotically efficient.

We now state our approach. Suppose we have a large number of independent observations $X_1, \dots, X_n$ with a common distribution function F. Fix a high threshold u, let N denote the number of exceedances of u and let $Y_1, \dots, Y_N$ denote the excesses. That is, $Y_i = X_j - u$, where j is the index of the ith exceedance. Conditionally on N, the excesses are i.i.d. with distribution function F_u. Our proposal is to approximate F_u by the GPD, $G(\cdot; \sigma, k)$ say, estimating σ and k by the maximum-likelihood estimates, $\hat{\sigma}_N$ and $\hat{k}_N$, based on $Y_1, \dots, Y_N$. Combined with N/n as an estimator of $1 - F(u)$, this gives the tail estimator

$$1 - \hat{F}(u + y) = n^{-1}N\left(1 - \hat{k}_N y/\hat{\sigma}_N\right)^{1/\hat{k}_N}, \tag{1.2}$$

for $0 < y < \infty$ (if $\hat{k}_N < 0$) or $0 < y < \hat{\sigma}_N/\hat{k}_N$ (if $\hat{k}_N > 0$).

This general approach allows us to derive a number of other estimators as special cases. If $\hat{k}_N > 0$, then $u + \hat{\sigma}_N/\hat{k}_N$ is an estimator of the endpoint of the distribution. This is effectively the same estimator as that of Hall (1982b) and Smith and Weissman (1985), but the present derivation seems more intuitive. If $\hat{k}_N < 0$, then $\hat{\alpha}_N = -1/\hat{k}_N$ is an estimator of the index of regular variation which is different from Hill's (1975) estimator. Hill assumed

$$\{1 - F(x)\}/\{1 - F(u)\} = (x/u)^{-\alpha}, \qquad x > u, \tag{1.3}$$

and derived from this the maximum-likelihood estimator of α, which in our notation is

$$\bar{\alpha}_N = N\left\{\sum_{i=1}^{N} \log(1 + Y_i/u)\right\}^{-1} \tag{1.4}$$

If (1.3) is satisfied exactly, then Hill's estimator is asymptotically efficient, but if (1.3) is only an asymptotic relation, then it may be that $\hat{\alpha}_N$ is asymptotically more efficient than $\bar{\alpha}_N$; see Section 4. Csörgő, Deheuvels and Mason (1985) have generalised Hill's estimator in a quite different way, using *weighted* linear combinations of the ordered values of $\log(1 + Y_i/u)$.

The use of $\hat{\alpha}_N$ as an estimator of α is theoretically justified (via Pickands' result) by the fact that F is in the domain of attraction of Φ_α if and only if $\{1 - F(tx)\}/\{1 - F(x)\} \to t^{-\alpha}$ as $x \to \infty$ for $t > 0$. The third limit law, Λ, has been less extensively studied, though the idea of using the exponential distribution as an approximation to F_u has been developed by hydrologists under the name *peaks over threshold* (POT) method. Smith (1984) has included a discussion and references. Davis and Resnick (1984) took a different approach, showing that tail estimates based on Hill's estimator are consistent for many F in the domain of attraction of Λ. We shall compare both the POT approach and the Davis–Resnick method with ours based on the GPD.

The main theoretical question, other than the choice among different estimators, is the choice of threshold u. If u is too large, then N will be small and hence the estimators will have high variance. If u is too small, then the approximation of F_u by the GPD will be poor and the estimates correspondingly biased. Our results lead to a quantification of this bias versus variance trade-off and hence to asymptotically optimal choices of u.

A different, but general, approach to tail estimation has been taken by Breiman, Stone and Gins (1979, 1981). Recently Laycock and Phang (1986) have studied closely related methods for the prediction of extreme values, though their methods are based on classical extreme value theory rather than the threshold approach. For a more direct approach based on classical extreme value theory, see Cohen (1984, 1986). Joe (1987) has also considered tail estimation using classical extreme value theory as well as the GPD.

The plan of the paper is as follows. The basic idea underlying all our results is presented in Section 2. The detailed working out of this method, however, requires separate consideration of the different domains of attraction. The case of Φ_α is studied in Section 3 and a comparison between the GPD and Hill's estimators is made in Section 4. Then we turn to Ψ_α, which is further subdivided into cases $\alpha > 2$, $\alpha = 2$, and $\alpha < 2$. The case $\alpha > 2$ (Section 5) is similar in character to the previous results, but the cases $\alpha = 2$ and $\alpha < 2$ (Sections 6 and 7) are quite different in character and we treat these in a different way from the rest of the paper. Returning to our main theme, we consider the consequences of the preceding results for the estimation of tail probabilities (Section 8) and finally turn to the domain of attraction of Λ (Section 9). Some conclusions and possibilities for further work are given in Section 10.

In presenting limit theorems, the role of N (the number of exceedances) is primary but that of n (the size of the original sample) is secondary. For these reasons, most of our results are conditional on both N and u and are presented as limit theorems as $N \to \infty$ and $u \equiv u_N \to x_0$, simultaneously. Such results may also be readily interpreted as unconditional results in which either N or u is treated as fixed and the other random, depending on n.

As a final preliminary remark, although all our results are for the case of i.i.d. observations, the threshold approach is not restricted to this case and, indeed, has many possibilities for dealing with dependent and/or nonidentically distributed observations. Some of the possibilities have been discussed by Davison (1984) and Smith (1984).

2. Estimation using the GPD. Let $g(y; \sigma, k) = (\partial/\partial y)G(y; \sigma, k)$ denote the GPD density. For reference, the first- and second-order derivatives of $\log g(\cdot\,; \sigma, k)$ and their expectations are derived in the Appendix.

First, consider the case where $Y_1, \dots, Y_N$ are i.i.d. with exact GPD density. Let $L_n(\sigma, k) = \sum_{i=1}^N \log g(Y_i; \sigma, k)$ denote the log likelihood and define $\mathbf{U}_N(\sigma, k)$ as the 2-vector with components $-\sigma\,\partial L_N/\partial\sigma$, $-\partial L_N/\partial k$ and $I_N(\sigma, k)$ as the 2×2 matrix with diagonal elements $-\sigma^2\partial^2 L_N/\partial\sigma^2$, $-\partial^2 L_N/\partial k^2$ and off-diagonal elements $-\sigma\,\partial^2 L_N/\partial\sigma\,\partial k$. Except for the factor, σ, these are the negative score statistic and the observed information matrix for (σ, k). Provided $k < \frac{1}{2}$, we have $E\,\mathbf{U}_N = \mathbf{0}$,

$$N^{-1}E\,\mathbf{U}_N\mathbf{U}_N^T = N^{-1}EI_N = M,$$

where

$$M = \frac{1}{(1-2k)(1-k)}\begin{bmatrix} 1-k & -1 \\ -1 & 2 \end{bmatrix}, \qquad M^{-1} = (1-k)\begin{bmatrix} 2 & 1 \\ 1 & 1-k \end{bmatrix}.$$

Standard arguments [e.g., Cox and Hinkley (1974), Chapter 9] suggest that there exists, with probability tending to 1 as $N \to \infty$, a local maximum $(\hat{\sigma}_N, \hat{k}_N)$ of L_N, satisfying

$$\begin{bmatrix} \hat{\sigma}_N/\sigma - 1 \\ \hat{k}_N - k \end{bmatrix} = -I_N^{-1}\mathbf{U}_N(1 + o(1)) \tag{2.1}$$

and from this it follows that

$$N^{1/2}\begin{bmatrix} \hat{\sigma}_N/\sigma - 1 \\ \hat{k}_N - k \end{bmatrix} \to_d \mathscr{N}\left[\begin{bmatrix} 0 \\ 0 \end{bmatrix}, M^{-1}\right], \tag{2.2}$$

provided $k < \frac{1}{2}$. Here and everywhere, subsequently, $\mathscr{N}$ denotes the (univariate or bivariate) normal distribution and $\to_d$ denotes convergence in distribution. We shall also use $\to_p$ to denote convergence in probability. For the case $k > 0$ Cramér's regularity conditions are violated, but a rigorous proof of these statements, and extensions to $k \geq \frac{1}{2}$, are given by Smith (1985).

Now suppose that $Y_1, \dots, Y_N$ are drawn not from the GPD, but from one of the F_u's. If (1.1) holds, we will typically be able to determine a remainder function ϕ such that

$$\sup_{0<y<x_0-u} |F_u(y) - G(y; \sigma(u), k)| = O(\phi(u)), \tag{2.3}$$

where $\phi(u) \to 0$ as $u \to x_0$. In this case, we might expect (2.1) to remain valid, the effect of ϕ being to introduce a bias of $O(\phi(u))$ in $\mathbf{U}_N$. This will lead to a bias in (2.2).

To make this more precise, suppose that $N \to \infty$, $u = u_N \to x_0$ and $\sigma = \sigma_N = \sigma(u_N)$ and suppose $\phi(u_N) = O(N^{-1/2})$. We will typically be able to show that

$$N^{-1/2}E\{\mathbf{U}_N(\sigma_N, k)\} \to \mathbf{b}, \quad \text{for some finite vector } \mathbf{b}, \tag{2.4}$$

$$NI_N^{-1}(\sigma', k') \to_p M^{-1}, \quad \text{uniformly over } (\sigma', k'), \tag{2.5}$$

in a neighbourhood of the form $|\sigma'/\sigma_N - 1| < \varepsilon_N$, $|k_N - k| < \varepsilon_N$, where $N^{1/2}\varepsilon_N \to +\infty$. It will then follow that

$$N^{1/2}\begin{bmatrix} \hat{\sigma}_N/\sigma_N - 1 \\ \hat{k}_N - k \end{bmatrix} \to_d \mathscr{N}[-M^{-1}b, M^{-1}]. \tag{2.6}$$

In many cases we will have $\mathbf{b} = \mathbf{0}$, but much of the interest in the theory lies in the possibility of developing limit theorems when $\mathbf{b} \neq \mathbf{0}$, since this corresponds to a nonnegligible bias due to the approximation of F_u by the GPD.

Our main results are based on this argument, but the details are different for each of the three limit laws so we consider them separately.

3. Limit law Φ_α. Let F be in the domain of attraction of Φ_α for some $\alpha > 0$. Then $L(x) = x^\alpha(1 - F(x))$ is *slowly varying* at ∞. We shall assume L satisfies one of the following conditions:

SR1. $L(tx)/L(x) = 1 + O(\phi(x))$, as $x \to \infty$ for each $t > 0$,

SR2. $L(tx)/L(x) = 1 + k(t)\phi(x) + o(\phi(x))$, as $x \to \infty$ for each $t > 0$,

where $\phi(x) > 0$ and $\phi(x) \to 0$ as $x \to \infty$. In the case SR2, if we exclude trivial cases, ϕ must be regularly varying with some index $\rho \le 0$ (notation: $\phi \in \mathbb{R}_\rho$) and $k(t)$ is then of the form $k(t) = ch_\rho(t)$, $h_\rho(t) = \int_1^t u^{\rho-1}\,du$. We shall not attempt a detailed justification of these conditions [see Goldie and Smith (1987)] but shall remark that these are very general conditions. It is easily seen that $L(x) = C(1 + Dx^{-\beta} + o(x^{-\beta}))$ [assumed by Hall (1982a)] satisfies SR2 with $\beta = -\rho$ and many other forms of remainder terms are also encompassed by the SR2 condition. For an application to rates of convergence in extreme value theory, see Smith (1982).

The following result of Goldie and Smith (1987) will be used repeatedly:

PROPOSITION 3.1. *Let L satisfy* SR1 *or* SR2 *with remainder function ϕ, and let v be a real-valued function on* $(1, \infty)$. *Suppose, for some $\rho \le 0$, either* SR2 *with $\phi \in \mathbb{R}_\rho$, or* SR1 *with ϕ nonincreasing and*

$$\phi(tx)/\phi(x) \le Ct^\rho, \qquad x \ge x_0 > 0,\ t > 1,\ C < \infty. \tag{3.1}$$

If either (a) $\rho = 0$ *and* $\int_1^\infty |v(t)| t^\varepsilon\,dt < \infty$ *for some* $\varepsilon > 0$, *or* (b) $\rho < 0$ *and v is integrable, then*

$$\int_1^\infty v(t)\{L(tx)/L(x)\}\,dt = \int_1^\infty v(t)\,dt + O(\phi(x)).$$

In case of SR2, *we also have*

$$\int_1^\infty v(t)\{L(tx)/L(x)\}\,dt = \int_1^\infty v(t)\,dt + \phi(x)\int_1^\infty v(t)k(t)\,dt + o(\phi(x)).$$

Our main result in this section is the following

THEOREM 3.2. *Suppose L satisfies* SR2. *Let $Y_1, \dots, Y_N$ be i.i.d. from F_{u_N}, where $N \to \infty$, $u_N \to \infty$ such that*

$$N^{1/2}c\phi(u_N)/(\alpha - \rho) \to \mu, \qquad -\infty < \mu < \infty. \tag{3.2}$$

Define $k = -1/\alpha$, $\sigma_N = u_N/\alpha$. Then there exists, with probability tending to 1, a local maximum $(\hat\sigma_N, \hat k_N)$ of the GPD likelihood function, such that

$$N^{1/2}\begin{bmatrix} \hat\sigma_N/\sigma_N - 1 \\ \hat k_N - k \end{bmatrix} \to_d \mathcal{N}\left[\begin{bmatrix} \dfrac{\mu(1-k)(1+2k\rho)}{1-k+k\rho} \\ \dfrac{\mu(1-k)k(1+\rho)}{1-k+k\rho} \end{bmatrix}, M^{-1}\right].$$

If L satisfies only SR1 *with ϕ nonincreasing, and if $N^{1/2}\phi(u_N) \to 0$, then the same result holds with $\mu = 0$.*

REMARK. The precise form of (3.2) is chosen for consistency with equation (4.3.3) of Goldie and Smith (1987).

PROOF. We give the proof only for SR2. The method follows Section 2, and the principal part of the proof is the calculation of **b**.

Suppose $Y \sim F_u$ and $r > 0$. Integrating by parts, we have by Proposition 3.1 that

$$E\{(1 + Y/u)^{-r}\} = 1 - r\int_1^{\infty} t^{-r-1-\alpha}\frac{L(ut)}{L(u)}\,dt$$

$$= \frac{\alpha}{r+\alpha} - \frac{cr\phi(u)}{(r+\alpha)(r+\alpha-\rho)} + o(\phi(u))$$

and a similar expression for $E\{\log(1 + Y/u)\}$. Using the expressions in the Appendix with $\sigma = u/\alpha$ and $k = -1/\alpha$, the first two moments of $-\sigma(\partial/\partial\sigma)\log g(Y; \sigma, k)$ are

$$-\frac{c\phi(u)}{\alpha+1-\rho} + o(\phi(u)), \qquad \frac{\alpha}{\alpha+2} + O(\phi(u))$$

and those of $-(\partial/\partial k)\log g(Y; \sigma, k)$ are

$$\frac{\alpha c\phi(u)}{(\alpha-\rho)(\alpha+1-\rho)} + o(\phi(u)), \qquad \frac{2\alpha^2}{(\alpha+1)(\alpha+2)} + O(\phi(u)).$$

The third absolute moments of both these quantities are also bounded as $u \to \infty$.

Under assumption (3.2), we than have (2.4) with

$$\mathbf{b} = \frac{\mu}{\alpha+1-\rho}\begin{bmatrix} -\alpha+\rho \\ \alpha \end{bmatrix} = \frac{\mu}{1-k+k\rho}\begin{bmatrix} -1-k\rho \\ 1 \end{bmatrix}.$$

The remainder of the proof requires (a) the Lyapunov CLT for $\mathbf{U}_N$—this follows easily from the preceding remarks—and (b) (2.5), which is most easily verified by bounding the expected value of the third derivatives of the log likelihood. We omit the details of these operations.

In application, it is usual to think of either N or u_N being fixed with respect to total sample size n and the other being random. Then the left-hand side of (3.2) is random, but it may still be the case that (3.2) holds in probability and then Theorem 3.2 remains valid. This is obvious from a Skorohod-type construction of probability spaces on which (3.2) holds almost surely. For example:

1. Fix threshold v_n such that $v_n \to \infty$ as $n \to \infty$, let $N = N_n$ denote the number of exceedances of v_n in a total sample of size n and let $u_N = v_n$. If $n\{1 - F(v_n)\} \to \infty$, $n^{1/2}\{1 - F(v_n)\}^{1/2}c\phi(v_n) \to \mu(\alpha - \rho)$, then $Nn^{-1}\{1 - F(v_n)\}^{-1} \to_p 1$ and (3.2) holds in probability. Theorem 3.2 holds conditionally on $\{N_n\}$, hence also unconditionally.

2. Fix $N = N_n$ so that $N \to \infty$ and $N/n \to 0$. Let u_N denote the $(N+1)$st order statistic and define $v_n = F^*(n/N)$, where $F^*(t) = \inf\{x\colon 1 - F(x) < t^{-1}\}$.

Suppose SR2 with $N^{1/2}c\phi(v_n) \to \mu(\alpha - \rho)$. We claim that (3.2) holds in probability, or equivalently $\phi(v_n)/\phi(u_N) \to_p 1$. To see this, write $u_n = F^*(W_n^{-1})$, where W_n is a beta r.v. satisfying $nW_n/N \to_p 1$ (Chebyshev); note that F^*, ϕ and hence $\phi \circ F^*$ are each regularly varying; use the fact that if $\{X_n\}$ and $\{Y_n\}$ are sequences of random variables, each tending to ∞ with $X_n/Y_n \to_p 1$ and if R is regularly varying, then $R(X_n)/R(Y_n) \to_p 1$. [This argument depends on ϕ being regularly varying and hence is only valid for SR2. Different kinds of conditions on F are assumed by Haeusler and Teugels (1985) and by Csörgő and Mason (1985).]

Under Hall's (1982a) assumption $L(x) = C\{1 + Dx^{-\beta} + o(x^{-\beta})\}$, we have $\phi(x) = x^{-\beta}$, $\rho = -\beta$ and $c = -\beta D$. If we consider $v_n = An^{\gamma}$ and let

$$(3.3) \qquad \gamma = (\alpha + 2\beta)^{-1}, \qquad \mu = -A^{-\alpha/2-\beta}C^{1/2}D\beta(\alpha + \beta)^{-1},$$

then (3.2) holds in probability and Theorem 3.2 follows. □

4. Comparison with Hill's estimator. In the notation of Section 3, $\hat{\alpha}_N = -1/\hat{k}_N$ is an estimator of α. The conclusion of Theorem 3.2 gives

$$(4.1) \qquad N^{1/2}(\hat{\alpha}_N - \alpha) \to_d \mathscr{N}\left[-\frac{\mu\alpha(\alpha+1)(1+\rho)}{\alpha + 1 - \rho}, \alpha^2(\alpha+1)^2\right].$$

This may be compared with the corresponding result for Hill's estimator (1.4),

$$(4.2) \qquad N^{1/2}(\bar{\alpha}_N - \alpha) \to_d \mathscr{N}[-\mu\alpha, \alpha^2]$$

[Goldie and Smith (1987)]. Comparing (4.1) and (4.2), we see that the asymptotic variance of $\hat{\alpha}_N$ is higher than that of $\bar{\alpha}_N$, but the bias may well be smaller.

To make a more in depth comparison suppose SR2 holds, that v_n is the threshold associated with sample size n and let N denote the number of exceedances over $u_N = v_n$. Assuming (3.2) in probability, (4.2) gives the asymptotic mean-squared error of $\bar{\alpha}_N$ as

$$\alpha^2(1 + \mu^2)/N \sim \alpha^2(1 + \mu^2)/\{n(1 - F(v_n))\}.$$

The denominator is asymptotically $\mu^{2\alpha/(\alpha-2\rho)}$ times a fixed function of n, so as a function of μ this is proportional to

$$\alpha^2(1 + \mu^2)\mu^{-2\alpha/(\alpha-2\rho)}.$$

Similarly, the asymptotic mean-squared error of α_N is proportional to

$$\alpha^2(\alpha+1)^2\mu^{-2\alpha/(\alpha-2\rho)}\left\{1 + \mu^2(1+\rho)^2(\alpha + 1 - \rho)^{-2}\right\}.$$

Choosing μ (separately) to minimise each of the last two expressions, the ratio of the two becomes

$$\lim \frac{\text{minimum m.s.e. for } \hat{\alpha}_N}{\text{minimum m.s.e. for } \bar{\alpha}_N} = (\alpha+1)^2\left(\frac{1+\rho}{\alpha + 1 - \rho}\right)^{2\alpha/(\alpha-2\rho)}$$

This expression is 1 for $\rho = 0$, decreases with $|\rho|$ to be 0 at $\rho = -1$, thereafter increases to $(\alpha+1)^2$ as $|\rho| \to \infty$. The broad conclusion is that Hill's estimator is superior for $|\rho|$ large (when the Pareto tail is a very good fit) but the GPD estimator is superior when $|\rho|$ is small. The conservative choice might be to use the GPD estimator on the grounds that this gives greater protection against a poor Pareto fit.

This comparison is artificial in that it assumes the threshold to be chosen optimally, which would be hard to achieve in practice, but there are other considerations which might influence the practical statistician. For example, Hill's estimator is not location invariant and its use therefore seems unnatural when the origin of the scale of measurement is arbitrary. The case $\rho = -1$ is peculiar: An appropriate change of origin removes the leading error term and so makes possible an improved overall rate of convergence. A similar phenomenon has been observed in connection with the rate of convergence of $F^n(a_n x + b_n)$ to $\Phi_\alpha(x)$ [Smith (1982)]. Note that this requires SR2 not SR1 and therefore does not contradict the main result of Hall and Welsh (1984).

5. Limit law Ψ_α, $\alpha > 2$. The distribution function F is in the domain of attraction of Ψ_α, $\alpha > 0$, if and only if

$$(5.1)\quad x_0 = \sup\{x: F(x) < 1\} < \infty \quad \text{and} \quad L(x) = x^\alpha\{1 - F(x_0 - x^{-1})\}$$

is slowly varying at ∞. Our results for the three cases $\alpha > 2$, $\alpha = 2$ and $\alpha < 2$ are quite different in character, so they are dealt with in three separate sections. Here we consider $\alpha > 2$.

The basic procedure is again to fit GPD to the N excesses over a threshold u, obtaining maximum-likelihood estimates $\hat{\sigma}_N$, $\hat{k}_N$. If $\hat{k}_N > 0$, then $u + \hat{\sigma}_N/\hat{k}_N$ is an estimator of x_0. This estimator, formulated slightly differently, was shown to be consistent by Smith and Weissman (1985) provided $\alpha > 1$, and asymptotically normal by Hall (1982b) provided $\alpha > 2$. Some of Hall's results also cover the case $1 < \alpha \le 2$. In this section we generalise Hall's result for $\alpha > 2$ in two principal ways, first by allowing SR1 or SR2 (see Section 3) in place of the narrower conditions assumed by Hall, and second by weakening the conditions on N and u to include the case $\mathbf{b} \neq \mathbf{0}$, in the notation of Section 2.

THEOREM 5.1. *Assume* (5.1) *with L obeying* SR2 *for $\phi \in \mathbb{R}_\rho$, $\rho \le 0$, and $k(t) = ch_\rho(t)$. Let $N \to \infty$, $u_N \to x_0$ such that*

$$(5.2)\qquad N^{1/2} c\phi\left((x_0 - u_N)^{-1}\right)/(\alpha - \rho) \to \mu.$$

For each N, let $Y_1, \dots, Y_N$ be i.i.d. with d.f. F_{u_N}. Then, with probability tending to 1, *there exists a local maximum $(\hat{\sigma}_N, \hat{k}_N)$ of the GPD likelihood, such that*

$$(5.3)\qquad N^{1/2}\begin{bmatrix} \hat{\sigma}_N/\sigma_N - 1 \\ \hat{k}_N - k \end{bmatrix} \to_d \mathcal{N}\left[\begin{bmatrix} \dfrac{\mu(1-k)(1-2k\rho)}{(1-k-k\rho)} \\ \dfrac{\mu k(1-k)(1-\rho)}{1-k-k\rho} \end{bmatrix}, M^{-1}\right].$$

Here, $k = 1/\alpha$ and $\sigma_N = (x_0 - u_N)/\alpha$. If L satisfies only SR1 *with ϕ nonincreasing, and $N^{1/2}\phi((x_0 - u_N)^{-1}) \to 0$, then the same result holds with $\mu = 0$.*

SKETCH OF PROOF. Assume SR2. If $Y \sim F_u$, then by similar arguments to those in Section 3 the means of $-\sigma(\partial/\partial\sigma)\log g(Y;\sigma,k)$, $-(\partial/\partial k)\log g(Y;\sigma,k)$ evaluated at $\sigma = (x_0 - u)/\alpha$, $k = 1/\alpha$, are, respectively,

$$-\frac{c}{\alpha-\rho-1}\phi\big((x_0-u)^{-1}\big)(1+o(1)),$$

$$\frac{\alpha c}{(\alpha-\rho)(\alpha-\rho-1)}\phi\big((x_0-u)^{-1}\big)(1+o(1)),$$

$u \to x_0$. Calculations as in Section 2 then lead to (5.3), with a similar conclusion for SR1.

ESTIMATION OF ENDPOINT. If we define $\theta_N = \sigma_N/k$ and $\hat{\theta}_N = \hat{\sigma}_N/\hat{k}_N$, then rearrangement of (5.3) leads to

$$(5.4)\quad N^{1/2}\big(\hat{\theta}_N/\theta_N - 1\big) \to_d \mathscr{N}\left[\frac{\mu(\alpha-1)(\alpha-2)\rho}{\alpha(\alpha-\rho-1)}, \frac{(\alpha-2)(\alpha-1)^2}{\alpha}\right].$$

We compare this with Hall (1982b), Theorem 6. Let $L(x) = C(1 + Dx^{-\beta} + o(x^{-\beta}))$ and consider total sample size n, threshold $v_n = x_0 - A^{-1}n^{-\gamma}$, where $0 < \gamma < \alpha^{-1}$ and the number of exceedances N and $u_N = v_n$. Then $NA^{\alpha}C^{-1}n^{\alpha\gamma-1} \to_p 1$ and (5.2) is equivalent to (3.3). We estimate x_0 by $\hat{x}_0 = u_N + \hat{\theta}_N$ and a simple calculation shows $(N/n)^{1/\alpha} \sim C^{1/\alpha}(x_0 - u_N)$ in probability. Consequently, (5.4) implies

$$(5.5)\quad \begin{aligned} N^{1/2}(N/N)^{1/\alpha}(\hat{x}_0 - x_0) &\sim_p N^{1/2}C^{-1/\alpha}\theta_N^{-1}\big(\hat{\theta}_N - \theta_N\big) \\ &\to_d \mathscr{N}\left[-\frac{\mu(\alpha-1)(\alpha-2)\beta}{C^{1/\alpha}\alpha(\alpha+\beta-1)}, \frac{(\alpha-2)(\alpha-1)^2}{C^{2/\alpha}\alpha}\right]. \end{aligned}$$

In this case the mean-squared error of $\hat{x}_0$ is $O(N^{-1}\theta_N^2) = O(n^{-(2+2\beta)/(\alpha+2\beta)})$. Hall's estimator differed only in that $N = N_n$ was treated as a deterministic sequence and u_N defined to be the Nth largest order statistic from the full sample. Hall obtained the $\mu = 0$ case of (5.5) under the condition $L(x) = C + O(x^{-\beta})$ and the additional condition $N_n = o(n^{2/3})$.

6. Limit law Ψ_α, $\alpha = 2$.

In this section we consider the case $\alpha = 2$ and make the simplification of assuming that α is known to the statistician. We could also consider the case in which α is treated as an unknown, but results of Smith (1985) suggest that we would get the same result for the asymptotic estimation of x_0, so we do not consider this case in detail.

We assume that the endpoint x_0 is finite and that $1 - F(x_0 - x) = x^2L(x^{-1})$, where L is slowly varying at ∞. If $u = x_0 - \theta$, $\theta > 0$, and $Y_1,\dots,Y_N$ are i.i.d. excesses over u, then the common distribution function of Y_i, $1 \le i \le N$, is

$$F_u(y) = \frac{F(u+y) - F(u)}{1 - F(u)} \sim 1 - \left(1 - \frac{y}{\theta}\right)^2, \qquad \theta \downarrow 0,\quad 0 < y < \theta.$$

An approximate log likelihood for θ is then

$$l_N(\theta) = \sum_{i=1}^{N} \log(\theta - Y_i) - 2N \log \theta + \text{const.}, \tag{6.1}$$

with derivatives

$$-\theta l_N'(\theta) = 2N - \sum \frac{\theta}{\theta - Y_i}, \qquad -\theta^2 l_N''(\theta) = -2N + \sum \left(\frac{\theta}{\theta - Y_i} \right)^2. \tag{6.2}$$

We may therefore define an approximate maximum-likelihood estimator of θ (based on the excesses over u) as that value $\bar{\theta}_N$ which maximises $l_N(\theta)$. In turn, $\bar{x}_0 = u + \bar{\theta}_N$ is an estimator of x_0.

Previous results on this problem are due to Woodroofe (1972), Weiss and Wolfowitz (1973) and Hall (1982b). The first two papers assumed L known, so that x_0 is the only unknown parameter and is a location parameter, and proved asymptotic normality and asymptotic efficiency of the usual maximum-likelihood estimator. Hall considered essentially the same estimator as ours, but only for a much narrower class of slowly varying functions L, which does not bring out the full range of possibilities. Our result is

THEOREM 6.1. *Suppose L satisfies* SR1 *with a nonincreasing remainder function ϕ. Suppose $N \to \infty$ and $u_N \to x_0$ such that $N^{1/2}\phi((x_0 - u_N)^{-1})$ is bounded. Let $\theta_N = x_0 - u_N$. Define*

$$L_1(x) = \int_1^x L(y) y^{-1} \, dy,$$

$$\mu(x, \theta) = 1 - \frac{L(\theta^{-1}x)}{L(\theta^{-1})} + 2\frac{L_1(\theta^{-1}x) - L_1(\theta^{-1})}{L(\theta^{-1})}.$$

Let $Y_{N,1}, \dots, Y_{N,N}$ be i.i.d. from F_{u_N} and define $l_N(\theta)$ to be $\sum \log(\theta - Y_{N,i}) - 2N\log\theta$.

Then, there exists a nonrandom sequence $\{s_N\}$ such that

$$N s_N^{-2} \mu(s_N, \theta_N) \to 1, \tag{6.3}$$

$$s_N^2 / N \to \infty \tag{6.4}$$

and, with probability tending to 1, there exists a local maximum $\bar{\theta}_N$ of $l_N(\theta)$ such that

$$\frac{s_N}{\theta_N}(\bar{\theta}_N - \theta_N) \to_d \mathcal{N}(0,1). \tag{6.5}$$

Before proving this, we make some remarks about its relation to earlier work. The generality of Theorem 6.1 stems from the fact that L is not assumed to converge at ∞. If we do assume this, however, the result may be simplified. Suppose $L(x) \to \alpha/2$ as $x \to \infty$; in turn, a sufficient condition for that is $F''(x) \to \alpha$ as $x \to x_0$, cf. Woodroofe (1972). Then $L_1(x) \sim \alpha \log x/2$, as $x \to \infty$ and (6.3) may be satisfied by choosing s_N so that

$$s_N^2 \sim N \log N, \qquad N \to \infty.$$

Let us now interpret the result as an unconditional result about the estimation of the endpoint x_0. We have $x_0 = u_N + \theta_N$, $\bar{x}_0 = u_N + \bar{\theta}_N$, say. Denote the total sample size by n, the threshold u_n instead of u_N and let $N \sim n\{1 - F(u_n)\} \sim \alpha n(x_0 - u_n)^2/2$ denote the random number of exceedances over the threshold. If we write α_n in place of s_N/θ_N, then (6.5) becomes

$$\alpha_n(\bar{x}_0 - x_0) \to_d \mathcal{N}(0,1), \qquad n \to \infty.$$

But $\alpha_n^2 \sim N \log N/\theta_N^2 \sim \alpha n \log N/2$ exactly, as in Woodroofe (1972). This leads to the following conclusions.

1. Our estimator $\bar{x}_0$ has exactly the same asymptotic distribution as the maximum-likelihood estimator derived under the assumption that L is known everywhere, since this is essentially the case dealt with by Woodroofe (1972). Thus, there is no loss of information as a result of not knowing L.

2. The only restriction on the sequence of thresholds is that $N^{1/2}\phi((x_0 - u_N)^{-1})$ must be bounded—this is a condition which says that N must not grow too quickly. Thus, there is no optimum rate of convergence in this case and N may be chosen so as to grow only very slowly. Hall (1982b) also remarked on this point.

PROOF OF THEOREM 6.1. The proof is split into a series of lemmas. First, define $X_{N,i} = \theta_N/(\theta_N - Y_{N,i})$ and note that

$$P\{X_{N,i} > x\} = x^{-2}L(\theta_N^{-1}x)/L(\theta_N^{-1}), \tag{6.6}$$

$$\mu(x, \theta_N) = E\{X_{N,i}^2; 1 < X_{n,i} \le x\}. \tag{6.7}$$

LEMMA 6.2. *The sequence* $\{s_N\}$ *exists, satisfying* (6.3), (6.4) *and*

$$\lim_{N\to\infty} NP\{X_{N,1} > \varepsilon s_N\} = 0, \quad \textit{for all } \varepsilon > 0. \tag{6.8}$$

PROOF. By de Haan (1970), Section 1.2, L_1 is slowly varying and $L_1(x)/L(x) \to \infty$ as $x \to \infty$. By this and (6.7), μ is a nondecreasing and slowly varying function of x for each θ. For fixed x, Fatou's lemma gives

$$\liminf_{\theta\to 0} \mu(x, \theta) = 2 \liminf_{\theta\to 0} \int_1^x \frac{L(\theta^{-1}y)}{L(\theta^{-1})} \frac{dy}{y} \ge 2 \log x. \tag{6.9}$$

Thus, $x^{-2}\mu(x, \theta)$ is bounded away from 0 as $\theta \to 0$ for any fixed $x > 1$, but tends to 0 as $x \to \infty$ for fixed θ. For sufficiently large N, we can find s_N such that $\mu(s_N - \varepsilon, \theta_N) < s_N^2/N < \mu(s_N + \varepsilon, \theta_N)$ for any $\varepsilon > 0$. Suppose $s_N \nrightarrow \infty$. Then $s_N \le \tau < \infty$ infinitely often, and $\mu(\tau - \varepsilon, \theta_N) \to 0$ along a subsequence. This is impossible by (6.9). Therefore, $s_N \to \infty$. For fixed $\gamma > 0$, we have $\mu(s_N\gamma, \theta_N)/\mu(s_N, \theta_N) \to 1$ by slow variation of L and L_1; hence, (6.3) follows. Suppose $s_N^2/N \nrightarrow \infty$. Then there exists $M < \infty$ such that $\mu(s_N, \theta_N) < M$ infinitely often. But for any fixed τ, $s_N > \tau$ eventually, so by (6.9), $M \ge 2 \log \tau$. This contradiction proves (6.4). Finally, suppose (6.8) fails. Using (6.6) and (6.3), it follows that there is a subsequence along which $\mu(s_N, \theta_N)L(\theta_N^{-1})/L(\theta_N^{-1}s_N)$ is bounded. In particular, $\{L_1(\theta_N^{-1}s_N) - L_1(\theta_N^{-1})\}/L(\theta_N^{-1}s_N)$ is bounded. But the last expression is eventually at least as large as $\{L_1(\theta_N^{-1}s_N) - L_1(\theta_N^{-1}s_N x^{-1})\}/L(\theta_N^{-1}s_N)$ for any fixed $x > 1$ and by Fatou's lemma the lim inf of this is at least $\log x$. Hence, (6.8) is proved by contradiction and the proof of the lemma is complete. □

LEMMA 6.3. $s_N^{-1}\theta_N l_N'(\theta_N) \to_d \mathscr{N}(0,1)$, $s_N^{-2}\theta_N^2 l_N''(\theta_N) \to_p 1$.

PROOF. We have $E\{X_{N,i}\} = 2 + O(\phi(\theta_N^{-1})) = 2 + o(s_N^{-1})$ using Proposition 3.1, the assumptions of Theorem 6.1 and (6.4). This condition, together with (6.3), (6.7) and (6.8), suffices for Theorem 16, Chapter 4 of Petrov (1975) and hence via (6.2) for the first statement of the lemma. The second statement follows similarly from the weak law of large numbers for triangular arrays [Petrov (1975), Chapter 9, Theorem 3] or from Raikov's theorem [see Hall and Heyde (1980), Section 3.4]. □

LEMMA 6.4. *There exists a sequence* $\{t_N\}$ *such that* $s_N t_N^{-1} \to 0$, $s_N^3 t_N^{-1} \to \infty$ *and* $t_N^{-1}\Sigma_i X_{N,i}^3 \to_p 0$. *With* $X_{N,1}^* \geq \cdots \geq X_{N,N}^*$ *the order statistics of* $X_{N,i}$, $1 \leq i \leq N$, *we then have*

$$t_N^{-1}\sum_{i=2}^{N}\left\{(X_{N,i}^*)^{-1} - (X_{N,1}^*)^{-1}\right\}^{-3} \to_p 0. \tag{6.10}$$

PROOF. Choose $t_N = \delta_N s_N^3$ so that $\delta_N \to 0$ and

$$Nt_N^{-2/3}L\left(\theta_N^{-1}t_N^{1/3}\right)/L\left(\theta_N^{-1}\right) \to 0, \qquad s_N t_N^{-1} \to 0. \tag{6.11}$$

To see that this is possible, observe that if $\delta_N \equiv \delta > 0$, then (6.11) is true by Lemma 6.2. But then it follows, by a diagonalisation argument, that it must be possible to choose $\delta_N \to 0$ so that (6.11) remains true.

To show that $t_N^{-1}\Sigma X_{N,i}^3 \to_p 0$, it suffices by Theorem 3, Chapter 9 of Petrov (1975) to prove that

$$NP\left\{X_{N,1}^3 > \varepsilon t_N\right\} \to 0, \quad \text{for all } \varepsilon > 0, \tag{6.12a}$$

$$Nt_N^{-1}E\left\{X_{N,1}^3;\, X_{N,1}^3 < t_N\right\} \to 0, \tag{6.12b}$$

$$Nt_N^{-2}E\left\{X_{n,1}^6;\, X_{N,1}^3 < t_N\right\} \to 0. \tag{6.12c}$$

Now (6.12a) is just (6.8). For (6.12b, c) observe that, for $r > 2$,

$$\begin{aligned} E\{X_{N,1}^r;\, X_{N,1} < x\} &= 1 - x^{r-2}\frac{L\left(\theta_N^{-1}x\right)}{L\left(\theta_N^{-1}\right)} + r\int_1^x y^{r-3}\frac{L\left(\theta_N^{-1}y\right)}{L\left(\theta_N^{-1}\right)}\,dy \\ &< 1 + r\theta_N^{r-2}\frac{R\left(\theta_N^{-1}x\right) - R\left(\theta_N^{-1}\right)}{L\left(\theta_N^{-1}\right)}, \end{aligned}$$

where $R(x) = \int_1^x y^{r-3}L(y)\,dy \sim x^{r-2}L(x)/(r-2)$ by de Haan (1970), Section

1.2. Hence,

$$t_N^{-r/3}E\{X_{N,1}^r;\ X_{N,1}^3 < t_N\} = O\left\{t_N^{-2/3}\frac{L(\theta_N^{-1}t_N^{1/3})}{L(\theta_N^{-1})}\right\}$$

and (6.12b, c) follow from (6.12a).

To show (6.10), write

$$\{(X_{N,i}^*)^{-1} - (X_{N,1}^*)^{-1}\}^{-3} = (X_{N,i}^*)^3\left[1 - \frac{X_{N,i}^*}{X_{N,1}^*}\right]^{-3} \le (X_{N,i}^*)^3\left[1 - \frac{X_{N,2}^*}{X_{N,1}^*}\right]^{-3}.$$

But it is easy to show that $X_{N,2}^*/X_{N,1}^*$ is bounded away from 1 (in probability) as $N \to \infty$ and the result then follows from the first part of the lemma.

We can now complete the proof of Theorem 6.1. By familiar arguments, $\bar{\theta}_N$ exists and satisfies $\bar{\theta}_N - \theta_N = -l_N'(\theta_N)/l_N''(\theta_N^*)$, where θ_N^* lies between θ_N and $\bar{\theta}_N$. In view of Lemma 6.3, it suffices to show that there exists a sequence $\{\varepsilon_N\}$ with $\varepsilon_N > 0$, $\varepsilon_N s_N/\theta_N \to \infty$ such that

$$s_N^{-2}\theta_N^2 \sup_{|\theta-\theta_N|<\varepsilon_N} |l_N''(\theta) - l_N''(\theta_N)| \to_p 0. \tag{6.13}$$

Choose $\varepsilon_N = \theta_N(s_N/t_N)^{1/2}$. Taking a Taylor expansion in θ, we have to show

$$Ns_N^{-2}\theta_N^2\varepsilon_N\theta^{-3} \to 0, \qquad s_N^{-2}\theta_N^2\varepsilon_N\sum(\theta - Y_{N,i})^{-3} \to_p 0, \tag{6.14}$$

uniformly over $|\theta - \theta_N| < \varepsilon_N$. The first part is automatic from (6.4) and $s_N t_N^{-1} \to 0$. For the second part, it suffices to restrict ourselves to $\theta < Y_{N,1}^*$ and hence to show

$$\begin{aligned} &s_N^{-2}\theta_N^2\varepsilon_N\sum_{i=2}^{N}(Y_{N,1}^* - Y_{N,i}^*)^{-3} \to_p 0, \\ &s_N^{-2}\theta_N^2\varepsilon_N(\theta_N - \varepsilon_N - Y_{N,1}^*)^{-1} \to_p 0. \end{aligned} \tag{6.15}$$

The second limit in (6.15) holds because $Y_{N,1}^* = o_p(\theta_N)$. The first limit follows from Lemma 6.4 after substituting $Y_{N,i}^* = \theta_N - \theta_N/X_{N,i}^*$. With this, the proof of Theorem 6.1 is complete. □

7. Limit law Ψ_α, $0 < \alpha < 2$. In this section we assume that $1 - F(x) \sim c(x_0 - x)^\alpha$ as $x \uparrow x_0$, where $c > 0$, $0 < \alpha < 2$. The simplest case is when x_0 is a pure location parameter, so that $F(x - x_0)$ is fully specified. In this case, the density $f = F'$ is typically J-shaped when $\alpha \le 1$, so that a local maximum-likelihood estimator exists only when $\alpha > 1$. Woodroofe (1974) showed that when $\alpha > 1$, the maximum-likelihood estimator of x_0 is $n^{1/\alpha}$-consistent, and he obtained the (nonnormal) limiting distribution. Akahira (1975) showed, for the full range $0 < \alpha < 2$, that $n^{1/\alpha}$ is the optimum rate of consistency and Ibragimov and Has'minskii (1981) obtained the asymptotic distributions of both maximum-likelihood and Bayes estimators in a variety of cases, although they did not obtain explicit formulas for these asymptotic distributions.

Smith (1985) studied estimation of θ in the case

$$f(x;\theta,\phi) = (x-\theta)^{\alpha-1} g(x-\theta;\phi), \qquad x < \theta,$$

where ϕ is a vector of nuisance parameters, $\alpha \equiv \alpha(\phi)$ may be known or unknown and g is a known function tending to a constant as $x - \theta \to 0$. In this case, for $1 < \alpha < 2$, the maximum-likelihood estimator was shown to have the same asymptotic distribution as when ϕ is known. Similar results were also presented in the case $0 < \alpha \leq 1$. The problem considered in this section is a kind of infinite-dimensional extension: The function g is assumed completely unknown, subject only to smoothness and integrability constraints. As in Section 6, our eventual conclusion is that ignorance about g has no effect on the asymptotic estimation of x_0. We show this by adapting some results of Ibragimov and Has'minskii.

It is convenient to recast the problem in terms of estimating a lower endpoint. We shall assume that $F(x) \sim C(x-\theta)^\alpha$ as $x \downarrow \theta$, where θ, c and α are unknown. Since θ is a location parameter and the proposed estimator is location invariant, there is no loss of generality in assuming the true value of θ to be 0.

We assume that

$$F(x) = Cx^\alpha\{1 + Dx^\beta + o(x^\beta)\}, \quad \text{as } x \downarrow 0, \qquad C > 0, \beta > 0, 0 < \alpha < 2, \tag{7.1}$$

and that the same relation remains valid under two differentiations, viz.,

$$f(x) = F'(x) = \alpha C x^{\alpha-1}\{1 + D\alpha^{-1}(\alpha+\beta)x^\beta + o(x^\beta)\}, \tag{7.2}$$

$$f'(x) = \alpha(\alpha-1)Cx^{\alpha-2} \times \{1 + D\alpha^{-1}(\alpha-1)^{-1}(\alpha+\beta)(\alpha+\beta-1)x^\beta + o(x^\beta)\}. \tag{7.3}$$

We also assume that f'' exists and that

$$\limsup_{\varepsilon\to 0} \int_u^\infty \frac{f''(x+\varepsilon u)}{f(x+\varepsilon u)} f(x)\,dx = O(u^{\alpha-2}), \qquad u \downarrow 0, \tag{7.4}$$

$$\limsup_{\varepsilon\to 0} \int_u^\infty \left\{\frac{f'(x+\varepsilon u)}{f(x+\varepsilon u)}\right\}^2 f(x)\,dx = O(u^{\alpha-2}), \qquad u \downarrow 0, \tag{7.5}$$

$$\int_0^\infty |f'(x)|\,dx < \infty \qquad (\alpha = 1), \qquad \int_0^\infty x^\delta f(x)\,dx < \infty, \quad \text{for some } \delta > 0 \qquad (\alpha \neq 1). \tag{7.6}$$

These assumptions are considerably stronger than those in the preceding sections. They are needed for application of the Ibragimov–Has'minskii results. N. H. Bingham has pointed out to me that, by the results in Section 7 of Balkema, Geluk and de Haan (1979), it is possible to sandwich any regularly varying function F between two "smoothly" varying functions, say $F_1 \leq F \leq F_2$, whose derivatives exist and have the anticipated limits. Since $Z_{n,3}$ (defined below) is a monotone function of the order statistics, it could also be sandwiched between the corresponding functions defined for random variables from F_1 and F_2. Moreover, $\tilde{\alpha}_n$ (also defined below) can be shown to be a consistent estimator of α under much weaker conditions than (7.1)–(7.6), so that this result could be extended to $Z_{n,4}$. Thus, we might extend the following results so that they hold under weaker conditions. We shall not attempt to fill in the details of this.

Consequences of (7.1)–(7.3) are the relations

$$(7.7)\qquad \begin{aligned} f(x) &= \exp\left\{-\int_x^\infty \eta(y)y^{-1}\,dy\right\},\\ F(x) &= \exp\left\{-\int_x^\infty \eta_0(y)y^{-1}\,dy\right\},\end{aligned}$$

where

$$(7.8)\qquad \begin{aligned} \eta(y) &= yf'(y)/f(y) = \alpha - 1 + O(y^\beta),\\ \eta_0(y) &= yf(y)/F(y) = \alpha + O(y^\beta).\end{aligned}$$

Suppose the total sample size is n and that there are N $(= N_n)$ observations below a threshold u $(= u_n)$. We assume

$$(7.9)\qquad u_n > 0, \qquad nu_n^\alpha \to \infty, \qquad u_n \sim \gamma n^{-1/(\alpha+2\beta)}, \qquad 0 \le \gamma < \infty$$

and, consequently,

$$(7.10)\qquad \begin{aligned} &N_n \sim C\gamma^\alpha n^{2\beta/(\alpha+2\beta)}, \quad \text{in probability},\\ &N^{1/2}u_n^\beta \to_p C^{1/2}\gamma^{\beta+\alpha/2}.\end{aligned}$$

Let the observations be $X_1, \dots, X_n$, ordered as $X_{n,1}^* \le \cdots \le X_{n,n}^*$. Define

$$\tilde{\alpha}(= \tilde{\alpha}_n) = N\left[\sum_{i=2}^N \log\{(u - X_{n,1}^*)/(X_{n,i}^* - X_{n,1}^*)\}\right]^{-1}.$$

Consider the four likelihood ratio processes,

$$\begin{aligned} Z_{n,1}(t) &= \prod_{i=1}^n \left\{\frac{f(X_i + tn^{-1/\alpha})}{f(X_i)}\right\},\\ Z_{n,2}(t) &= \prod_{i:\,X_i<u} \left\{\frac{f(X_i + tn^{-1/\alpha})}{f(X_i)}\,\frac{F(u)}{F(u + tn^{-1/\alpha})}\right\},\\ Z_{n,3}(t) &= \prod_{i:\,X_i<u} \left\{\left(1 + tn^{-1/\alpha}X_i^{-1}\right)^{\alpha-1}\left(1 + tn^{-1/\alpha}u^{-1}\right)^{-\alpha}\right\},\\ Z_{n,4}(t) &= \prod_{i:\,X_i<u} \left\{\left(1 + tn^{-1/\alpha}X_i^{-1}\right)^{\tilde{\alpha}-1}\left(1 + tn^{-1/\alpha}u^{-1}\right)^{-\tilde{\alpha}}\right\},\end{aligned}$$

each defined for $t > n^{-1/\alpha}X_{n,1}^*$. The motivation for these definitions is as follows. $Z_{n,1}(t)$ is the likelihood ratio statistic for testing $\theta = 0$ versus $\theta = -tn^{-1/\alpha}$ based on the full sample, assuming that f is fully specified but for the unknown θ. This is viewed as a stochastic process in t. $Z_{n,2}(t)$ is the likelihood ratio statistic based on observations censored at u; $Z_{n,3}$ and $Z_{n,4}$ are approximations to $Z_{n,2}$ in which α is assumed known but nothing else ($Z_{n,3}$) and in which α is also unknown and estimated by $\tilde{\alpha}(Z_{n,4})$. In $Z_{n,4}$, we do not change the $n^{-1/\alpha}$ because this is a normalising constant which, as it turns out, is not affected by α being unknown. Thus, $Z_{n,2}$, $Z_{n,3}$ and $Z_{n,4}$ are successive approximations to $Z_{n,1}$ in which more and more information is discarded.

In the terminology of Ibragimov and Has'minskii (1981), Chapters 5–6, the density f possesses a singularity of order $\alpha - 1$ at 0. Their condition (1.3), page 282, is satisfied provided $\alpha < \lambda < \alpha + \beta$, by our (7.2) and (7.3). The process $Z_{n,1}$ defined above, then coincides with the process Z_n, defined by Ibragimov and Has'minskii. They show (Chapter 5, Theorem 2.1 and Chapter 6, Theorems 2.1 and 2.3) that the finite-dimensional distributions of Z_n converge to those of a stochastic process Z which they define. They also show (Chapter 5, Theorem 4.2 and Chapter 6, Theorem 6.2) that $n^{1/\alpha}(\tilde{\theta}_n - \theta)$ converges to a random variable defined from Z. Here $\tilde{\theta}_n$ is a Bayes estimator with respect to a continuous positive prior and a suitable loss function. In the case $\alpha > 1$ they prove a corresponding result for the maximum-likelihood estimator (Chapter 6, Theorem 6.4). These results assume (7.6). Note that the asymptotic distribution of the maximum-likelihood estimator was found in a completely different form by Woodroofe (1974).

Our principal result is

THEOREM 7.1.

$$\sup|Z_{n,1}(t)/Z_{n,2}(t) - 1| \to_p 0, \qquad \sup|Z_{n,2}(t)/Z_{n,3}(t) - 1| \to_p 0,$$

where the suprema are taken over $-n^{-1/\alpha}X_{n,1}^* < t \le M$ *for some arbitrary* $M < \infty$. *Furthermore,*

$$\sup|Z_{n,3}(t)/Z_{n,4}(t) - 1| \to_p 0,$$

where the supremum is taken over $-n^{-1/\alpha}X_{n,1}^* + mN^{-\gamma} < t \le M$, *for arbitrary* $\gamma < \infty$, $m > 0$ *and* $M < \infty$.

Theorem 7.1 may be interpreted as saying that, so long as attention is restricted to a compact set of values of t, all four processes $Z_{n,1}$–$Z_{n,4}$ converge to the same limit Z. Consequently, estimators constructed from these four processes have the same asymptotic distributions. This includes maximum-likelihood estimators (when $\alpha > 1$), Bayes estimators and also Pitman estimators which, as a referee has pointed out, are asymptotically efficient under squared error loss.

The estimation of α is a separate problem, of independent interest. The estimator $\tilde{\alpha}_n$ was proposed, in slightly different form, by Smith and Weissman (1985). Concerning its rate of consistency, we have

THEOREM 7.2. (a) *If* $0 < \alpha < 2$ *and the assumptions of this section hold, then*

$$N^{1/2}(\tilde{\alpha}_n - \alpha) \to_d \mathcal{N}\left(\gamma^{\beta+\alpha/2}C^{1/2}D\alpha\beta(\alpha+\beta)^{-1}, \alpha^2\right).$$

(b) *If the same assumptions hold with* $2 \le \alpha < \infty$, *then we still have* $N^{1/\alpha}(\tilde{\alpha}_n - \alpha) = O_p(1)$.

As a consequence of Theorem 7.2, the estimator $\tilde{\alpha}_n$ may be used to distinguish between the cases $\alpha < 2$, $\alpha > 2$. Thus, if $\tilde{\alpha}_n < 2$, we take $\tilde{\alpha}_n$ as our estimate of α and proceed to construct an estimator of θ based on $Z_{n,4}$. If $\tilde{\alpha}_n > 2$, we then switch to the methods of Sections 2–5 and estimate both α and θ using the GPD.

PROOF OF THEOREM 7.2. If it were known that $\theta = 0$, then we could define

$$\bar{\alpha}(=\bar{\alpha}_n) = N\left[\sum_{i=1}^{N} \log(u_N/X_{n,i}^*)\right]^{-1},$$

which is just Hill's estimator for $1 - F(1/x)$ (see Section 3). By (4.2) or Hall (1982a), we have

$$N^{1/2}(\bar{\alpha} - \alpha) \to_d \mathcal{N}\left(\gamma^{\beta+\alpha/2}C^{1/2}D\alpha\beta(\alpha+\beta)^{-1}, \alpha^2\right).$$

The proof now rests on comparing $\bar{\alpha}$ and $\tilde{\alpha}$. By adapting the proofs of Lemmas 1–3 in Smith (1985), it is possible to show

$$\log(1 - X_{n,1}^*/u_n) = O_p(N^{-1/\alpha}),$$

$$\sum_{i=2}^{N} \log(1 - X_{n,1}^*/X_{n,i}^*) = \begin{cases} O_p(N^{1-1/\alpha}), & \alpha > 1, \\ O_p(\log N), & \alpha \le 1. \end{cases} \tag{7.11}$$

It then follows that $\tilde{\alpha}^{-1} - \bar{\alpha}^{-1}$ is $O_p(N^{-1}\log N)$ when $\alpha \le 1$, $O_p(N^{-1/\alpha})$, when $\alpha > 1$. This suffices for the result. □

PROOF OF THEOREM 7.1. First, we compare $Z_{n,1}$ with $Z_{n,2}$. From the definitions, it follows that

$$\log\left\{\frac{Z_{n,1}(t)}{Z_{n,2}(t)}\right\} = \sum_{i:\, X_i > u} \log\left\{\frac{f(X_i + tn^{-1/\alpha})}{f(X_i)}\right\} + N\log\left\{\frac{F(u + tn^{-1/\alpha})}{F(u)}\right\}. \tag{7.12}$$

Now,

$$N\log\left\{\frac{F(u + tn^{-1/\alpha})}{F(u)}\right\} = Ntn^{-1/\alpha}\frac{f(u)}{F(u)} + \frac{N}{2}t^2n^{-2/\alpha}\left[\frac{f'(u^*)}{F(u^*)} - \left\{\frac{f(u^*)}{F(u^*)}\right\}^2\right],$$

where u^* is between u and $u + tn^{-1/\alpha}$. By (7.8) $f(u)/F(u) = u^{-1}\{\alpha + O(u^{\beta})\}$, while the expression in square brackets is $O(u^{-2})$. But $Nn^{-1/\alpha}u^{-1+\beta} \to_p 0$, $Nn^{-2/\alpha}u^{-2} \to_p 0$, so

$$N\log\left\{\frac{F(u + tn^{-1/\alpha})}{F(u)}\right\} = n^{1-1/\alpha}u^{\alpha-1}tC\alpha + o_p(1). \tag{7.13}$$

We apply the same techniques to the first term in (7.12), using (7.4) and (7.5) to deduce that

$$\sum_{i:\,X_i>u} \log\left\{\frac{f(X_i + tn^{-1/\alpha})}{f(X_i)}\right\} = tn^{-1/\alpha} \sum_{i:\,X_i>u} \frac{f'(X_i)}{f(X_i)} + o_p(1).$$

But the mean of $\{f'(X_i/f(X_i)\}I(X_i > u)$ is $-f(u) = -\alpha C u^{\alpha-1}(1 + O(u^{\beta}))$ and the second moment is $O(u^{\alpha-2})$ by (7.4). Hence,

$$\sum_{i:\,X_i>u} \log\left\{\frac{f(X_i + tn^{-1/\alpha})}{f(X_i)}\right\} = -n^{1-1/\alpha}u^{\alpha-1}tC\alpha + o_p(1). \tag{7.14}$$

Comparing (7.13) with (7.14), we deduce that $Z_{n,1}(t)/Z_{n,2}(t) \to 1$, uniformly over finite ranges of t, in probability.

Now we compare $Z_{n,2}$ with $Z_{n,3}$. We first show that

$$N\{\log F(u) - \log F(u + tn^{-1/\alpha}) + \alpha \log(1 + tn^{-1/\alpha}u^{-1})\} \to 0, \tag{7.15}$$

$$\sum_{i=2}^{N} \{\log f(X_{n,i}^* + tn^{-1/\alpha}) - \log(X_{n,i}^*) - (\alpha - 1)\log(1 + tn^{-1/\alpha}/X_{n,i}^*)\} \to 0, \tag{7.16}$$

each uniformly in probability over $-n^{1/\alpha}X_{n,1}^* < t \le M$. By (7.7) and (7.8), the left-hand side of (7.15) is

$$N\int_1^{1+tn^{1/\alpha}u^{-1}} \{\eta_0(uy) - \alpha\}y^{-1}\,dy = O_p(nu^{\alpha}n^{-1/\alpha}u^{-1}u^{\beta}) \to 0.$$

A similar argument gives (7.16) via (7.7), (7.8) and (7.11). Thus,

$$\frac{Z_{n,3}(t)}{Z_{n,2}(t)} = \frac{(X_{n,1}^* + tn^{-1/\alpha})^{\alpha-1}}{f(X_{n,1}^* + tn^{-1/\alpha})} \frac{f(X_{n,1}^*)}{(X_{n,1}^*)^{\alpha-1}}\{1 + o_p(1)\},$$

uniformly. But $X_{n,1}^* \to_p 0$, $X_{n,1}^* + tn^{-1/\alpha} \to_p 0$, uniformly in t, hence, $Z_{n,3}(t)/Z_{n,2}(t) \to 1$, uniformly over the required range of t, in probability.

The argument for $Z_{n,4}/Z_{n,3}$ is almost identical, since by Theorem 7.2 we have $\tilde{\alpha} - \alpha = O_p(N^{-1/2}) = O_p(u^{\beta})$. We deduce that

$$\frac{Z_{n,4}(t)}{Z_{n,3}(t)} = (X_{n,1}^* + tn^{-1/\alpha})^{\tilde{\alpha}-\alpha}(1 + o_p(1)),$$

from which the result follows. □

8. Estimating tail probabilities. So far, our results have been concerned with estimating parameters, rather than with the tail distribution function itself. In this section we consider the latter problem. There is also the closely related inverse problem of estimating quantiles, but we shall not consider that explicitly.

Suppose we have a total of n observations from the unknown F. Fix a threshold, and let N denote the random number of observations above the threshold. The threshold is denoted u_N, the suffix N being for consistency with earlier notation. Suppose we wish to estimate $1 - F(u_N + y_N)$, for some $y_N \ge 0$. The estimator we propose to use is N/n as an estimator of $1 - F(u_N)$, and the

GPD for $\{1 - F(u_N + y_N)\}/\{1 - F(u_N)\}$. Thus,

$$1 - \hat{F}(u_N + y_N) = n^{-1}N\left(1 - \hat{k}_N y_N/\hat{\sigma}_N\right)^{1/\hat{k}_N},$$

where $\hat{\sigma}_N$, $\hat{k}_n$ are estimators of the GPD parameters. Where there is no ambiguity about N, we shall drop the suffix and write $\hat{\sigma}$, $\hat{k}$.

Suppose σ_N, k are defined and that

$$N^{1/2}\begin{bmatrix} \hat{\sigma}/\sigma_N - 1 \\ \hat{k} - k \end{bmatrix} \to_d \mathscr{N}[\mathbf{b}, S], \tag{8.1}$$

for some vector $\mathbf{b}$ and matrix S. It is not necessary that $\hat{\sigma}$, $\hat{k}$ be the maximum-likelihood estimator. Suppose $z > 0$ is fixed and $y_N = \sigma_N(1 - z)/k$.

Write $\{1 - \hat{F}(u_N + y_N)\}/\{1 - \hat{F}(u_N)\}$ as $\exp\{-h(\hat{\sigma}, \hat{k}, y_N)\}$, where $h(\sigma, k, y) = -k^{-1}\log(1 - ky/\sigma)$. Define $\mathbf{c}^T$ to be $(\sigma\,\partial h/\partial\sigma, \partial h/\partial k)$ evaluated at (σ_N, k, y_N). Thus,

$$\mathbf{c}^T = \left(-k^{-1}(z^{-1} - 1), k^{-2}\log z + k^{-2}(z^{-1} - 1)\right). \tag{8.2}$$

Then

$$N^{1/2}\left\{h\left(\hat{\sigma}, \hat{k}, y_N\right) - h(\sigma_N, k, y_N)\right\} \to_d \mathscr{N}(\mathbf{c}^T\mathbf{b}, \mathbf{c}^T S\mathbf{c}). \tag{8.3}$$

Suppose

$$z^{-1/k}\{1 - F(u_N + y_N)\}/\{1 - F(u_N)\} = 1 + N^{-1/2}b' + o(N^{-1/2}), \tag{8.4}$$

for some b' depending on z. We also have

$$N^{1/2}\left[n^{-1}N\{1 - F(u_N)\}^{-1} - 1\right] \to_d \mathscr{N}(0,1), \tag{8.5}$$

which follows directly from the binomial distribution of N. Writing $\{1 - \hat{F}(u_N + y_N)\}/\{1 - F(u_N + y_N)\}$ as

$$\frac{N}{n\{1 - F(u_N)\}} z^{1/k} \frac{1 - F(u_N)}{1 - F(u_N + y_N)} \exp\left\{-h\left(\hat{\sigma}, \hat{k}, y_N\right) + h(\sigma_N, k, y_N)\right\}$$

and applying (8.3)–(8.5), we obtain

$$N^{1/2}\left\{\frac{1 - \hat{F}(u_N + y_N)}{1 - F(u_N + y_N)} - 1\right\} \to_d \mathscr{N}(-b' - \mathbf{c}^T\mathbf{b}, 1 + \mathbf{c}^T S\mathbf{c}). \tag{8.6}$$

This is the key result. Note that it is formulated in terms of relative error rather than absolute error, which we believe to be appropriate for applications. Davis and Resnick (1984) proved the consistency of their tail estimators uniformly in $y_N \geq 0$, but from the point of view of absolute error.

To illustrate (8.6), consider the limit law Φ_α. We assume SR2 and all the other conditions required for Theorem 3.2. Then the bias term $-b' - \mathbf{c}^T\mathbf{b}$ turns out to be

$$\begin{aligned} -\mu(\alpha - \rho)h_\rho(z) &+ \mu(\alpha + 1)(\alpha + 1 - \rho)^{-1} \\ &\times\left\{(z^{-1} - 1)(2 + \alpha)\rho + (1 + \rho)\alpha\log z\right\}, \end{aligned}$$

which vanishes if ρ is 0 or -1. The variance $\mathbf{c}^T S\mathbf{c}$ may be similarly calculated, but we shall not write it out.

The idea of using Hill's estimator to estimate the tail of the distribution, as advocated by Davis and Resnick (1984), may be brought within this framework by defining $\hat{\sigma} = u/\bar{\alpha}$, $\hat{k} = -1/\bar{\alpha}$, where $\bar{\alpha}$ is Hill's estimator of α. As a consequence of (4.2), (8.1) holds with

$$\mathbf{b} = \begin{bmatrix} \mu \\ -\mu/\alpha \end{bmatrix}, \qquad S = \begin{bmatrix} 1 & -1/\alpha \\ -1/\alpha & 1/\alpha^2 \end{bmatrix}.$$

The fact that S is now singular makes no difference to (8.6). In this case $-b' - \mathbf{c}^T\mathbf{b} = -\mu(\alpha - \rho)h_\rho(z) + \alpha\mu \log z$, which again vanishes if $\rho = 0$.

It is of some interest to extend these results to allow $z = z_N \to \infty$, thus providing a statistical counterpart to Anderson's (1978, 1984) theory of large deviations in extreme value theory. If $\hat{\sigma}$, $\hat{k}$ are the maximum-likelihood estimators using the GPD, then $-b' - c^T b \sim \mu\alpha(\alpha + 1)(1 + \rho)(\alpha + 1 - \rho)^{-1}\log z_N$ if $\rho < 0$, and $\mathbf{c}^T S\mathbf{c} \sim \alpha^2(\alpha + 1)^2 \log z_N$. Our result is

THEOREM 8.1. *Suppose L satisfies* SR2 *and that the other conditions of Theorem* 3.2 *hold. Suppose*

$$z_N \to +\infty, \qquad N^{-1/2}\log z_N \to 0, \tag{8.7}$$

$$z_N^{-s\rho}\phi(u_N z_N^s)/\phi(u_N) \to 1, \quad \text{for } 0 \le s \le 1. \tag{8.8}$$

Then

$$\frac{N^{1/2}}{\log z_N}\left[\frac{1 - \hat{F}(u_N z_N)}{1 - F(u_N z_N)} - 1\right] \to_d \mathcal{N}(\nu, \tau^2),$$

where ν is 0 *for $\rho = 0$, $\mu\alpha(\alpha + 1)(1 + \rho)(\alpha + 1 - \rho)^{-1}$ for $\rho < 0$ and $\tau^2 = \alpha^2(\alpha + 1)^2$.*

REMARK. Condition (8.7) is necessary as well as sufficient for the result, and this effectively answers the question, "How far can we extrapolate into the tails?" from the point of view of controlling relative error. Condition (8.8) is a form of a "super slow variation" (SSV) condition on $x^{-\rho}\phi(x)$. See Anderson (1978, 1984) and Goldie and Smith (1987) for more information about SSV. Our condition differs from Anderson's in that the SSV is imposed on the remainder function rather than on L itself. Note that, in simple cases, we often have $\phi(x) = x^\rho$, in which case (8.8) is automatic. Some other aspects of large deviations have been studied by Smith and Weissman (1987).

PROOF OF THEOREM 8.1. Let $y_N = \sigma_N(1 - z_N)/k$. By analogy with (8.3)–(8.6), the result follows if we can establish

$$\begin{aligned} &N^{1/2}(\log z_N)^{-1}\left[\exp\{-h(\hat{\sigma}, \hat{k}, y_N) + h(\sigma_N, k, y_N)\} - 1\right] \\ &\qquad \to_d \mathcal{N}\left[\mu\alpha(\alpha + 1)(1 + \rho)(\alpha + 1 - \rho)^{-1}, \alpha^2(\alpha + 1)^2\right], \end{aligned} \tag{8.9}$$

$$\begin{aligned} z_N^{-1/k}&\{1 - F(u_N z_N)\}/\{1 - F(u_N)\} \\ &= L(u_N z_N)/L(u_N) \\ &= \begin{cases} 1 + o(N^{-1/2}\log z_N), & \rho < 0, \\ 1 + \mu\alpha N^{-1/2}\log z_N + o(N^{-1/2}\log z_N), & \rho = 0, \end{cases} \end{aligned} \tag{8.10}$$

$$N^{1/2}(\log z_N)^{-1}\left[n^{-1}N\{1 - F(u_N)\}^{-1} - 1\right] \to_p 0. \tag{8.11}$$

Equation (8.9) follows from the same argument that led to (8.3), given (8.7). (8.11) is trivial. To show (8.10), use the representations

$$L(x)=\begin{cases} C\{1+c\rho^{-1}\phi(x)+o(\phi(x))\}, & \rho<0,\\ \exp\left[C+o(\phi(x))+\int_1^x\{c+o(1)\}\phi(t)t^{-1}\,dt\right], & \rho=0\end{cases}$$

[Goldie and Smith (1987)]. In the case $\rho<0$, (8.10) follows at once via (8.8). In the case $\rho=0$, (8.8) and monotonicity of ϕ give

$$\int_{u_N}^{u_N z_N}\phi(t)t^{-1}\,dt\sim\phi(u_N)\log z_N,$$

from which (8.10) follows easily.

If Hill's estimator is used in place of the GPD, then Theorem 8.1 remains valid with $\nu=0$ $(\rho=0)$, $\nu=\mu\alpha$ $(\rho\neq0)$ and $\tau^2=\alpha^2$. Note the similarity with the result in Sections 3 and 4 for the estimation of α.

We briefly treat the limit law Ψ_α. First, assume (5.3) for $\alpha>2$. This defines $\mathbf{b}$ and S in (8.1). In this case the bias $-b'-\mathbf{c}^T\mathbf{b}$ is

$$-\mu(\alpha-\rho)h_\rho(z^{-1})-\mu(\alpha-1)(\alpha-1-\rho)^{-1}$$
$$\times\left[(2-\alpha)\rho(z^{-1}-1)+\alpha(1-\rho)\log z\right].$$

Again, this vanishes when $\rho=0$. The same result is valid for $\alpha=2$, the only qualitative difference being that S is singular. Now suppose $\alpha<2$. Defining $\tilde\alpha$ as in Section 7, we have $N^{1/2}(\tilde\alpha-\alpha)\to_d\mathscr{N}(-\mu\alpha,\alpha^2)$. (This is true without making the detailed assumptions in Section 7.) Let $\tilde\theta$ be any estimator of θ such that $n^{1/\alpha}(\tilde\theta/\theta_N-1)$ converges. For example, we could take $\tilde\theta$ to be the sample maximum. Defining $\hat k=1/\tilde\alpha$ and $\hat\sigma=\tilde\theta/\tilde\alpha$ we have (8.1) with

$$\mathbf{b}=\begin{bmatrix}\mu\\ \mu/\alpha\end{bmatrix},\qquad S=\begin{bmatrix}1 & 1/\alpha\\ 1/\alpha & 1/\alpha^2\end{bmatrix}.$$

Then $-b'-\mathbf{c}^T\mathbf{b}=-\mu(\alpha-\rho)h_\rho(z^{-1})-\alpha\mu\log z$, and (8.6) again holds. □

9. Limit law Λ. We now treat the case where F is in the domain of attraction of Λ. As in Sections 1–5, we consider procedures based on N excesses $Y_1,\dots,Y_N$ over a threshold u (or u_N) and use the GPD as an approximation for F_u, the common distribution function of $Y_1,\dots,Y_N$. We consider three procedures.

(i) Estimate the GPD parameters σ and k by maximum-likelihood estimators $\hat\sigma_N$, $\hat k_N$.

(ii) Assume $k=0$, i.e., approximate $F_u(y)$ by the exponential distribution $1-\exp(-y/\sigma)$ and estimate σ by $\overline{Y}_N$, the sample mean of $Y_1,\dots,Y_N$.

(iii) Estimate σ and k by $\bar\sigma_N=u/\bar\alpha_n$, $\bar k_N=-1/\bar\alpha_N$, where $\bar\alpha_N$ is Hill's estimator (1.4).

Procedure (i) is the basic procedure studied in Sections 1–5. Procedure (ii) is included because the exponential distribution is already widely used by hydrologists [see Smith (1984) for further discussions] and because the exponential distribution is the appropriate form of the GPD when F is in the domain of attraction of Λ. Procedure (iii) is the Davis–Resnick (1984) method: Although they did not formulate it in this way, it is easy to see that the tail estimates, so constructed, are the same as those of Davis and Resnick. They proved, under conditions broad enough to include most practical examples, that the tail estimates constructed in this way are consistent. It is therefore of interest to determine the asymptotic properties in greater detail.

The plan of this section is as follows. Section 9.1 gives technical preliminaries about the domain of attraction of Λ. In Section 9.2, asymptotic results about the three estimators of σ and k are given. In Section 9.3, these results are extended, in the manner of Section 8, to the estimators of the tail distribution function. Section 9.4 gives examples and discussion. All proofs in this section are deferred to Section 9.5.

9.1. *Technical preliminaries.* If F is in the domain of attraction of Λ, then there exists a representation

$$1 - F(x) = c(x)\exp\left\{-\int_{-\infty}^{x}\frac{dt}{\phi(t)}\right\}, \qquad x < x_0, \tag{9.1}$$

where $c(x) \to 1$ as $x \to x_0 \le \infty$, ϕ is a positive differentiable function and $\phi'(x) \to 0$. This was proved by Balkema and de Haan (1972) extending de Haan (1970).

PROPOSITION 9.1. *Suppose* (9.1) *holds and*

$$\phi'(u + y\phi(u))/\phi'(u) \to 1, \quad \text{as } u \to x_0, \text{ uniformly over } 0 \le y \le -K\log|\phi'(u)|, \quad \text{for some } K > 1, \tag{9.2a}$$

$$c(u) - 1 \sim s\phi'(u), \quad \text{as } u \to x_0 \text{ for finite } s. \tag{9.2b}$$

Then, for each $\delta > 0$, *there exist* $u_\delta \le x_0$ *and a function* ε_u *tending to* 0 *as* $u \to x_0$, *such that for* $u \ge u_\delta$, $0 \le y < x_0 - u$,

$$\left|1 - F_u(y\phi(u)) - e^{-y}\{1 + \phi'(u)y^2/2\}\right| < \varepsilon_u\phi'(u)\min(1, y^{-\delta}). \tag{9.3}$$

PROPOSITION 9.2. *Suppose* (9.1) *holds and*

$$\phi''(u + y\phi(u))/\phi''(u) \to 1, \quad \text{as } u \to x_0, \text{ uniformly over } 0 \le y \le -K\log|\phi'(u)|, \quad \text{for some } K < 2, \tag{9.4a}$$

$$c(u) - 1 \sim s\left\{(\phi'(u))^2 + |\phi(u)\phi''(u)|\right\}, \quad \text{as } u \to x_0 \text{ for finite } s, \tag{9.4b}$$

$$\phi(u)\phi''(u)\log|\phi'(u)|/\phi'(u) \to 0, \quad \text{as } u \to x_0. \tag{9.4c}$$

Then, for each $\delta > 0$, *there exist* $u_\delta < 0$ *and a function* ε_u *tending to* 0 *as* $u \to x_0$, *such that for* $u \ge u_\delta$, $0 \le y \le x_0 - u$,

$$\Big|1 - F_u(y\phi(u)) - e^{-y}\big[1 + y^2\phi'(u)/2 - y^3\{2(\phi'(u))^2 - \phi(u)\phi''(u)\}/6 + y^4(\phi'(u))^2/8\big]\Big| < \varepsilon_u\left[(\phi'(u))^2 + |\phi(u)\phi''(u)|\right]\min(1, y^{-\delta}). \tag{9.5}$$

REMARK. Equation (9.2) essentially defines Cohen's (1982b) class N, as simplified by Anderson (1984). A sufficient condition for (9.2a), due to Cohen, is that ϕ' be eventually one-signed with either $|\phi'(x)|$ (in the case $x_0 = \infty$) or $|\phi'(x_0 - x^{-1})|$ (in the case $x_0 < \infty$), regularly varying at $x = \infty$. The additional conditions in (9.4) are similar to those of Cohen's Theorem 9. Proposition 9.1 is due, with minor variations, to C. W. Anderson. A proof of Proposition 9.2 is given in Section 9.5..

9.2. *Estimation of σ and k.* We follow the scheme outlined in Section 2, letting $g(\cdot;\sigma,k)$ denote the GPD density. For fixed u set $\sigma = \phi(u)$, $k = 0$ so that $-\sigma(\partial/\partial\sigma)\log g(Y;\sigma,k) = 1 - Y/\sigma(u)$, $-(\partial/\partial k)\log g(Y;\sigma,k) = Y^2/2\phi^2(u) - Y/\phi(u)$. Suppose $Y \sim F_u$. If (9.4) holds, then by (9.5) we have as $u \to x_0$,

$$\begin{aligned}E\{Y/\phi(u)\} &= 1 + \phi'(u) + (\phi'(u))^2 + \phi(u)\phi''(u)\\ &\quad + o\{(\phi'(u))^2 + |\phi(u)\phi''(u)|\},\\ E\{Y^2/\phi^2(u)\} &= 2 + 6\phi'(u) + 14(\phi'(u))^2\\ &\quad + 8\phi(u)\phi''(u) + o\{(\phi'(u))^2 + |\phi(u)\phi''(u)|\}.\end{aligned}$$

Thus, defining $\mathbf{U}_N$ as in Section 2,

$$N^{-1}E\{\mathbf{U}_N(\sigma,k)\} = \begin{bmatrix} -\phi'(u) - (\phi'(u))^2 - \phi(u)\phi''(u) \\ 2\phi'(u) + 6(\phi'(u))^2 + 3\phi(u)\phi''(u)\end{bmatrix} + o\{(\phi'(u))^2 + |\phi(u)\phi''(u)|\}.$$

If only (9.2) holds instead of (9.4), the same expression is valid up to the term in $\phi'(u)$.

THEOREM 9.3. *Suppose* (9.1) *and* (9.2) *hold and* $N \to \infty$, $u_N \to x_0$,

$$N^{1/2}\phi'(u_N) \to \mu. \tag{9.6}$$

Then

$$N^{1/2}\begin{bmatrix}\hat{\sigma}_N/\phi'(u_N) - 1\\ \hat{k}_N\end{bmatrix} \to_d \mathcal{N}\left[\begin{bmatrix}0\\ -\mu\end{bmatrix}, \begin{bmatrix}2 & 1\\ 1 & 1\end{bmatrix}\right], \tag{9.7}$$

$$N^{1/2}\{\bar{Y}_N/\phi(u_N) - 1\} \to_d \mathcal{N}(\mu, 1). \tag{9.8}$$

If further

$$N^{1/2}\phi(u_N)/u_N \to \nu, \tag{9.9}$$

then

$$N^{1/2}\begin{bmatrix}\bar{\sigma}_N/\phi(u_N) - 1\\ \bar{k}_N\end{bmatrix} \to_d \mathcal{N}\left[\begin{bmatrix}\mu - \nu\\ -\nu\end{bmatrix}, \begin{bmatrix}1 & 0\\ 0 & 0\end{bmatrix}\right]. \tag{9.10}$$

In the case of $(\hat{\sigma}_N, \hat{k}_N)$, we can achieve a finer result by lowering the sequence of thresholds:

THEOREM 9.4. *Suppose* (9.1') *and* (9.4) *hold and* $N \to \infty$, $u_N \to x_0$,

$$N^{1/2}(\phi'(u_N))^2 \to \mu_1, \qquad N^{1/2}\phi(u_N)\phi''(u_N) \to \nu_1. \tag{9.11}$$

Define $b_N = N^{1/2}\phi'(u_N)$. *Then*

$$N^{1/2}\begin{bmatrix} \hat{\sigma}_N/\phi(u_N) - 1 \\ \hat{k}_N \end{bmatrix} + \begin{bmatrix} 0 \\ b_N \end{bmatrix} \to_d \mathcal{N}\left[\begin{bmatrix} -4\mu_1 - \nu_1 \\ -5\mu_1 - 2\nu_1 \end{bmatrix}, \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix}\right].$$

REMARK. Many well-known distributions satisfy (9.4), including the normal and lognormal and the gamma and Weibull distributions with shape parameter $\neq 1$. A curious exception, however, is the exponential distribution; in this case $\phi' \equiv 0$ and neither (9.2) nor (9.4) is satisfied. Because of this, Cohen (1982b) [following Anderson (1976)] defined a second class of distributions which he called class E. It is possible to prove similar results to Theorems 9.3 and 9.4 for class E as well, but we shall not go in the details of this.

9.3. *Estimation of tail probabilities*. It is possible to make some comparisons among procedures (i)–(iii) using the results of Section 9.2, but this is not very satisfactory because it is not clear, at this stage, what we are estimating. We therefore proceed directly to the estimation of tail probabilities. For this we follow Section 8, pointing out the changes that are needed. We consider estimators of $F(u_N + y)$, $y \geq 0$, corresponding to our three procedures for estimating σ and k:

$$F_{(\mathrm{i})}(u_N + y) = 1 - n^{-1}N(1 - \hat{k}_N y/\hat{\sigma}_N)^{1/\hat{k}_N},$$

$$F_{(\mathrm{ii})}(u_N + y) = 1 - n^{-1}N\exp(-y/\bar{Y}_N),$$

$$F_{(\mathrm{iii})}(u_n + y) = 1 - n^{-1}N(1 - \bar{k}_N y/\bar{\sigma}_N)^{1/\bar{k}_N} = 1 - n^{-1}N(1 + y/u_N)^{-\bar{\alpha}_N}.$$

Assume (8.1) holds with $k = 0$, where $(\hat{\sigma}, \hat{k})$ is any of $(\hat{\sigma}_N, \hat{k}_N)$, $(\bar{y}_n, 0)$ and $(\bar{\sigma}_N, \bar{k}_N)$. Assume $y = y_N$ is given by $y_N = \sigma_N z$ for fixed $z > 0$. Then (8.2) must be replaced by

$$c^T = (-z, z^2/2). \tag{9.12}$$

Equations (8.3) and (8.5) hold without change. In place of (8.4) we assume

$$e^z\{1 - F(u_N + y_N)\}/\{1 - F(u_N)\} = 1 + N^{-1/2}b' + o(N^{-1/2}), \tag{9.13}$$

for some b' depending on z. Then (8.6) holds.

The following theorem is proved by a routine working-out of this procedure.

THEOREM 9.5. *Suppose the conditions of Theorem* 9.3 *are satisfied. Then for fixed* $z > 0$,

$$N^{1/2}\left\{\frac{1 - F_{(i)}(u_N + z\phi(u_N))}{1 - F(u_N + z\phi(u_N))} - 1\right\} \to_d \mathcal{N}\left[0, 1 + 2z^2 - z^3 + \frac{z^4}{4}\right],$$

$$N^{1/2}\left\{\frac{1 - F_{(ii)}(u_N + z\phi(u_N))}{1 - F(u_N + z\phi(u_N))} - 1\right\} \to_d \mathcal{N}\left[\mu z - \frac{\mu z^2}{2}, 1 + z^2\right],$$

$$N^{1/2}\left\{\frac{1 - F_{(iii)}(u_N + z\phi(u_N))}{1 - F(u_N + z\phi(u_N))} - 1\right\} \to_d \mathcal{N}\left[(\mu - \nu)\left(z - \frac{z^2}{2}\right), 1 + z^2\right].$$

We would also like to study $F_{(i)}$ when the conditions of Theorem 9.4 are satisfied. The changes we need to make are as follows. Defining

$$\mathbf{b} = \begin{bmatrix} -4\mu_1 - \nu_1 \\ -5\mu_1 - 2\nu_1 \end{bmatrix}, \qquad S = \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix},$$

(8.3) must be replaced by

$$(9.14)\quad N^{1/2}\{h(\hat{\sigma}_N, \hat{k}_N, y_N) - h(\sigma_N, 0, y_N) + b_N z^2/2\} \to_d \mathcal{N}(\mathbf{c}^T\mathbf{b}, \mathbf{c}^T S\mathbf{c}).$$

In place of (8.4), we have

$$(9.15)\quad \begin{aligned} &e^z\{1 - F(u_N + y_N)\}/\{1 - F(u_n)\} \\ &\quad = 1 + N^{-1/2}(z^2 b_N/2 + b') + o(N^{-1/2}), \end{aligned}$$

where $b' = -z^3(2\mu_1 - \nu_1)/6 + z^4\mu_1/8$. Putting (9.14), (9.15) and (8.5) together, the terms involving b_N cancel and we again get (8.6). Thus we have

THEOREM 9.6. *Under the conditions of Theorem* 9.4, *for* $z > 0$,

$$\begin{aligned} &N^{1/2}\left\{\frac{1 - F_{(i)}(u_N + z\phi(u_N))}{1 - F(u_N + z\phi(u_N))} - 1\right\} \\ &\quad \to_d \mathcal{N}\left[-z(4\mu_1 + \nu_1) + \frac{z^2(5\mu_1 + 2\nu_1)}{2} + \frac{z^3(2\mu_1 - \nu_1)}{6} - \frac{z^4\mu_1}{8},\right. \\ &\qquad\qquad \left. 1 + 2z^2 - z^3 + \frac{z^4}{4}\right]. \end{aligned}$$

9.4. *Discussion and examples.* The results of this section may be used to determine both the asymptotically optimal threshold and the best rate of convergence of the estimators, for each of the three procedures considered. Suppose we have a total sample size n and threshold v_n. Then the number of exceedances N_n satisfies $N_n \sim n\{1 - F(v_n)\}$ in probability, provided $N_n \to \infty$. Identifying N with N_n and u_N with v_n, Theorems 9.3–9.6 hold conditionally on the sequence $\{N_n\}$, hence, also unconditionally. The conditions (9.6), (9.9) and (9.11) are valid with u_n replaced by v_n, N replaced by $n\{1 - F(v_n)\}$. Assuming (9.4), we can then say

$$(9.6) \Rightarrow \text{error in } F_{(\text{ii})} \text{ is } O_p\left(N_n^{-1/2}\right),$$

$$(9.6) \text{ and } (9.9) \Rightarrow \text{error in } F_{(\text{iii})} \text{ is } O_p\left(N_n^{-1/2}\right),$$

$$(9.11) \Rightarrow \text{error in } F_{(\text{i})} \text{ is } O_p\left(N_n^{-1/2}\right).$$

These error rates cannot be improved unless the bias term is 0. The only nontrivial case is when $\mu = \nu$ in (9.9), a possibility we shall discuss in a moment.

For example, if F is standard normal, then (9.1) holds with $c(x) = 1$, $\phi(x) = x^{-1} - x^{-3} + 3x^{-5} - \cdots$ which satisfies (9.4) with $\phi'(x) \sim -x^{-2}$. For (9.6) we require

$$-n^{1/2}\{1 - F(v_n)\}^{1/2} v_n^{-2} \to \mu, \qquad -\infty < \mu \le 0,$$

which, except in the case $\mu = 0$, implies $N_n \sim (\log n)^2$. Thus, the optimal rate of convergence using procedures (ii) and (iii) is $O\{(\log n)^{-1}\}$. For procedure (i), however, we can do better by using condition (9.11) in place of (9.6); the rate of convergence is then $O\{(\log n)^{-2}\}$. There is a parallel here with Cohen's (1982a) results on penultimate approximations. If $\Lambda(x)$ is used as an approximation to $F^n(a_n x + b_n)$, for suitable a_n and b_n, then the best rate of convergence is $O\{(\log n)^{-1}\}$, but by using a different (penultimate) extreme value approximation, this can be improved to $O\{(\log n)^{-2}\}$. The analogous comparison is between fixing k at its limiting value 0 [procedure (ii)] and allowing k to vary with n [procedure (i)].

For the gamma distribution with density $x^{\alpha-1}e^{-x}/\Gamma(\alpha)$, $\alpha \ne 1$, we have (9.1) with $c(x) = 1$, $\phi(x) = 1 + (\alpha - 1)x^{-1} + (\alpha - 1)(\alpha - 2)x^{-2} + \cdots$. The best rates of convergence for procedures (i), (ii) and (iii) are $(\log n)^{-3}$, $(\log n)^{-2}$ and $(\log n)^{-1}$, respectively. In this case (iii) does worse because (9.6) $\not\Rightarrow$ (9.9). Again, the rates of convergence for (ii) and (iii) are the same as for the ultimate and penultimate approximations in extreme value theory [Cohen (1982b)].

The connections with Cohen's results are not accidental. Suppose the conditions of Cohen's (1982b) Theorem 2 and our (9.2) are satisfied, with (9.2a) extended to $K \log|\phi'(u)| \le y \le 0$, $K > 2$, (this is automatic if Cohen's Lemma 1 is satisfied). Cohen defines b_n so that $n\{1 - F(b_n)\} \to 1$, $a_n = \phi(b_n)$ and shows that $F^n(a_n x + b_n) - \Lambda(x) = O\{\phi'(b_n)\}$, uniformly in x. Let $v_n = b_n + 2\phi(b_n)\log|\phi'(b_n)/\mu|$ for fixed μ. By expanding $\int dt/\{\phi(t)\}$ (see the proof of Proposition 9.2) and using (9.2) we may deduce that

$$\lim n^{1/2}\{1 - F(v_n)\}^{1/2}\phi'(v_n) = \lim n^{1/2}\{1 - F(v_n)\}^{1/2}\phi'(b_n) = \mu.$$

Defining $N \sim n\{1 - F(v_n)\}$ and $u_N = v_n$, Theorem 9.5 applies to procedures (i) and (ii). In particular, the relative error is $O\{\phi'(v_n)\} = O\{\phi'(b_n)\}$. Now, however, suppose Cohen's Theorem 9 and our (9.4) are satisfied, with (9.4a) extended to $K \log|\phi'(u)| \leq y \leq \theta$, $K > 4$. By defining w_n to be either $b_n + 2\phi(b_n)\log|\{\phi'(b_n)\}^2/\mu_1|$ or $b_n + 2\phi(b_n)\log|\phi(b_n)\phi''(b_n)/\nu_1|$ and using Theorem 9.6, the relative error in $F_{(i)}$ with respect to $\{w_n\}$ is found to be $O[\{\phi'(b_n)\}^2 + |\phi(b_n)\phi''(b_n)|]$ matching Cohen's Theorem 9. Thus, it is true quite generally that the best rates of convergence for procedures (ii) and (i) are the same as Cohen's rates of convergence for the ultimate and penultimate approximations. In particular, columns 2 and 3 of Cohen's Table 1 may be interpreted in this way.

Suppose Cohen's (1982b) Lemma 1 is satisfied, i.e., either $|\phi'(x)|$ (in the case $x_0 = \infty$) or $|\phi'(x_0 - x^{-1})|$ (in the case $x_0 < \infty$) is in $\mathbb{R}_\rho$, $\rho \leq 0$. If $x_0 = \infty$ and either (a) $\rho > -1$ or (b) $\rho < -1$, $\phi(+\infty) = 0$, then $x\phi'(x)/\phi(x) \to \rho + 1 \neq 0$ by de Haan (1970), Theorem 1.2.1. In this case, (9.6) $\Rightarrow$ (9.9) with $\nu = \mu/(1+\rho)$. Thus, whenever Theorem 9.5 applies to procedure (ii) it also applies to (iii). In this sense, (iii) is as good as (ii) and, in the case $\rho = 0$ (so $\mu = \nu$), is better than (ii). Note that $\rho = 0$ corresponds to ϕ' decreasing as slowly as possible, i.e., the most long-tailed case in the domain of attraction of Λ. On the other hand, if $\rho = -1$ or $\rho < -1$ and $\phi(+\infty) > 0$, then $x\phi'(x)/\phi(x) \to 0$ by de Haan (1970), and (9.9) is a strictly stronger condition than (9.6). In the same sense as before, (iii) is now a worse procedure than (ii). The gamma distribution is of this form. If $x_0 < \infty$, then $\phi'(x)/\phi(x) \to \infty$ as $x \to x_0$, so (9.6) $\Rightarrow$ (9.9) with $\nu = 0$. In this case, (ii) and (iii) are equally good from the point of view of rates of convergence, but we would still expect (i) to do better in most cases. The one case where procedure (ii) appears to be the best of the three procedures is when dealing with class E, which was mentioned briefly in Section 9.2 but not pursued. This is, however, rather a small class and is of much narrower applicability than class N.

9.5. *Proofs.*

Proof of Proposition 9.2. First note that (9.4a) implies

$$\phi'(u + y\phi(u))/\phi'(u) \to 1, \tag{9.16}$$

$$\phi(u + y\phi(u))/\phi(u) \to 1, \tag{9.17}$$

each uniformly over $0 \leq y \leq -K \log|\phi'(u)|$. For (9.16), write $\phi'(u + y\phi(u))/\phi'(u) = 1 + \{y\phi(u)\phi''(u)/\phi'(u)\}\{1 + o(1)\}$ and use (9.4c). (9.17) is similar.

Assume $0 \leq y \leq -K \log|\phi'(u)| = z_u$, say. Using (9.4b) we have

$$c(u + y\phi(u))/c(u) = 1 + o\left\{(\phi'(u))^2 + |\phi(u)\phi''(u)|\right\}, \tag{9.18}$$

uniformly in y. Writing $\Psi(x) = 1/\phi(x)$, we have

$$\int_u^{u+y\phi(u)} \frac{dt}{\phi(t)} = \int_0^y \frac{\Psi(u + s\phi(u))}{\Psi(u)} ds$$

$$= y + \frac{y^2}{2} \frac{\Psi'(u)}{\Psi^2(u)} + \frac{y^3\Psi''(u + \theta\phi(u))}{6\Psi^3(u)},$$

where $0 \le \theta \equiv \theta(u, y) \le y$. Using (9.4a), (9.16) and (9.17), this is

$$y - \frac{y^2}{2}\phi'(u) + \frac{y^3}{6}\{2(\phi'(u))^2 - \phi(u)\phi''(u)\} + o\{(1 + y^3)R(u)\},$$

uniformly in $0 \le y \le z_u$, where $R(u) = \{\phi'(u)\}^2 + |\phi(u)\phi''(u)|$. Thus, we have

$$\exp\left\{-\int_u^{u+y\phi(u)} \frac{dt}{\phi(t)}\right\} = \exp\{-y + r(u, y)\}, \tag{9.19}$$

where $r \to 0$ as $u \to x_0$, uniformly in $y \le z_u$. Thus, $\exp\{r(u, y)\} = 1 + r + r^2(1 + o(1))/2$, uniformly, and so

$$\begin{aligned}\exp\{r(u, y)\} &= 1 + \frac{y^2}{2}\phi'(u) - \frac{y^3}{6}\{2(\phi'(u))^2 - \phi(u)\phi''(u)\}\\ &\quad + \frac{y^4}{8}(\phi'(u))^2 + o\{(1 + y^4)R(u)\},\end{aligned} \tag{9.20}$$

uniformly on $y \le z_u$. Combining (9.18)–(9.20), since $(1 + y^4)e^{-y} = o(y^{-\delta})$ as $y \to \infty$, gives the result for $0 \le y \le z_u$.

To extend this to $z_u < y < \infty$, we proceed as follows. Given $\varepsilon > 0$, for all large u we have $|\phi'(x)| < \varepsilon$, $x > u$; hence, $\phi(u + x\phi(u))/\phi(u) < 1 + \varepsilon x$; hence, for $y > z \ge 0$,

$$\int_{u+z\phi(u)}^{u+y\phi(u)} \frac{dt}{\phi(t)} \ge \int_z^y \frac{dt}{1 + \varepsilon x} \ge \frac{1}{\varepsilon}\log\left[\frac{\varepsilon y}{1 + \varepsilon z}\right].$$

Writing $\varepsilon = 1/\delta$ and using $c(x) \to 1$, we have

$$\frac{1 - F(u + y\phi(u))}{1 - F(u + z\phi(u))} \le Ky^{-\delta}\left(1 + \frac{z}{\delta}\right)^{\delta}, \qquad y > z \ge 0,\ u \ge u_\delta,$$

for appropriate u_δ and K. So

$$\frac{1 - F(u + y\phi(u))}{1 - F(u)} \le Ky^{-\delta}\left(1 + \frac{z_u}{\delta}\right)^{\delta} \frac{1 - F(u + z_u\phi(u))}{1 - F(u)}.$$

But $\{1 - F(u + z_u\phi(u))\}/\{1 - F(u)\} = o(z_u^4 e^{-z_u})$ and $z_u^{4+\delta}e^{-z_u} = o\{(\phi'(u))^2\}$ since $K > 2$, so the last expression is $o\{y^{-\delta}(\phi'(u))^2\}$, uniformly in $y \ge z_u$,

$u \geq u_\delta$. We also have

$$e^{-y}\left[1 + \frac{y^2}{2}\phi'(u) - \frac{y^3}{6}\{2(\phi'(u))^2 - \phi''(u)\} + \frac{y^4}{8}(\phi'(u))^2\right]$$
$$= o\{y^{-\delta}(\phi'(u))^2\},$$

uniformly in $y \geq z_u$, $u \geq u_\delta$. This completes the proof. □

PROOF OF THEOREM 9.3. Equation (9.7) is proved essentially by verifying (2.4)–(2.6). We omit all the technical details, which are similar to those in Section 3. Equation (9.8) is just the central limit theorem. By the inequality $|\log(1 + x) - x + x^2/2| < x^3/3$, $x > 0$, we have

$$N^{1/2}\{\bar{\sigma}_N/\phi(u_N) - 1\} = -N^{1/2}\{u_N\bar{k}_N/\phi(u_N) + 1\}$$
$$= N^{1/2}\sum\left\{\frac{Y_{N,i}}{\phi(u_N)} - 1\right\} - \frac{N^{1/2}\phi(u_N)}{u_N}\frac{1}{2N}\sum\left\{\frac{Y_{N,i}}{\phi(u_N)}\right\}^2$$
$$+ O_P\left[N^{1/2}\frac{\phi^2(u_N)}{u_N}\frac{1}{N}\sum\left\{\frac{Y_{N,i}}{\phi(u_N)}\right\}^3\right]$$
$$= \mathscr{N}(\mu, 1) - \nu + o_p(1).$$

This also shows that $N^{1/2}\bar{k}_N \to_p -\nu$, hence (9.10). □

PROOF OF THEOREM 9.4. $N^{-1}E\{\mathbf{U}_N\mathbf{U}_N^T\} = M + o(1)$ as usual, and Lyapunov's central limit theorem implies asymptotic normality of $N^{-1/2}\mathbf{U}_N^T + (b_N - 2b_N)$. The result then follows by another application of the argument in Section 2. □

10. Concluding remarks. In this paper we have examined in detail the properties of estimators based on the GPD and, in some cases, have compared them with competing estimators. In Section 9, in particular, we compared our procedure [procedure (i)] with two others which have been suggested, and showed that in a very wide class of cases, it has a faster rate of convergence, provided the threshold is chosen optimally. A practical advantage of the GPD is the avoidance of any need to decide a priori among the three limiting types.

Two questions which we have not discussed are whether our procedures have any optimality properties in the class of all possible statistical procedures and the practical choice of threshold. These questions have been discussed, respectively, by Hall and Welsh (1984), Hall and Welsh (1985) for the estimation of an index of regular variation. The latter paper proposes an adaptive rule for choosing the threshold, which is quite different from the goodness-of-fit procedures of Pickands (1975) and Hill (1975). Pickands' procedure is essentially to choose the threshold to minimise a goodness-of-fit statistic based on the observed and fitted distributions of the exceedances. Heuristic arguments suggest that this procedure will lead to N, the number of exceedances, the same order of magnitude as its optimal value, but multiplied by a nondegenerate random constant. There is obviously scope for further investigation on this point.

APPENDIX

Elementary properties of the GPD. From the density $g(y;\sigma,k) = \sigma^{-1}(1-ky/\sigma)^{1/k-1}$, $ky<\sigma$, we obtain

$$-\frac{\partial \log g}{\partial \sigma} = \frac{1}{\sigma k} - \frac{1}{\sigma}\left(\frac{1}{k}-1\right)\left(1-\frac{ky}{\sigma}\right)^{-1},$$

$$-\frac{\partial \log g}{\partial k} = \frac{1}{k^2}\log\left(1-\frac{ky}{\sigma}\right) - \frac{1}{k}\left(\frac{1}{k}-1\right)\left\{1-\left(1-\frac{ky}{\sigma}\right)^{-1}\right\},$$

$$-\frac{\partial^2 \log g}{\partial \sigma^2} = -\frac{1}{\sigma^2 k} + \frac{1}{\sigma^2}\left(\frac{1}{k}-1\right)\left(1-\frac{ky}{\sigma}\right)^{-2},$$

$$-\frac{\partial^2 \log g}{\partial \sigma\,\partial k} = -\frac{1}{\sigma k^2} + \frac{2-k}{\sigma k^2}\left(1-\frac{ky}{\sigma}\right)^{-1} - \frac{1-k}{\sigma k^2}\left(1-\frac{ky}{\sigma}\right)^{-2},$$

$$-\frac{\partial^2 \log g}{\partial k^2} = -\frac{2}{k^3}\log\left(1-\frac{ky}{\sigma}\right) + \frac{3-k}{k^3} - \frac{2(2-k)}{k^3}\left(1-\frac{ky}{\sigma}\right)^{-1} + \frac{1-k}{k^3}\left(1-\frac{ky}{\sigma}\right)^{-2}.$$

If Y has the density $g(y;\sigma,k)$, then

$$E\left\{\left(1-\frac{kY}{\sigma}\right)^{-r}\right\} = \frac{1}{1-kr} \quad \text{(provided } kr<1\text{)},$$

$$E\left[\left\{-\log\left(1-\frac{kY}{\sigma}\right)\right\}^{s}\right] = k^s\Gamma(s+1) \quad \text{(integer } s\text{)}.$$

Thus, the Fisher information matrix M_0 is (for $k<\frac{1}{2}$)

$$\begin{bmatrix} \dfrac{1}{\sigma^2(1-2k)} & -\dfrac{1}{\sigma(1-k)(1-2k)} \\ -\dfrac{1}{\sigma(1-k)(1-2k)} & \dfrac{2}{(1-k)(1-2k)} \end{bmatrix}.$$

Acknowledgments. The original stimulus for this paper was the Symposium on Order Statistics, Quantile Processes and Extreme Values at Oberwolfach, March 1984, organised by R. D. Reiss and W. van Zwet. The bulk of the paper was written during a visit to the Australian National University supported by a Visiting Fellowship. I am grateful to Clive Anderson for sending me unpublished material containing an earlier version of Proposition 9.1 and to N. H. Bingham, Peter Hall and the referees of the first version of the paper for their helpful comments and suggestions.

REFERENCES

AKAHIRA, M. (1975). Asymptotic theory for estimation of location in nonregular cases. II. Bounds of asymptotic distributions of consistent estimates. *Rep. Statist. Appl. Res. Un. Japan Sci. Engrs.* **22** 99–115.

ANDERSON, C. W. (1976). Extreme value theory and its approximations. *Proc. Symp. Reliability Technology.* U.K. Atomic Energy Authority, Bradford.

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.

ARIYAWANSA, K. A. and TEMPLETON, J. G. C. (1983). On statistical control of optimization. *Math. Operationsforsch. Statist. Ser. Optim.* **14** 393–410.

ARIYAWANSA, K. A. and TEMPLETON, J. G. C. (1986). Structural inference for parameters of a power function distribution. Preprint, Washington State Univ., Pullman.

BALKEMA, A. A. and DE HAAN, L. (1972). On R. von Mises' condition for the domain of attraction of $\exp(-e^{-x})$. *Ann. Math. Statist.* **43** 1352–1354.

BALKEMA, A. A., GELUK, J. L. and DE HAAN, L. (1979). An extension of Karamata's Tauberian theorem and its connection with complementary convex functions. *Quart. J. Math. Oxford Ser.* (2) **30** 385–416.

BREIMAN, L., STONE, C. J. and GINS, J. D. (1979). New methods for estimating tail probabilities and extreme value distributions. Preprint, Univ. of California, Berkeley.

BREIMAN, L., STONE, C. J. and GINS, J. D. (1981). Further development of new methods for estimating tail probabilities and extreme value distributions. Preprint, Univ. of California, Berkeley.

COHEN, J. P. (1982a). The penultimate form of approximation to normal extremes. *Adv. in Appl. Probab.* **14** 324–339.

COHEN, J. P. (1982b). Convergence rates for the ultimate and penultimate approximations in extreme-value theory. *Adv. in Appl. Probab.* **14** 833–854.

COHEN, J. P. (1984). The asymptotic behaviour of the maximum likelihood estimates for univariate extremes. In *Statistical Extremes and Applications* (J. Tiago de Oliveira, ed.) 435–442. Reidel, Dordrecht.

COHEN, J. P. (1986). Large sample theory for fitting an approximate Gumbel model to maxima. *Sankhya Ser. A* **48** 372–392.

COOKE, P. J. (1979). Statistical inference for bounds of random variables. *Biometrika* **66** 367–374.

COOKE, P. J. (1980). Optimal linear estimation of bounds of random variables. *Biometrika* **67** 257–258.

COX, D. R. and HINKLEY, D. V. (1974). *Theoretical Statistics.* Chapman and Hall, London.

CSÖRGŐ, S., DEHEUVELS, P. and MASON, D. M. (1985). Kernel estimates of the tail index of a distribution. *Ann. Statist.* **13** 1050–1077.

CSÖRGŐ, S. and MASON, D. M. (1985). Central limit theorems for sums of extreme values. *Math. Proc. Cambridge Philos. Soc.* **98** 547–558.

DAVIS, R. A. and RESNICK, S. I. (1984). Tail estimates motivated by extreme value theory. *Ann. Statist.* **12** 1467–1487.

DAVISON, A. (1984). Modelling excesses over high thresholds, with an application. In *Statistical Extremes and Applications* (J. Tiago de Oliveira, ed.) 461–482. Reidel, Dordrecht.

DE HAAN, L. (1970). *On Regular Variation and Its Application to the Weak Convergence of Sample Extremes*. Mathematical Centre Tracts 32, Amsterdam.

DE HAAN, L. (1981). Estimation of the minimum of a function using order statistics. *J. Amer. Statist. Assoc.* **76** 467–469.

DE HAAN, L. and RESNICK S. I. (1980). A simple asymptotic estimate for the index of a stable distribution. *J. Roy. Statist. Soc. Ser. B* **42** 83–87.

DUMOUCHEL, W. (1983). Estimating the stable index α in order to measure tail thickness. *Ann. Statist.* **11** 1019–1036.

GALAMBOS, J. (1978). *The Asymptotic Theory of Extreme Order Statistics*. Wiley, New York.

GOLDIE, C. M. and SMITH, R. L. (1987). Slow variation with remainder: A survey of the theory and its applications. *Quart. J. Math. Oxford. Ser.* (2) **38** 45–71.

HAEUSLER, E. and TEUGELS, J. L. (1985). On asymptotic normality of Hill's estimator for the exponent of regular variation. *Ann. Statist.* **13** 743–756.

HALL, P. (1982a). On some simple estimates of an exponent of regular variation. *J. Roy. Statist. Soc. Ser. B* **44** 37–42.

HALL, P. (1982b). On estimating the endpoint of a distribution. *Ann. Statist.* **10** 556–568.

HALL, P. and HEYDE, C. C. (1980). *Martingale Limit Theory and Its Application*. Academic, New York.

HALL, P. and WELSH, A. H. (1984). Best attainable rates of convergence for estimates of parameters of regular variation. *Ann. Statist.* **12** 1079–1084.

HALL, P. and WELSH, A. H. (1985). Adaptive estimates 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.

IBRAGIMOV, I. A. and HAS'MINSKII, R. Z. (1981). *Statistical Estimation: Asymptotic Theory*. Springer, New York.

JOE, H. (1987). Estimation of quantiles of the maximum of N observations. *Biometrika*. To appear.

LAYCOCK, P. J. and PHANG, R. (1986). Block size and prediction for extreme values. Dept. Mathematics, UMIST, Manchester. Unpublished.

LEADBETTER, M. R., LINDGREN, G. and ROOTZÉN, H. (1983). *Extremes and Related Properties of Random Sequences and Processes*. Springer, New York.

MANDELBROT, B. B. (1982). *The Fractal Geometry of Nature*. Freeman, New York.

PATEL, N. R. and SMITH, ROBERT L. (1983). The asymptotic extreme value distribution of the sample minimum of a concave function under linear constraints. *Oper. Res.* **31** 789–794.

PETROV, V. V. (1975). *Sums of Independent Random Variables*. Springer, New York.

PICKANDS, J. (1975). Statistical inference using extreme order statistics. *Ann. Statist.* **3** 119–131.

ROBSON, D. S. and WHITLOCK, J. H. (1964). Estimation of a truncation point. *Biometrika* **51** 33–39.

SMITH, R. L. (1982). Uniform rates of convergence in extreme value theory. *Adv. in Appl. Probab.* **14** 600–622.

SMITH, R. L. (1984). Threshold methods for sample extremes. In *Statistical Extremes and Applications* (J. Tiago de Oliveira, ed.) 621–638. Reidel, Dordrecht.

SMITH, R. L. (1985). Maximum likelihood estimation in a class of non-regular cases. *Biometrika* **72** 67–92.

SMITH, R. L. and WEISSMAN, I. (1985). Maximum likelihood estimation of the lower tail of a probability distribution. *J. Roy. Statist. Soc. Ser. B* **47** 285–298.

SMITH, R. L. and WEISSMAN, I. (1987). Large deviations of tail estimators based on the Pareto approximation *J. Appl. Probab.* To appear.

TEUGELS, J. L. (1981). Limit theorems on order statistics. *Ann. Probab.* **9** 868–880.

TEUGELS, J. L. (1984). Extreme values in insurance mathematics. In *Statistical Extremes and Applications* (J. Tiago de Oliveira, ed.) 253–259. Reidel, Dordrecht.

WEISS, L. (1971). Asymptotic inference about a density function at the end of its range. *Naval Res. Logist. Quart.* **18** 111–114.

WEISS, L. and WOLFOWITZ, J. (1973). Maximum likelihood estimation of a translation parameter of a truncated distribution. *Ann. Statist.* **1** 944–947.

WEISSMAN, I. (1978). Estimation of parameters and large quantiles based on the k largest observations. *J. Amer. Statist. Assoc.* **73** 812–815.

WOODROOFE, M. (1972). Maximum likelihood estimation of a translation parameter of a truncated distribution. *Ann. Math. Statist.* **43** 113–122.

WOODROOFE, M. (1974). Maximum likelihood estimation of translation parameter of truncated distribution. II. *Ann. Statist.* **2** 474–488.

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF SURREY
GUILDFORD SURREY GU2 5XH
ENGLAND