The Annals of Statistics
1985, Vol. 13, No. 1, 331–341

ADAPTIVE ESTIMATES OF PARAMETERS OF REGULAR VARIATION

By Peter Hall and A. H. Welsh

Australian National University

The problem of estimating shape and scale parameters for a distribution with regularly varying tails is related to that of nonparametrically estimating a density at a fixed point, in that optimal construction of the estimators depends substantially upon unknown features of the distribution. We show how to overcome this problem by using adaptive methods. Our main results hold very generally, for a large class of adaptive estimators. Later we consider specific versions of adaptive estimators, and describe their performance both in theory and by means of simulation studies. We also examine a technique proposed by Hill (1975) for solving similar problems.

1. Introduction. This paper deals with the estimation of shape and scale parameters for a distribution with regularly varying tails. Following Hill (1975), Hall (1982) and Welsh (1985) we assume a general nonparametric model, in which the only available information is in the form of asymptotic properties of the distribution's tail. Therefore inference has to be based on extreme values from the tail of the sample. Hill (1975) and Hall (1982) examined the number of extreme values required to achieve optimal performance, and showed that in general this number depends on unknown properties of the tail. Therefore the size of the extreme subsample used to construct the estimators must itself be estimated from the sample. Our main aim in the present paper is to show that there exists a large class of adaptive, "asymptotically optimal" methods for estimating the size of the subsample of extremes. We study one of these methods in detail, first in theory and then by means of simulation studies, and show that it does indeed possess nearly optimal properties. We also examine a technique proposed by Hill (1975) for solving similar problems.

We may assume without loss of generality that the regularly varying tail is at the origin. Following Hall (1982), suppose

$$F(x) = Cx^{\alpha}[1 + Dx^{\beta} + o(x^{\beta})] \tag{1.1}$$

as $x \downarrow 0$, where $\alpha > 0$, $C > 0$, $\beta > 0$ and D is a real number which we take to be nonzero. Hill (1975) and Hall (1982) proposed the following classes of estimators of α and C: let $X_{n1} \le \cdots \le X_{nn}$ denote the ordered n-sample, let $1 \le r \le n-1$ and put

$$\hat{\alpha}_r = (\log X_{n,r+1} - r^{-1} \textstyle\sum_{i=1}^{r} \log X_{ni})^{-1} \quad \text{and} \quad \hat{C}_r = r/nX_{n,r+1}^{\hat{\alpha}_r}. \tag{1.2}$$

Hall (1982) showed that optimal performance is achieved by taking $r = r_0 \equiv \lambda n^{2\beta/(2\beta+\alpha)}$, for some $\lambda > 0$, and Hall and Welsh (1984) proved that $\hat{\alpha}_{r_0}$ and $\hat{C}_{r_0}$

Received September 1983.

AMS 1980 *subject classifications.* Primary 60G05, 60E05; secondary 62F35, 60F17.

Key words and phrases. Adaptive estimator, order statistics, regular variation, scale parameter, shape parameter, stable laws.

possess rates of convergence which are optimal in the class of all possible estimators of α and C. In view of these estimators' optimal properties and attractively simple construction, we shall take them as our basic estimates of α and C. We shall show how to construct estimates $\hat{r}_0$ of r_0, and prove that $\hat{\alpha}_{\hat{r}_0}$ and $\hat{C}_{\hat{r}_0}$ share optimal convergence rates with $\hat{\alpha}_{r_0}$ and $\hat{C}_{r_0}$. These results will be presented in Sections 3, 4 and 5. Specifically, Section 3 provides an invariance principle for the estimators $\hat{\alpha}_r$ and $\hat{C}_r$; Section 4 suggests estimates of r_0, and applies the results of Section 3 to obtain limit theorems for $\hat{\alpha}_{\hat{r}_0}$ and $\hat{C}_{\hat{r}_0}$; and Section 5 reports simulation studies of our estimators.

In slightly different circumstances, Hill (1975) proposed a simple and attractive sequential decision procedure for estimating r. In Section 2, we consider this procedure in detail and show that at least for large samples, it is not appropriate for models like (1.1). We conclude this section by giving an heuristic argument describing why the procedure fails to give satisfactory results.

Large-sample properties of $\hat{\alpha}_r$ and $\hat{C}_r$ are based on the fact that rescaled logarithms of ratios of extreme order statistics, specifically

$$A_s \equiv \alpha s \log(X_{n,s+1}/X_{ns}),$$

are approximately distributed as centered exponential variables. The approximation is very good for r close to 1, but deteriorates as r increases. Indeed, it is largely for this reason that performance of the estimators $\hat{\alpha}_r$ and $\hat{C}_r$ becomes worse as r increases beyond the optimum, r_0. Therefore we might carry out a sequence of goodness-of-fit tests on the "exponential" samples $\mathscr{S}_r = \{A_s, 1 \le s \le r\}$, $r \ge 1$, and stop when the samples start to fail the test. We could take $\hat{r}_0$ equal to the largest value of r which provides a satisfactory exponential fit. However, the deterioration of the exponential approximation is very gradual. As r increases beyond r_0, we are adding a small number of nonexponential "drops" to an "ocean" of very nearly exponential random variables A_i in $\mathscr{S}_r$. We must add a very large number before the nonexponential A_i's become sufficiently many to swamp the nearly exponential values and lead us to reject the hypothesis of exponentiality. By that time, we will have substantially overestimated r_0. This type of behaviour is described quantitatively in Section 2, where for the sake of simplicity we consider a single test, rather than a sequential test. It is easy to see that if a single test fails to pick up the fact that r is an order of magnitude too large, then the sequential test will lead us to overestimate r_0.

2. Goodness-of-fit approach. Hill (1975) suggested an approach to estimating r_0 which is based on sequentially testing appropriate functions of the observations for exponentiality. Our aim in this section is to show that Hill's method does not perform well in the case of models like (1.1). We choose a simple special case of (1.1) and treat it in detail. We stress that versions of the results below can be derived in a very general context, although with more complicated analysis.

For simplicity assume that

$$F^{-1}(u) = u(1+u) \quad \text{for} \quad 0 < u \le u_0, \tag{2.1}$$

some $u_0 > 0$. This entails $F(x) = x\{1 - x + O(x^2)\}$ as $x \downarrow 0$. Thus, $\alpha = \beta = C = 1$, and the "optimal" r in the sense of minimising integrated square error, is $r = r_0 \equiv 2^{1/3}n^{2/3}$ (see Hall, 1982). We draw an independent n-sample from this distribution and denote the order statistics by $X_{n1} \le \cdots \le X_{nn}$. For each n set

$$(2.2) \qquad Y_i = Y_i(n) = i \log(X_{n,i+1}/X_{ni}), \quad 1 \le i \le r,$$

and $\bar{Y} = r^{-1} \sum_{i=1}^{r} Y_i$. If the factor $(1 + u)$ were to be omitted from (2.1) then the variables Y_i, $1 \le i \le r$, would be independent and exponentially distributed. Hill (1975) suggested increasing r until $Y_1, \cdots, Y_r$ fail a test for exponentiality. We fix the sample size and investigate the asymptotic effect which nonexponentiality has on the procedure.

There is of course a wide variety of statistics which can be used to test exponentiality. We chose to avoid the chi-squared goodness-of-fit test so as not to become involved in the controversy over the choice of cells. Instead, we examined the well-known statistics considered by Bartholomew (1957), Proschan and Pyke (1964) and other authors. Bartholomew's results show that of the three statistics he examines, the quantity

$$(2.3) \qquad S = \bar{Y}^{-2} \sum_{i=1}^{r} Y_i^2$$

is the most sensitive to departure from exponentiality in the upper tail. The statistic S admits comparatively simple asymptotic analysis, and in our case is equivalent to the statistic $H^{(r)}$ suggested by Hill (1975, page 1172). For these reasons we concentrate on S.

The main result of this section is the following theorem.

THEOREM 2.1 *Define S as in* (2.3), *using the Y_i's given in* (2.2). *Assume $r = r(n) = o(n^{4/5})$ as $n \to \infty$, and that $r \to \infty$. Then one may write*

$$(2.4) \qquad S = S_0 + o_p(r^{1/2}),$$

where S_0 has the distribution S would have under the null hypothesis of exponentiality.

It may be proved that $r^{-1/2}(S_0 - 2r)$ is asymptotically normally distributed with finite nonzero variance; see Proschan and Pyke (1964) for a unified account of limit theory for statistics like S. The remainder term in (2.4) is of *smaller* order than $r^{1/2}$, and so for $r \sim \lambda n^{\alpha}$, with $0 < \gamma < 4/5$, the chance of rejecting the null hypothesis in an asymptotic test of exponentiality will converge to the significance level of the test. Since the optimal r is of order $n^{2/3}$, this approach to selecting r is likely to result in r of too large an order of magnitude. The test is certainly of no benefit in selecting the right multiplicative constant in the optimal formula, $r \sim \lambda n^{2/3}$.

PROOF OF THEOREM 2.1. In view of (2.1), $X_{ni} = U_{ni}(1 + U_{ni})$ for uniform order statistics U_{ni}. Moreover, by Rényi's representation of order statistics (see David, 1981, page 21), we may write

$$U_{ni} = \exp\{-\textstyle\sum_{j=1}^{n-i+1} Z_j/(n - j + 1)\}, \quad 1 \le i \le n,$$

where $Z_1, \cdots, Z_n$ are independent exponential random variables. Then we may write (2.2) as

$$Y_i = Z_{n-i+1} + i\{\log(1 + U_{n,i+1}) - \log(1 + U_{ni})\} \\ = Z_{n-i+1} + i(U_{n,i+1} - U_{ni})(1 + U^*_{ni})^{-1}, \tag{2.5}$$

where U^*_{ni} lies between U_{ni} and $U_{n,i+1}$. Let $\mathbf{Y} = (Y_1, \cdots, Y_r)$, $\bar{Y} = r^{-1} \sum_{i=1}^r Y_i$, $\mathbf{Z} = (Z_n, Z_{n-1}, \cdots, Z_{n-r+1})$ and $\bar{Z} = r^{-1} \sum_{i=1}^r Z_{n-i+1}$. Expanding $S(\mathbf{Y})$, defined in (2.3), in a Taylor expansion about $\mathbf{Z}$ and writing $S_0 \equiv S(\mathbf{Z})$, we obtain

$$S(\mathbf{Y}) = S_0 + T_{1r} + R_r,$$

where

$$T_{1r} = \sum_{i=1}^r q_i(\mathbf{Z})(Y_i - Z_{n-i+1}),$$

$$R_r = \tfrac{1}{2} \sum_{i=1}^r \sum_{j=1}^r Q_{ij}(\mathbf{Y}^*)(Y_i - Z_{n-i+1})(Y_j - Z_{n-j+1}),$$

$$q_i(\mathbf{Y}) \equiv \partial S(\mathbf{Y})/\partial Y_i = 2\bar{Y}^{-3}(Y_i\bar{Y} - r^{-1} \sum_{j=1}^r Y_j^2),$$

$$Q_{ij}(\mathbf{Y}) \equiv \partial^2 S(\mathbf{Y})/\partial Y_i \partial Y_j = 2\bar{Y}^{-4}\{3r^{-2} \sum_{k=1}^r Y_k^2 - 2r^{-1}\bar{Y}Y_i - 2r^{-1}\bar{Y}Y_j + I(i = j)\,\bar{Y}^2\}$$

and the ith element of $\mathbf{Y}^*$ lies between Y_i and Z_{n-i+1}. In view of (2.5),

$$| \bar{Y}^* - \bar{Z} | \le r^{-1} \sum_{i=1}^r i(U_{n,i+1} - U_{ni}) = O_p(r/n)$$

and

$$r^{-1/2} | (\sum_{i=1}^r Y_i^{*2})^{1/2} - (\sum_{i=1}^r Z_{n-i+1}^2)^{1/2} | \le [r^{-1} \sum_{i=1}^r \{i(U_{n,i+1} - U_{ni})\}^2]^{1/2} = O_p(r/n).$$

Therefore

$$| Q_{ij}(\mathbf{Y}^*) | = O_p(1)\{r^{-1}(1 + Y_i + Y_j + Z_{n-i+1} + Z_{n-j+1}) + I(i = j)\}$$

uniformly in $1 \le i, j \le r$, whence

$$| R_r | = O_p(1)[\sum_{i=1}^r \sum_{j=1}^r \{r^{-1}(1 + Y_i + Y_j + Z_{n-i+1} + Z_{n-j+1}) + I(i = j)\} \cdot i(U_{n,i+1} - U_{ni})j(U_{n,j+1} - U_{nj})] = O_p(r^3/n^2).$$

Thus, we have

$$S = S_0 + T_{1r} + O_p(r^3/n^2). \tag{2.6}$$

Let $T_{2r} = \sum_{i=1}^r q_i(\mathbf{Z})i(U_{n,i+1} - U_{ni})(1 + U_{n,i+1})^{-1}$. Since

$$| (1 + U^*_{ni})^{-1} - (1 + U_{n,i+1})^{-1} | \le 2(U_{n,i+1} - U_{ni}),$$

then

$$| T_{1r} - T_{2r} | \le 2 \sum_{i=1}^r | q_i(\mathbf{Z}) | i(U_{n,i+1} - U_{ni})^2 \\ = O_p(1) \sum_{i=1}^r (1 + Z_{n-i+1})i(U_{n,i+1} - U_{ni})^2 \\ = O_p(r^2/n^2). \tag{2.7}$$

From the inequality $|x^{-1}(1-e^{-x})-1| \le x$, we obtain

$$U_{n,i+1} - U_{ni} = U_{n,i+1}\{1 - \exp(-Z_{n-i+1}/i)\}$$
$$= U_{n,i+1}(Z_{n-i+1}/i)(1 + R_{1ni}),$$

where $|R_{1ni}| \le Z_{n-i+1}/i$. Furthermore, $\sup_{1\le i\le n}(n/i)U_{ni} = O_p(1)$. Thus, with

$$T_{3r} = \sum_{i=1}^{r} q_i(\mathbf{Z})U_{n,i+1}Z_{n-i+1}(1 + U_{n,i+1})^{-1}$$

and

$$T_{4r} = \sum_{i=1}^{r} q_i(\mathbf{Z})U_{n,i+1}Z_{n-i+1},$$

we have

$$\begin{aligned} |T_{2r} - T_{3r}| &= \sum_{i=1}^{r} |q_i(\mathbf{Z})| U_{n,i+1}Z_{n-i+1}^2/i \\ &= O_p(1/n)\sum_{i=1}^{r}(1 + Z_{n-i+1})Z_{n-i+1}^2 = O_p(r/n) \end{aligned} \tag{2.8}$$

and

$$\begin{aligned} |T_{3r} - T_{4r}| &\le \sum_{i=1}^{r} |q_i(\mathbf{Z})| U_{n,i+1}^2 Z_{n-i+1} \\ &= O_p(1/n^2)\sum_{i=1}^{r} i^2(1 + Z_{n-i+1})Z_{n-i+1} = O_p(r^3/n^2). \end{aligned} \tag{2.9}$$

Combining (2.6)–(2.9), we have

$$S = S_0 + T_{4r} + O_p(r^3/n^2). \tag{2.10}$$

Next we show that $T_{4r} = O_p(r^{3/2}/n)$. Note that with

$$c_{ni} = \exp\{-\textstyle\sum_{j=1}^{n-i} 1/(n-j+1)\} \le Mi/n,$$
$$U_{ni} = c_{ni}\exp\{-\textstyle\sum_{j=1}^{n-i}(Z_j - 1)/(n-j+1)\} = c_{ni}(1 + R_{2ni}),$$

where

$$\begin{aligned} |R_{2ni}| &\le \left|\sum_{j=1}^{n-i}\frac{Z_j - 1}{n-j+1}\right| \exp\left\{\sup_{1\le k\le n}\left|\sum_{j=1}^{k}\frac{Z_j - 1}{n-j+1}\right|\right\} \\ &= O_p(1)\left|\sum_{j=1}^{n-i}\frac{Z_j - 1}{n-j+1}\right| \end{aligned}$$

uniformly in i. Hence

$$|U_{ni} - c_{ni}| = O_p(1)(i/n)\left|\textstyle\sum_{j=1}^{n-i}(Z_j - 1)/(n-j+1)\right|,$$

so with $T_{5r} = \sum_{i=1}^{r} q_i(\mathbf{Z})c_{ni}Z_{n-i+1}$, we have

$$\begin{aligned} |T_{4r} - T_{5r}| &= O_p(1)\sum_{i=1}^{r}(1 + Z_{n-i+1})Z_{n-i+1}(i/n)\left|\sum_{j=1}^{n-i}\frac{Z_j - 1}{n-j+1}\right| \\ &= O_p\Big[n^{-1}\sum_{i=1}^{r} i\Big\{\sum_{j=1}^{n-i} 1/(n-j+1)^2\Big\}^{1/2}\Big] = O_p(r^{3/2}/n). \end{aligned} \tag{2.11}$$

Furthermore, since $r^{1/2}|\bar{Z} - 1| + r^{1/2}\,|r^{-1}\sum_{i=1}^{r} Z_{n-i+1}^2 - 2| = O_p(1)$, then

$$T_{5r} = \sum_{i=1}^{r}(Z_{n-i+1} - 2)Z_{n-i+1}c_{ni} + O_p(r^{3/2}n^{-1}). \tag{2.12}$$

From (2.11) and (2.12) we obtain $T_{4r} = O_p(r^{3/2}n^{-1})$, and combining this estimate with (2.10), we have $S = S_0 + O_p(r^{3/2}/n + r^3/n^2)$. If $r = o(n^{4/5})$ then the remainder here is $o_p(r^{1/2})$, which proves Theorem 2.1.

3. Invariance principles. In this section we derive invariance principles for $\hat{\alpha}_r$ and $\hat{C}_r$. These results are applied in Section 4 to yield limit theorems for $\hat{\alpha}_{\hat{r}_0}$ and $\hat{C}_{\hat{r}_0}$.

Let $H = [h_1, h_2]$, where $0 < h_1 < h_2 < \infty$, and write $r = r(h, n) = [hn^{2\rho/(2\rho+1)}]$ (the integer part of $hn^{2\rho/(2\rho+1)}$). Consider the stochastic processes

$$A_n^*(h) = n^{\rho/(2\rho+1)}(\hat{\alpha}_r - \alpha)$$

and

$$C_n^*(h) = n^{\rho/(2\rho+1)}(\log n)^{-1}(\hat{C}_r - C),$$

where $\rho = \beta/\alpha$, $h \in H$ and $\hat{\alpha}_r$ and $\hat{C}_r$ are defined in (1.2). Define

$$G(h) = W(h)/h + DC^{-\rho}\rho(1 + \rho)^{-1}h^\rho,$$

where $W(\cdot)$ is standard Brownian motion. We prove weak convergence of A_n^* and C_n^* to G in the Skorohod topology on $D[h_1, h_2]$.

THEOREM 3.1. *If* (1.1) *holds, then* $A_n^*(\cdot)$ *converges weakly to* $\alpha G(\cdot)$ *and* $C_n^*(\cdot)$ *converges weakly to* $C(2\rho + 1)^{-1}G(\cdot)$ *as* $n \to \infty$.

PROOF. We shall need

LEMMA 3.1. *Let* $\{Z_j\}$ *be independent exponential random variables. If* $r = r(h, n) = [hm]$, $h \in H$, *with* $m = m(n) \to \infty$ *and* $m/n \to 0$, *then*

$$\exp[-\rho \textstyle\sum_{j=1}^{n-r} Z_j/(n - j + 1)] = (r/n)^\rho\{1 + \Delta_1(r)\} \tag{3.1}$$

and

$$\exp\left[-\rho \sum_{j=1}^{n-r} \frac{Z_j}{n - j + 1}\right]\left\{1 - r^{-1} \sum_{i=0}^{r-1} \exp\left[-\rho \sum_{j=n-r+1}^{n-i} \frac{Z_j}{n - j + 1}\right]\right\} = \rho(1 + \rho)^{-1}(r/n)^\rho\{1 + \Delta_2(r)\}. \tag{3.2}$$

where $\sup_{h\in H}|\Delta_\ell(r)| = O_p(m^{-1/2})$, $\ell = 1, 2$.

PROOF. Let $S_r = \sum_{j=1}^{n-r} (Z_j - 1)/(n - j + 1)$. Then

$$\exp[-\rho \textstyle\sum_{j=1}^{n-r} Z_j/(n - j + 1)] = (r/n)^\rho\{1 + \Delta_1(r)\},$$

where $|\Delta_1(r)| \le M[r^{-1} + |S_r| \exp\{\rho|S_r|\}]$. It follows from the invariance principle for tail sequences (see for example, Heyde, 1977) that

$$\sup_{h\in H} r^{1/2}|S_r| = O_p(1), \tag{3.3}$$

as $n \to \infty$, and so (3.1) obtains.

Next, write

$$S_{ri} = \sum_{j=n-r+1}^{n-i} (Z_j - 1)/(n - j + 1)$$
$$= \sum_{j=1}^{n-i} (Z_j - 1)/(n - j + 1) - \sum_{j=1}^{n-r} (Z_j - 1)/(n - j + 1) = S_{ni} - S_r,$$

and $s_{ri} = \sum_{j=i+1}^{r} j^{-1}$, for $i < r$. By the invariance principle for tail sequences again,

$$\sup_{0 \le i \le n-1} (i + 1)^{1/2} | S_{ni} | = O_p(1), \tag{3.4}$$

as $n \to \infty$. Also, as $n \to \infty$,

$$\sup_{h \in H} r^{-1} \sum_{i=0}^{r-1} (i + 1)^{-1/2} \{(i + 1)/r\}^{\rho} = O(m^{-1/2}) \tag{3.5}$$

and

$$\sup_{h \in H} r^{-1} \sum_{i=0}^{r-1} \{(i + 1)/r\}^{\rho} = (1 + \rho)^{-1} + O(m^{-1}). \tag{3.6}$$

Hence, by (3.3)–(3.6),

$$\sup_{h \in H} r^{-1} \sum_{i=0}^{r-1} \{(i + 1)/r\}^{\rho} | S_{ri} | = O_p(m^{-1/2}), \tag{3.7}$$

as $n \to \infty$. Now with $| R_{ri} | \le M | S_{ri} | \exp\{\rho | S_{ri} |\}$,

$$\begin{aligned} r^{-1} \textstyle\sum_{i=0}^{r-1} \exp\{-\rho \sum_{j=n-r+1}^{n-i} Z_j/(n - j + 1)\} & \\ &= r^{-1} \textstyle\sum_{i=0}^{r-1} \exp(-\rho s_{ri})(1 + R_{ri}) \\ &= r^{-1} \textstyle\sum_{i=0}^{r-1} \{(i + 1)/r\}^{\rho}[1 + O\{(i + 1)^{-1}\}](1 + R_{ri}) \\ &= (1 + \rho)^{-1} + \Delta(r), \end{aligned} \tag{3.8}$$

where $\sup_{h \in H} | \Delta(r) | = O_p(m^{-1/2})$, by (3.6) and (3.7). Condition (3.2) obtains from (3.1) and (3.8).

To continue the proof of Theorem 3.1, observe that for a distribution function F satisfying (1.1), we have

$$\log F^{-1}(e^{-z}) = \log C^{-1/\alpha} - \alpha^{-1}\{z + DC^{-\rho}e^{-\rho z} + o(e^{-\rho z})\}, \tag{3.9}$$

as $z \to \infty$. Let $\{Z_j\}$ be standard exponential random variables. Using Rényi's representation of order statistics we may write

$$X_{ni} = F^{-1}[\exp\{- \textstyle\sum_{j=1}^{n-i+1} Z_j/(n - j + 1)\}], \quad 1 \le i \le n. \tag{3.10}$$

Combining (1.2), (3.9) and (3.10) we have

$$\begin{aligned} & n^{\rho/(2\rho+1)}\alpha(\hat{\alpha}_r^{-1} - \alpha^{-1}) \\ &= n^{\rho/(2\rho+1)}\{r^{-1} \textstyle\sum_{i=0}^{r-1} \sum_{j=1}^{n-i} Z_j/(n - j + 1) - \sum_{j=1}^{n-r} Z_j/(n - j + 1) - 1\} \\ &\quad - n^{\rho/(2\rho+1)}DC^{-\rho}\exp\{-\rho \textstyle\sum_{j=1}^{n-r} Z_j/(n - j + 1)\} \\ &\quad \cdot[1 - r^{-1} \textstyle\sum_{i=0}^{r-1} \exp\{-\rho \sum_{j=n-r+1}^{n-i} Z_j/(n - j + 1)\}] + R_1(r), \end{aligned}$$

where $\sup_{h\in H}|R_1(r)| = o_p(1)$, by Lemma 3.1. Applying Lemma 3.1,

$$n^{\rho/(2\rho+1)}\alpha(\hat{\alpha}_r^{-1} - \alpha^{-1})$$
$$= h^{-1}n^{-\rho/(2\rho+1)} \sum_{j=n-r+1}^{n} (Z_j - 1) - DC^{-\rho}\rho(1+\rho)^{-1}h^{\rho} + R_2(r),$$

where $\sup_{h\in H}|R_2(r)| = o_p(1)$.

It follows from Donsker's theorem that $n^{-\rho/(2\rho+1)} \sum_{j=n-r+1}^{n} (Z_j - 1)$ converges weakly to $W(\cdot)$, and hence that $n^{\rho/(2\rho+1)}\alpha(\hat{\alpha}_r^{-1} - \alpha^{-1})$ converges weakly to $-G(\cdot)$. Thus $A_n^*(\cdot)$ converges weakly to $\alpha G(\cdot)$. Finally,

$$n^{\rho/(2\rho+1)}(\log n)^{-1}(\log \hat{C}_r - \log C) = \alpha^{-1}(2\rho+1)^{-1}A_n^*(h) + R_3(r),$$

where $\sup_{h\in H}|R_3(r)| = o_p(1)$, and so Theorem 3.1 is proved.

4. Adaptive estimates. The invariance principles of Theorem 3.1 lead immediately to the following limit theorems for $\hat{\alpha}_{\hat{r}}$ and $\hat{C}_{\hat{r}}$, in which $\hat{r}$ denotes a random variable taking integer values.

THEOREM 4.1. *If condition* (1.1) *holds, if* $r = r(n) = [\lambda n^{2\rho/(2\rho+1)}]$ *and* $\hat{r}/r \to 1$ *in probability, then*

$$n^{\rho/(2\rho+1)}(\hat{\alpha}_{\hat{r}} - \alpha) \to \alpha Z$$

and

$$n^{\rho/(2\rho+1)}(\log n)^{-1}(\hat{C}_{\hat{r}} - C) \to C(2\rho+1)^{-1}Z$$

in distribution as $n \to \infty$, *where* Z *denotes a normal variable with mean* $DC^{-\rho}\rho(1+\rho)^{-1}\lambda^{\rho}$ *and variance* λ^{-1}.

Note that the mean square error of each limit distribution in Theorem 4.1 is minimized by taking $\lambda = \lambda_0 \equiv \{C^{2\rho}(1+\rho)^2/2D^2\rho^3\}^{1/(2\rho+1)}$.

Theorem 4.1 holds very generally, in that it requires no assumptions about the nature of the random sequence $\hat{r}$, other than that $\hat{r}/r \to 1$ in probability. During the remainder of this section we develop a specific estimate, $\hat{r}_0$, of $r_0 \equiv \lambda_0 n^{2\rho/(2\rho+1)}$. Then we may deduce from Theorem 4.1 that the adaptive estimators $\hat{\alpha}_{\hat{r}_0}$ and $\hat{C}_{\hat{r}_0}$ perform as well as $\hat{\alpha}_{r_0}$ and $\hat{C}_{r_0}$, in the sense of minimizing asymptotic mean square error.

Distributions satisfying condition (1.1) usually arise as

(i) powers of smooth distributions, i.e. of distributions admitting a Taylor series expansion of at least three terms about the origin; or

(ii) extreme value distributions $F(x) = e^{-x^{-\alpha}}$, $x > 0$, or stable distributions with index $0 < \alpha < 1$; or

(iii) stable distributions with index $1 < \alpha < 2$.

In the first case ρ will usually equal 1, but may equal 2 if $F''(0) = 0$. In the second case $\rho = 1$, and in the third case, $\frac{1}{2} < \rho \leq 1$. (These inequalities also hold in many cases where $\alpha = 1$.) Thus for all the above cases, $\rho \in (\frac{1}{2}, 2]$. In fact, usually $\rho \in (\frac{1}{2}, 1]$, and we often have $\rho = 1$. Thus it is reasonable to assume that we know a (possibly conservative) interval (ρ_1, ρ_2) such that $\rho \in (\rho_1, \rho_2)$.

We now describe our estimators of ρ and λ_0. Suppose we know that $\rho \in (\rho_1, \rho_2)$. Choose $0 < \sigma < 2\rho_1/(2\rho_1 + 1)$ and $2\rho_2/(2\rho_2 + 1) < \tau_1 < \tau_2 < 1$ such that $2\rho_2(1 - \tau_1) < \sigma$. Set $s = [n^\sigma]$, $t_1 = [n^{\tau_1}]$ and $t_2 = [n^{\tau_2}]$. With $\hat{\alpha}_r$ defined as in (1.2), set

$$\hat{\rho} = |\log|(\hat{\alpha}_{t_1} - \hat{\alpha}_s)/(\hat{\alpha}_{t_2} - \hat{\alpha}_s)|/\log(t_1/t_2)|$$

and

$$\hat{\lambda}_0 = |\hat{\alpha}_s/(2\hat{\rho})^{1/2}(n/t_1)^{\hat{\rho}}(\hat{\alpha}_{t_1} - \hat{\alpha}_s)|^{2/(2\hat{\rho}+1)}.$$

Then put

$$\hat{r}_0 = [\hat{\lambda}_0 n^{2\hat{\rho}/(2\hat{\rho}+1)}].$$

THEOREM 4.2. *With the above definition of $\hat{r}_0$, we have $\hat{r}_0/r_0 \to 1$ in probability as $n \to \infty$.*

PROOF. It follows from the definition of s, t_1 and t_2 that $s \to \infty$, $s/n^{2\rho/(2\rho+1)} \to 0$, $t_j/n \to 0$, $t_j/n^{2\rho/(2\rho+1)} \to \infty$ $(j = 1, 2)$ and $n^\rho/t_1^\rho s^{1/2} \to 0$ as $n \to \infty$. It follows from Theorem 2 of Hall (1982) that

$$\hat{\alpha}_s - \alpha = O_p(s^{-1/2})$$

and

$$\hat{\alpha}_{t_j} - \alpha = \alpha D C^{-\rho}\rho(1 + \rho)^{-1}(t_j/n)^\rho\{1 + o_p(1)\}, \quad j = 1, 2.$$

Therefore

$$\begin{aligned}(\hat{\alpha}_{t_1} - \hat{\alpha}_s)/(\hat{\alpha}_{t_2} - \hat{\alpha}_s) &= (t_1/t_2)^\rho\{1 + o_p(1) + O_p(n^\rho/t_1^\rho s^{1/2})\} \\ &= (t_1/t_2)^\rho\{1 + o_p(1)\},\end{aligned}$$

so that

$$\hat{\rho} = \rho + o_p\{(\log n)^{-1}\}.$$

Hence

$$\{2\hat{\rho}/(2\hat{\rho} + 1) - 2\rho/(2\rho + 1)\}\log n = o_p(1),$$

which implies that

$$n^{2\hat{\rho}/(2\hat{\rho}+1)}/n^{2\rho/(2\rho+1)} \to 1$$

in probability, as $n \to \infty$. Next, notice that

$$\begin{aligned}(n/t_1)^\rho(\hat{\alpha}_{t_1} - \hat{\alpha}_s)/\hat{\alpha}_s &= (n/t_1)^\rho\{(\hat{\alpha}_{t_1} - \alpha)/\alpha + O_p(s^{-1/2})\} \\ &= DC^{-\rho}\rho(1 + \rho)^{-1}\{1 + o_p(1)\},\end{aligned}$$

TABLE 5.1
Relative root mean squared errors (RMSE) of estimators of $\alpha = 1$ estimated from 200 samples.

Sample size n	50	100	200	500
RMSE (THEOR)	.235	.187	.148	.109
RMSE (THEOR)/RMSE ($\hat{\alpha} - \alpha$):				
r KNOWN	1.11	1.18	1.16	1.09
ρ KNOWN	.66	.70	.75	.81
ρ UNKNOWN	.56	.51	.61	.62

whence

$$\begin{aligned}\log\{(n/t_1)^{\rho}(\hat{\alpha}_{t_1} - \hat{\alpha}_s)/\hat{\alpha}_s\} + (\hat{\rho} - \rho)\log(n/t_1) + \tfrac{1}{2}\log(2\hat{\rho}) \\ = \log\{DC^{-\rho}\rho(1+\rho)^{-1}\} + \tfrac{1}{2}\log 2\rho + o_p(1) \\ = -\{(2\rho+1)/2\}\log\lambda_0 + o_p(1).\end{aligned}$$

Thus

$$\log\hat{\lambda}_0 - \log\lambda_0 = o_p(1),$$

and the result obtains.

5. Simulation results. In order to examine the finite sample properties of our estimators, a simulation study (200 replications) was performed for samples of size $n = 50$, 100, 200 and 500. The samples were generated as follows. Let Y be an exponential random variable with $E(Y) = 1/C$. Then $X = Y^{1/\alpha}$ has a distribution function F satisfying (1.1) with $D = -C/2$ and $\beta = \alpha$. In this case, $\rho = 1$ and $r_0 = [2n^{2/3}]$ for all $\alpha > 0$ and $C > 0$. Notice that for the values of n in the simulation, $r_0 = 27$, 43, 68 and 125 respectively, so that the effective sample size for estimating α and C is quite small. We restrict attention to the case $\alpha = 1$ and consider separately the situations where r is known, where ρ is known and where ρ is unknown.

In the case where ρ is known, we set $s = [n^{0.6}]$ and $t = [n^{0.95}]$, while in the case where ρ is unknown we put $s = [n^{0.5}]$, $t_1 = [n^{0.9}]$ and $t_2 = [n^{0.95}]$. The results are given in Table 5.1, which lists first the root mean squared error (RMSE) of $(\hat{\alpha} - 1)$ predicted by Theorem 2.1, and then the square root of the relative efficiencies.

When r is unknown, the performance of $\hat{\alpha}$ depends on both n and the extent of knowledge about ρ. The estimator $\hat{r}_0$ typically underestimates r_0, which increases the variance of $\hat{\alpha}$. For the sample sizes considered, the standard deviation of $\hat{r}_0$ is of the order of half the mean in each case, and this affects the performance of $\hat{\alpha}$. When ρ is assumed known, $\hat{\alpha}$ performs well for moderate sample sizes.

REFERENCES

BARTHOLOMEW, D. J. (1957). Testing for departure from the exponential distribution. *Biometrika* **44** 253–257.

DAVID, H. A. (1981). *Order Statistics* 2nd ed. Wiley, New York.

DE HAAN, L. and RESNICK, S. I. (1980). A simple asymptotic estimate for the index of a stable distribution. *J. Roy. Statist. Soc.* B **42** 83–88.

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

HALL, P. and WELSH, A. H. (1984). Best attainable rates of convergence for estimates of parameters of regular variation. *Ann. Statist.* **12** 1079–1083

HEYDE, C. C. (1977). On central limit and iterated logarithm supplements to the martingale convergence theorem. *J. Appl. Probab.* **14** 758–775.

HILL, B. M. (1975). A simple general approach to inference about the tail of a distribution. *Ann. Math. Statist.* **3** 1163–1174.

PROSCHAN, F. and PYKE, R. (1964). Asymptotic normality of certain test statistics of exponentiality. *Biometrika* **51** 253–256.

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

WELSH, A. H. (1985). On the use of the empirical distribution and characteristic function to estimate parameters of regular variation. *Austral. J. Statist.* To appear.

DEPARTMENT OF STATISTICS
UNIVERSITY OF CHICAGO
5734 UNIVERSITY AVENUE
CHICAGO, IL 60637

DEPARTMENT OF STATISTICS
FACULTY OF ECONOMICS
AUSTRALIAN NATIONAL UNIVERSITY
GPO BOX 4, CANBERRA ACT 2601
AUSTRALIA