The Annals of Statistics
1982, Vol. 10, No. 2, 556–568

ON ESTIMATING THE ENDPOINT OF A DISTRIBUTION

By Peter Hall

Australian National University

We propose a method of estimating the endpoint, θ, of a distribution when only limited information is available about the behaviour of the distribution in the neighbourhood of θ. By using increasing numbers of extreme order statistics we obtain an estimator which improves on earlier estimators based on only a bounded number of extremes. In a certain particular model our estimator is equal to a maximum likelihood estimator, but it is robust against departures from this model.

1. Introduction and summary. Let f be a density with support confined to the positive half line and which satisfies $f(x) = x^k L(x)$ as $x \to 0^+$, where $k \geq 0$ and L is slowly varying at the origin. Define $f_\theta(x) = f(\theta - x)$, $-\infty < x < \infty$. The problem of estimating the location parameter θ has been addressed by several authors; see for example Polfeldt (1970a, 1970b), Woodroofe (1972, 1974), Weiss and Wolfowitz (1973), Akahira (1975a, 1975b) and Akahira and Takeuchi (1979). It has generally been assumed that the distribution is completely specified except for the parameter θ, and in this case the technique of maximum likelihood is a natural choice for estimation. However, if the density f can be described with some accuracy only in the neighbourhood of the origin, then the estimator of θ should be a function of a relatively small proportion of the sample.

Recently attention has centered on just this case. Cooke (1979, 1980) built upon work of Robson and Whitlock (1964) and proposed estimators based on linear combinations of a fixed number of extreme order statistics. However, when $k \geq 1$ Cooke's restriction to a bounded number of extreme values can waste valuable information about the parameter θ. We shall propose a very different solution to this problem, and advocate the use of increasing numbers of order statistics. Our estimator is asymptotically efficient in comparison to other estimators computed using the same amount of information.

Cooke's work is confined to the case where k is known, and it seems difficult to extend his approach beyond that situation. We shall examine the case of unknown k, and show that the penalty paid for this lack of information is to increase the variance of the estimator of θ by a factor of k^2 $(k \geq 1)$. Thus, if k is close to 1 there is a relatively small loss in not knowing the value of k.

We suggest that estimation proceed as though the density were given by $g(x) = c(k+1)x^k$ in some domain $(0, \varepsilon)$, rather than by f. (Here $c > 0$ is an unknown constant.) Our estimate of θ is that value which would maximise the likelihood of the r largest order statistics $X_{n,n-r+1}, \cdots, X_{nn}$, if the true density were $g_\theta(x) = c(k+1)(\theta - x)^k$. This fictitious assumption serves only to derive a formula for calculating the estimator $\hat{\theta}$. We shall show in Section 2 that the estimator's limiting distribution has minimum variance in a class of robust estimators based on the first r order statistics. Here the term "robust" is used to denote that the asymptotic distribution of the estimator is unchanged if the very specific model g is generalized to the model

$$(1) \qquad f(x) = c(k+1)x^k\{1 + O(x^\ell)\} \quad \text{as} \quad x \to O^+,$$

or more generally, $1 - F(x) = cx^{k+1}\{1 + O(x^\ell)\}$, where F is the distribution function.

Received February 1981; revised August 1981.

AMS 1980 *subject classification.* Primary 62F10, 62G05; secondary 60F05.

Key words and phrases. Central limit theorem, efficiency, endpoint, extreme order statistics, regularly varying, robust.

Under this assumption our estimators converge at a considerably faster rate than those proposed by Cooke (1979, 1980) for the case $k \geq 1$. Admittedly this improvement is achieved at the expense of some tightening of the restrictions on f, for we are assuming that the slowly varying function L has the form $L(x) = c + O(x^{\ell})$ as $x \downarrow 0$ for some constant c. However, as we shall show in Section 2, this assumption is a reasonable one in many cases, and indeed it will often be possible to obtain some information about ℓ. The parameter c is assumed to be unknown and may be estimated at the same time as θ, but for our purposes we shall regard c as a nuisance parameter.

The improvements we have just described hold in the case $k \geq 1$, and are not possible for $0 < k < 1$. In fact it follows from work of Polfeldt (1970a, 1970b) and Woodroofe (1974) that even if f_θ is completely specified on $(-\infty, \theta)$, the maximum likelihood estimator based on the entire sample will converge at the same rate as the "naive" estimator, $\tilde{\theta}_n = X_{nn}$, if $0 < k < 1$. In this situation only limited improvements can be achieved by using more than a fixed number of order statistics to construct the estimator, and we shall examine such "finite" estimators in Section 3. We shall propose estimators which are asymptotically unbiased. The results in Sections 2 and 3 cover the case where k is known, and in Section 4 we treat the case of unknown k. The proofs are placed together in Section 5.

Throughout this paper we shall use the notation $\nu = 1/(k+1)$, and let $\rightarrow_{\mathscr{D}}$, $\rightarrow_p$ and $\rightarrow_{\text{a.s.}}$ denote convergence in distribution, convergence in probability and almost sure convergence, respectively. As far as possible we have tried to adopt the notation of Cooke (1979, 1980).

2. The case $k \geq 1$, and k known. Let us assume for the time being that in some neighbourhood of θ, the true density is given by

$$f_\theta(x) = c(k+1)(\theta - x)^k, \quad \theta - \varepsilon < x < \theta, \quad \text{and} \quad f_\theta(x) = 0, \quad x \geq \theta, \tag{2}$$

where only $k \geq 1$ is known. (We do not specify f_θ for $x \leq \theta - \varepsilon$.) We propose that the estimator $\hat{\theta}$ be chosen to maximise the likelihood of the r largest order statistics $X_{n,n-r+1} < \cdots < X_{nn}$, where $r \geq 2$. A similar approach was employed by Hill (1975) when estimating the exponent of a distribution with regularly varying tails. If $X_{n,n-r+1} > \theta - \varepsilon$, the likelihood is given by

$$L_n(\mathbf{X} \mid \theta, c) = \{n!/(n-r)!\}\{c(k+1)\}^r\{\textstyle\prod_{j=1}^r (\theta - X_{n,n-j+1})^k\}\{1 - c(\theta - X_{n,n-r+1})^{k+1}\}^{n-r}.$$

It is readily deduced that $\hat{c} = r/n(\hat{\theta} - X_{n,n-r+1})^{k+1}$, and $\hat{\theta}$ is the solution of the equation

$$k \textstyle\sum_{j=1}^{r-1} (X_{n,n-j+1} - X_{n,n-r+1})/(\theta - X_{n,n-j+1}) = r, \quad \theta > X_{nn}. \tag{3}$$

Note that with probability one the left side is a strictly decreasing function of θ, and so the likelihood equation has a unique root.

Having derived the estimator $\hat{\theta}$ we now drop our fictitious assumptions about the underlying distribution. First of all we observe that under very general conditions, $\hat{\theta}$ is strongly consistent for θ.

THEOREM 1. (*Strong consistency.*) *Suppose $r \geq 2$ and $r/n \rightarrow 0$ as $n \rightarrow \infty$. If $F_\theta(x) < 1$ for $x < \theta$, if $F_\theta(\theta) = 1$ and F_θ is continuous at θ, then $\hat{\theta} \rightarrow_{\text{a.s.}} \theta$.*

Next we show that $\hat{\theta}$ is robust under departures of the form (1) from our imagined model (2). Note that in the following result we do not even assume that the underlying distribution is continuous.

THEOREM 2. *Suppose $1 - F_\theta(x) = c(\theta - x)^{k+1}\{1 + O((\theta - x)^{\ell})\}$ as $x \uparrow \theta$, where $\ell > 0$. Set $\nu = 1/(k+1)$ and $m = \min(1, \nu\ell)$. If $k > 1$, $r = r(n) \rightarrow \infty$ and $r = o(n^{m/(m+1/2)})$ then $n^\nu r^{1/2-\nu}(\hat{\theta} - \theta)$ is asymptotically normal $N(0, (1-2\nu)/c^{2\nu})$, while if $k = 1$ and $r = O(n^{m/(m+1/2)})$ then $(n \log r)^{1/2}(\hat{\theta} - \theta)$ is asymptotically normal $N(0, c^{-1})$.*

In many situations it will be reasonable to assume that the underlying distribution F_θ is that of $X = \theta - Z^\nu$, where Z is a positive random variable whose density h satisfies $h(0^+) > 0$ and $h'(x)$ exists near the origin and is bounded. Here we should take $\ell = k + 1$, and then $m/(m + \frac{1}{2}) = \frac{2}{3}$. In other cases it would be more correct to suppose that $1 - F_\theta(x) = (\theta - x)^{k+1}H(\theta - x)$ for $x < \theta$, where $H(0^+) > 0$ and $H'(x)$ exists and is bounded. Then $\ell = 1$ and $m/(m + \frac{1}{2}) = 2/(k + 3)$.

The problem investigated in Theorem 2 is similar in some respects to that of nonparametric density estimation. In both situations we are estimating a quantity which is defined in terms of the local behaviour of the distribution, and since we lack parametric knowledge about the behaviour of the distribution on the whole real line, we base our estimation on order statistics from the neighbourhood under investigation. The parameter r plays the same role as the "window size" in the context of density estimation, and the problems associated with its use are very similar to those encountered in density estimation. If we are prepared to strengthen the assumption in Theorem 2 to the model

$$1 - F_\theta(x) = c(\theta - x)^{k+1}\{1 + C(\theta - x)^{\ell} + o((\theta - x)^{\ell})\}$$

as $x \to \theta^-$, where C is a nonzero constant, then the techniques used to prove Theorem 2 may be extended to derive formulae for the bias and mean square error of $\hat{\theta}$, which are functions of C, c, k and ℓ as well as r and n. Such formulae are analogues of better known expressions in the case of density estimators; see, for example, equations (15) and (16) of Rosenblatt (1971). An asymptotically optimal formula for r can be determined by minimising the mean square error, but this expression depends on the unknown parameters C and c, and is consequently of little practical value.

Our maximum likelihood approach may be used when somewhat more information is available about f—for example, if we are prepared to assume that $f(-x) = c(k + 1)x^k\{1 + d_1x^{\ell_1} + \cdots + d_mx^{\ell_m} + O(x^{\ell_{m+1}})\}$, where $d_1, d_2, \cdots, d_m$ are unknown constants and $\ell_1 < \ell_2 < \cdots < \ell_{m+1}$ are known constants. However the likelihood equation loses its simple form, and the technique of maximum likelihood estimation is less attractive.

Let us temporarily revert to the assumption (2). If $k > 1$ then the equation

$$\int L_n(\mathbf{x} \mid \theta, c)\, d\mathbf{x} = 1 \tag{4}$$

may be differentiated once under the integral sign with respect to either θ or c. The classical argument leading to the derivation of the information matrix and the Cramér-Rao lower bound may be conducted as usual, even though our case is somewhat irregular in that an endpoint of the distribution depends on an unknown parameter. Note that a second differentiation of (4) with respect to θ will not be valid unless $k > 2$, and so the information matrix should be established in terms of the first derivatives rather than the more commonly used second derivatives.

If $k = 1$ then not even a first differentiation is possible, and we are in a situation similar to that described by Weiss and Wolfowitz (1973). Like these authors we shall establish a lower bound for the variance by proving that the maximum likelihood estimator is asymptotically equivalent to a maximum probability estimator. Now, there are two unknown parameters in our problem, and their maximum probability estimators converge at different rates. Indeed, in the notation of Weiss and Wolfowitz (1967) the vector $((n \log r)^{1/2}, r^{1/2})$ is a normalizing factor for $(\hat{\theta}, \hat{c})$. The mathematics of our problem is considerably more complicated than in Weiss and Wolfowitz (1973). However, it turns out that in the exceptional case $k = 1$ the maximum likelihood estimator of θ calculated with c known has asymptotically the same distribution as that calculated with c unknown. That is, information about c is of negligible assistance in estimating θ. (This is not true for $k > 1$, where the relative efficiency of the two estimators equals k^{-2}.) Therefore it is both more informative and simpler to establish a minimum variance bound for the estimation of θ when c is known.

THEOREM 3. *Suppose the density is given by* (2), *and* $r \to \infty$. *If* $k > 1$ *and* $r = o(n^{2/3})$ *then the Cramér-Rao lower bound to the variance of* $\hat{\theta}$ *is given by* $\{1 + o(1)\}(1 - 2\nu)/c^{2\nu}n^{2\nu}r^{1-2\nu}$, *and so equals the asymptotic variance derived under the more general model in Theorem* 2. *If* $k = 1$ *then this lower bound is not well defined, but when* $r = O(n^{2/3})$, $\hat{\theta}$ *is asymptotically equivalent to a maximum probability estimator calculated when c is known. Therefore when* $k = 1$, $\hat{\theta}$ *has asymptotically minimum variance in the class of asymptotically normal estimators.*

Theorem 2 states that the first order asymptotic properties of our estimator $\hat{\theta}$ are unaffected by departures of the type (1) from the model (2), provided $r/n \to 0$ sufficiently quickly. It follows from Theorem 3 that $\hat{\theta}$ is asymptotically efficient in the class of asymptotically normal, unbiased estimators with this property. To see this, let $\tilde{\theta}$ be such an estimator and suppose for convenience that $k > 1$. Then there exist normalising constants $d_n = d_n(r)$ such that for each distribution satisfying (1), and for r/n tending to zero sufficiently quickly, we have $(\tilde{\theta} - \theta)/d_n \to_{\mathscr{D}} N(0, 1)$. In particular this is true in the special case of the model (2), and so it follows from Theorem 3 that

$$\liminf_{n\to\infty} d_n^2/\{(1 - 2\nu)/c^{2\nu}n^{2\nu}r^{1-2\nu}\} \geq 1.$$

But for the special case $\tilde{\theta} = \hat{\theta}$ we know from Theorem 2 that we may take $d_n^2 = (1 - 2\nu)/c^{2\nu}n^{2\nu}r^{1-2\nu}$, and so the minimum variance bound is asymptotically attained by $\hat{\theta}$.

3. The case $k > 0$, and k known. Suppose

$$1 - F_\theta(x) \sim c(\theta - x)^{k+1} \quad \text{as} \quad x \uparrow \theta, \tag{5}$$

where $k > 0$ and $r \geq 2$ is fixed. Then it can be shown that the maximum likelihood estimator $\hat{\theta}$ defined by equation (3) satisfies the relation $n^\nu(\hat{\theta} - \theta) \to_{\mathscr{D}} Z_r$, where the random variable Z_r is the solution of the equation

$$k \sum_{j=1}^{r-1} \{(\sum_{i=1}^{r} Y_i)^\nu - (\sum_{i=1}^{j} Y_i)^\nu\}/\{c^\nu Z_r + (\sum_{i=1}^{j} Y_i)^\nu\} = r, \quad c^\nu Z_r > -Y_1^\nu, \tag{6}$$

and $Y_1, \cdots, Y_r$ are independent exponential random variables. It may be proved using the techniques of the proof of Theorem 2 that if $k > 1$, $r^{1/2-\nu}Z_r \to_{\mathscr{D}} N(0, (1 - 2\nu)/c^{2\nu})$, and if $k = 1$, $(\log r)^{1/2}Z_r \to_{\mathscr{D}} N(0, c^{-1})$. Therefore $Z_r \to_p 0$ as $r \to \infty$ if $k \geq 1$, and the precision of the estimator may be improved by using a larger value of r. However, this result does not carry over to the case $0 < k < 1$.

THEOREM 4. *If* $0 < k < 1$ *and* Z_r *is defined by* (6) *then* $Z_r \not\to_p 0$ *as* $r \to \infty$.

A similar argument will show that if we allow n and r to diverge to infinity together in the manner of Section 2, then $n^\nu(\hat{\theta} - \theta) \not\to_p 0$ in the case $0 < k < 1$. Therefore when $0 < k < 1$ there is no obvious advantage in using a larger number of order statistics to form the estimator. Robson and Whitlock (1964) and Cooke (1979, 1980) used the criterion of mean square error to compare different estimators. Note however that the asymptotic distributions of their estimators are very different from one another, and so other criteria could have been employed. For example, we could choose as an estimator a linear combination of the r largest order statistics which has the property that its asymptotic distribution has zero mean, and a variance which is a minimum among all such linear combinations.

Let $\Lambda = (\lambda_{ij})$ be the symmetric $r \times r$ matrix given by

$$\lambda_{ij} = \Gamma(2\nu + i)\Gamma(\nu + j)/\{\Gamma(\nu + i)\Gamma(j)\}, \quad j \leq i,$$

and define $\mathbf{1}$, $\mathbf{v}$ and $\mathbf{a}$ to be the column vectors of length r with ith elements equal to 1, $\Gamma(\nu + i)/\Gamma(i)$ and a_i, respectively, $1 \leq i \leq r$.

THEOREM 5. *Suppose condition* (5) *holds, and* $r \geq 2$ *is fixed. The estimator of the* *form* $\tilde{\theta} = \sum_{j=1}^{r} a_j X_{n,n-j+1}$ *which is consistent and whose limiting distribution has zero mean and minimum variance is obtained using the weights*

$$\mathbf{a} = \Lambda^{-1}\{(\mathbf{v}^T\Lambda^{-1}\mathbf{v})\mathbf{1} - (\mathbf{1}^T\Lambda^{-1}\mathbf{v})\mathbf{v}\}/\{(\mathbf{v}^T\Lambda^{-1}\mathbf{v})(\mathbf{1}^T\Lambda^{-1}\mathbf{1}) - (\mathbf{1}^T\Lambda^{-1}\mathbf{v})^2\}.$$

Further, $n^{\nu}(\tilde{\theta} - \theta) \to_{\mathscr{D}} -\sum_{j=1}^{r} a_j(\sum_{i=1}^{j} Y_i/c)^{\nu}$, *where* $Y_1, \cdots, Y_r$ *are independent exponential variables.*

4. The case $k > 1$, but k otherwise unknown. As in Section 2 we begin with the very restrictive model (2), and once we have derived an expression for our estimator we greatly relax our assumptions. The likelihood is now a function of three unknowns, and after eliminating c we find that the estimators of θ and k are the solutions of the equations

$$k = r/\{\textstyle\sum_{j=1}^{r-1} \log(1 + \xi_{nj})\} - 1 \tag{7}$$

and $k \sum_{j=1}^{r-1} \xi_{nj} = r$, where $\xi_{nj} = \xi_{nj}(\theta) = (X_{n,n-j+1} - X_{n,n-r+1})/(\theta - X_{n,n-j+1})$. Therefore the maximum likelihood estimator $\hat{\theta}$ is a solution of the equation

$$r\{1/\textstyle\sum_{j=1}^{r-1} \log(1 + \xi_{nj}) - 1/\sum_{j=1}^{r-1} \xi_{nj}\} = 1, \quad \theta > X_{nn}. \tag{8}$$

We shall show during the proof of Theorem 6 that under the conditions of Theorem 2, the probability that one or more solutions of equation (8) exist tends to 1 as $n \to \infty$. For the sake of definiteness, let us define $\hat{\theta}$ to be the smallest solution of (8) if one or more solutions exist, and $\hat{\theta} = X_{nn}$ otherwise. Define $\hat{k}$ by (7) with $\theta = \hat{\theta}$. Our next result describes the asymptotic behaviour of the estimators $\hat{\theta}$ and $\hat{k}$.

THEOREM 6. *Assume the conditions of Theorem* 2, *and suppose* $r = r(n) \to \infty$ *and* $r = o(n^{m/(m+1/2)})$. *If* $k > 1$ *then* $n^{\nu} r^{1/2-\nu}(\hat{\theta} - \theta) \to_{\mathscr{D}} N(0, (1 - 2\nu)(1 - \nu)^2/\nu^2 c^{2\nu})$ *and* $r^{1/2}(\hat{k} - k) \to_{\mathscr{D}} N(0, (1 - \nu)^2/\nu^4)$, *while if* $k = 1$ *then* $(n \log r)^{1/2}(\hat{\theta} - \theta) \to_{\mathscr{D}} N(0, c^{-1})$ *and* $r^{1/2}(\hat{k} - 1) \to_{\mathscr{D}} N(0, 4)$.

Thus, the estimator $\hat{\theta}$ computed with k unknown has much the same properties as that computed with k known, except that its variance is reduced by a factor of k^{-2}.

A lower bound for the quantity $m/(m + ½)$ is needed in order to carry out the estimation procedure in Theorem 6. However, as was pointed out following Theorem 2, there will be practical situations in which we know that $m/(m + ½)$ equals either ⅔ or $2/(k + 3)$, and then the procedure can be conducted with very limited knowledge of k.

5. Proofs. Here the symbol C denotes a generic positive constant.

PROOF OF THEOREM 1. If false then since $\hat{\theta} > X_{nn} \to_{\text{a.s.}} \theta$, there exist $\varepsilon > 0$ and a set A with $P(A) > 0$ such that $\hat{\theta} > \theta + \varepsilon$ infinitely often on A. Let $B = A \cap \{X_{n,n-r+1} \to \theta\}$. Then the left side of (3) is less than $k \sum_{j=1}^{r-1} (X_{n,n-j+1} - X_{n,n-r+1})/(½ \varepsilon)$ infinitely often on B, and this term is in turn dominated by $2kr(\theta - X_{n,n-r+1})/\varepsilon$, which is $o(r)$ on B. Therefore equation (3) is not always satisfied on B – a contradiction.

PROOF OF THEOREM 2. There is no loss of generality in assuming that $c = 1$, for the contrary case may be handled by making the obvious transformation. And since θ is a location parameter we may suppose that $\theta = 0$.

Let $Y_1, Y_2, \cdots$ be independent exponential variables with mean 1. The sequence $S_n = \{X_{n,n-j+1}, 1 \le j \le r\}$ has the same distribution as $\{F^{-1}[\exp\{-\sum_{i=1}^{j} Y_i/(n - i + 1)\}], 1 \le j \le r\}$ for each $n \ge 1$ (Rényi, 1953). Since we are only interested in the weak properties of S_n there is no real loss of generality in redefining

$$X_{n,n-j+1} = F^{-1}[\exp\{-\textstyle\sum_{i=1}^{j} Y_i/(n - i + 1)\}] \text{ for all } 1 \le j \le n < \infty.$$

Note that strong properties of this new sequence imply only weak properties of the original sequence.

Under the hypothesis on F we have $F^{-1}(y) = -(1-y)^{\nu} + O((1-y)^{\nu(\ell+1)})$ as $y \to 1$, and therefore

$$(9)\qquad \begin{aligned} X_{n,n-j+1} &= -[1 - \exp\{-\textstyle\sum_{i=1}^{j} Y_i/(n-i+1)\}]^{\nu} + O((j/n)^{\nu(\ell+1)}) \\ &= -\{\textstyle\sum_{i=1}^{j} Y_i/(n-i+1)\}^{\nu} + O((j/n)^{\nu+m}) \\ &= -(\textstyle\sum_{i=1}^{j} Y_i/n)^{\nu} + O((j/n)^{\nu+m}) \quad \text{a.s.} \end{aligned}$$

uniformly in $1 \le j \le r$. The likelihood equation (3) may now be rewritten as

$$(10)\quad k \textstyle\sum_{j=1}^{r-1} [\{(\sum_{i=1}^{r} Y_i)^{\nu} - (\sum_{i=1}^{j} Y_i)^{\nu} + O(r^{\nu+m}/n^m)\} \times \{n^{\nu}\hat{\theta} + (\sum_{i=1}^{j} Y_i)^{\nu} + O(j^{\nu+m}/n^m)\}^{-1}] = r.$$

Next we prove that under the conditions of Theorem 2,

$$(11)\qquad n^{\nu}\hat{\theta} \to_p 0.$$

Let $-\infty < \xi < \infty$ and $\xi \neq 0$, and suppose that $n^{\nu}\hat{\theta} > \xi$ (respectively, $< \xi$). Then the left side of (10) is less than (greater than)

$$(12)\quad \begin{aligned} &k \textstyle\sum_{j=1}^{r-1} \{(\sum_{i=1}^{r} Y_i/\sum_{i=1}^{j} Y_i)^{\nu} - 1 + O(r^{\nu+m}/j^{\nu}n^m)\}\{\xi/(\sum_{i=1}^{j} Y_i)^{\nu} + 1 + O((j/n)^m)\}^{-1} \\ &= k \textstyle\sum_{j=1}^{r-1} \{(\sum_{i=1}^{r} Y_i/\sum_{i=1}^{j} Y_i)^{\nu} - 1\} - k\xi \sum_{j=1}^{r-1} \{(\sum_{i=1}^{r} Y_i)^{\nu}/(\sum_{i=1}^{j} Y_i)^{2\nu} - (\sum_{i=1}^{j} Y_i)^{-\nu}\} \\ &\quad + O(\textstyle\sum_{j=1}^{r-1} \{r^{\nu+m}/j^{\nu}n^m + (r/j)^{\nu}(j/n)^m + (r/j)^{\nu}j^{-2\nu}\}) \quad \text{a.s.} \end{aligned}$$

The remainder here equals $O(r^{1+m}/n^m + r^{\nu} \sum_1^r j^{-3\nu})$ a.s. Let $U_j = \sum_{i=1}^{j} (Y_i - 1)$. Then

$$(\textstyle\sum_{i=1}^{j} Y_i)^{-\nu} = j^{-\nu} - \nu U_j/j^{1+\nu} + O(j^{-1-\nu} \log\log j)$$

and

$$(\textstyle\sum_{i=1}^{r} Y_i)^{\nu} = r^{\nu} + \nu U_r/r^{1-\nu} + O(r^{-1+\nu} \log\log r) \quad \text{a.s.,}$$

whence

$$\textstyle\sum_{j=1}^{r-1} (\sum_{i=1}^{j} Y_i)^{-\nu} = r^{1-\nu}/(1-\nu) - \nu \sum_{j=1}^{r} U_j/j^{1+\nu} + O(1) \quad \text{a.s.}$$

and

$$(13)\quad k \textstyle\sum_{j=1}^{r-1} \{(\sum_{i=1}^{r} Y_i/\sum_{i=1}^{j} Y_i)^{\nu} - 1\} = r - r^{\nu}(1-\nu) \sum_{j=1}^{r} U_j/j^{1+\nu} + U_r + O(r^{\nu}) \quad \text{a.s.}$$

Furthermore, $\sum_{j=1}^{r} U_j/j^{1+\nu} = \sum_{i=1}^{r} (Y_i - 1) \sum_{j=1}^{r} j^{-1-\nu}$.

Suppose first that $\nu = ½$. Then

$$\operatorname{var}(\textstyle\sum_{j=1}^{r} U_j/j^{1+\nu}) = \sum_{i=1}^{r} (\sum_{j=1}^{r} j^{-1-\nu})^2 = O(\log r)$$

(see condition (15) below), and also

$$\textstyle\sum_{j=1}^{r-1} \{(\sum_{i=1}^{r} Y_i)^{\nu}/(\sum_{i=1}^{j} Y_i)^{2\nu} - (\sum_{i=1}^{j} Y_i)^{-\nu}\} \sim r^{\nu} \sum_{j=1}^{r-1} j^{-1} \sim r^{\nu} \log r \text{ a.s.}$$

Therefore in view of (13) the quantity in (12) may be written as

$$r - k\xi r^{\nu}(\log r)\{1 + o_p(1)\} + O_p(r^{1+m}/n^m + r^{\nu}(\log r)^{1/2}).$$

Comparing this with the right side of (10) we deduce that if $r = O(n^{m/(m+1/2)})$ then for $\xi > 0$, $P(n^{\nu}\hat{\theta} > \xi) \to 0$, and for $\xi < 0$, $P(n^{\nu}\hat{\theta} < \xi) \to 0$.

If $\nu < ½$ then $\operatorname{var}(\sum_{j=1}^{r} U_j/j^{1+\nu}) = O(r^{1-2\nu})$ (see condition (15)) and

$$\textstyle\sum_{j=1}^{r-1} \{(\sum_{i=1}^{r} Y_i)^{\nu}/(\sum_{i=1}^{j} Y_i)^{2\nu} - (\sum_{i=1}^{j} Y_i)^{-\nu}\} \sim r^{1-\nu}\, \nu/(1-\nu)(1-2\nu) \quad \text{a.s.}$$

Therefore in view of (13) the quantity (12) may be written as

$$r - k\xi r^{1-\nu}\nu(1-\nu)^{-1}(1-2\nu)^{-1}\{1 + o_p(1)\} + O_p(r^{1+m}/n^m + r^{\nu} \textstyle\sum_1^r j^{-3\nu} + r^{1/2}),$$

and (11) follows as before.

In view of (11) the left side of (10) is equal to

$$\sum_{j=1}^{r-1} \{(\sum_{i=1}^{r} Y_i/\sum_{i=1}^{j} Y_i)^{\nu} - 1\}$$
$$- kn^{\nu}\hat{\theta} \sum_{j=1}^{r-1} \{(\sum_{i=1}^{r} Y_i)^{\nu}/(\sum_{i=1}^{j} Y_i)^{2\nu} - (\sum_{i=1}^{j} Y_i)^{-\nu}\} + O_p(r^{1+m}/n^m + r^{\nu}(n^{\nu}\hat{\theta})^2 \sum_{j=1}^{r} j^{-3\nu});$$

compare the expression (12). Using the estimates derived during the proof of (11) we may rewrite this as

$$(14) \qquad r - r^{\nu}(1-\nu) \sum_{j=1}^{r} U_j/j^{1+\nu} + U_r - kn^{\nu}\hat{\theta}a(r)\{1 + o_p(1)\} + O_p(r^{1+m}/n^m + r^{\nu}(n^{\nu}\hat{\theta})^2 \sum_{j=1}^{r} j^{-3\nu}),$$

where $a(r) = r^{1/2} \log r$ if $\nu = ½$, and $a(r) = r^{1-\nu}\nu/(1-\nu)(1-2\nu)$ if $\nu < ½$. The variance of $U_r - r^{\nu}(1-\nu) \sum_{j=1}^{r} U_j/j^{1+\nu}$ may be derived using condition (15) below, and then Lindeberg's theorem may be used to prove that this quantity is asymptotically normal distributed.

Next we prove that

$$(15) \qquad \operatorname{var}(\sum_{j=1}^{r} U_j/j^{1+\nu}) = \sum_{i=1}^{r} (\sum_{j=i}^{r} j^{-1-\nu})^2 \sim \begin{cases} 4 \log r & \text{if } \nu = ½ \\ 2r^{1-2\nu}/(1-\nu)(1-2\nu) & \text{if } \nu < ½. \end{cases}$$

Now,

$$\begin{aligned}\sum_{i=1}^{r} \sum_{j=i}^{r} \sum_{k=i}^{r} (jk)^{-1-\nu} &= \sum_{j=1}^{r} \sum_{i=1}^{j} \sum_{k=i}^{r} (jk)^{-1-\nu} \\ &= \sum_{j=1}^{r} (\sum_{k=1}^{j} \sum_{i=1}^{k} + \sum_{k=j+1}^{r} \sum_{i=1}^{j})(jk)^{-1-\nu} \\ &= \sum_{j=1}^{r} j^{-1-\nu} \sum_{k=1}^{j} k^{-\nu} + \sum_{j=1}^{r} j^{-\nu} \sum_{k=j+1}^{r} k^{-1-\nu} \\ &= 2 \sum_{j=1}^{r} j^{-1-\nu} \sum_{k=1}^{j} k^{-\nu} - \sum_{j=1}^{r} j^{-1-2\nu} \sim 2 \sum_{j=1}^{r} j^{-2\nu}/(1-\nu),\end{aligned}$$

as required.

Theorem 2 follows on equating the quantity (14) to r, and solving for $\hat{\theta}$. For example, when $\nu < ½$ we have

$$\begin{aligned}kn^{\nu}\hat{\theta}a(r) &= U_r - r^{\nu}(1-\nu) \sum_{j=1}^{r} U_j/j^{1+\nu} + o_p(r^{1/2} + n^{\nu}|\hat{\theta}|\, a(r)) \\ &= \sum_{i=1}^{r} (Y_i - 1)\{1 - r^{\nu}(1-\nu) \sum_{j=i}^{r} j^{-1-\nu}\} + o_p(r^{1/2} + n^{\nu}|\hat{\theta}|\, a(r)),\end{aligned}$$

and the series on the right has variance equal to

$$\begin{aligned}\sum_{i=1}^{r} \{1 - r^{\nu}(1-\nu) \sum_{j=i}^{r} j^{-1-\nu}\}^2 &= r - 2(1-\nu)r^{\nu} \sum_{j=1}^{r} j^{-\nu} \\ &\quad + (1-\nu)^2 r^{2\nu} \sum_{i=1}^{r} (\sum_{j=i}^{r} j^{-1-\nu})^2 \\ &= r/(1-2\nu) + o(r),\end{aligned}$$

using (15).

PROOF OF THEOREM 3. As in the proof of Theorem 2 we may assume that $c = 1$, $\theta = 0$ and $X_{n,n-j+1} = F^{-1}[\exp\{-\sum_{i=1}^{j} Y_i/(n-i+1)\}]$, $1 \le j \le n < \infty$. We consider first the case $k > 1$.

LEMMA 1. *If $r/n \to 0$ then*

$$(16) \qquad \partial \log L_n/\partial\theta = k \sum_{j=1}^{r} \{\sum_{i=1}^{j} Y_i/(n-i+1)\}^{-\nu} - (n-r)(k+1)\{\sum_{i=1}^{r} Y_i/(n-i+1)\}^{1-\nu} + R_1,$$

where $E(R_1^2) = O((r^{2-\nu}/n^{1-\nu})^2)$.

PROOF. In our case $X_{n,n-j+1} = -[1 - \exp\{-\sum_{i=1}^{j} Y_i/(n-i+1)\}]^{\nu}$, and so

$$\begin{aligned} \partial \log L_n/\partial\theta|_{\theta=0} &= k \textstyle\sum_{j=1}^{r} (-X_{n,n-j+1})^{-1} \\ &\quad - (n-r)(k+1)(-X_{n,n-r+1})^{k}\{1 - (-X_{n,n-r+1})^{k+1}\}^{-1} \\ &= k \textstyle\sum_{j=1}^{r} [1 - \exp\{-\sum_{i=1}^{j} Y_i/(n-i+1)\}]^{-\nu} \\ &\quad - (n-r)(k+1)[1 - \exp\{-\sum_{i=1}^{r} Y_i/(n-i+1)\}]^{k\nu} \\ &\quad \cdot \exp\{\textstyle\sum_{i=1}^{r} Y_i/(n-i+1)\}. \end{aligned} \tag{17}$$

Now for $x > 0$, $x^{-1} < (1 - e^{-x})^{-1} = e^x/(e^x - 1) < e^x/x < x^{-1} + e^x$, and therefore $x^{-\nu} < (1 - e^{-x})^{-\nu} < x^{-\nu}(1 + xe^x)^{\nu} < x^{-\nu} + x^{1-\nu}e^x$. Consequently

$$\textstyle\sum_{j=1}^{r} [1 - \exp\{-\sum_{i=1}^{j} Y_i/(n-i+1)\}]^{-\nu} = \sum_{j=1}^{r} \{\sum_{i=1}^{j} Y_i/(n-i+1)\}^{-\nu} + R_{11} \tag{18}$$

where

$$0 < R_{11} < \textstyle\sum_{j=1}^{r} \{\sum_{i=1}^{j} Y_i/(n-i+1)\}^{1-\nu} \exp\{\sum_{i=1}^{j} Y_i/(n-i+1)\} < Cr(S_r/n)^{1-\nu}\exp(CS_r/n)$$

with $S_r = \sum_1^r Y_i$. Furthermore, since $x > 1 - e^{-x} > x(1 - \frac{1}{2}x)$ and $k\nu < 1$ then $x^{k\nu} > (1 - e^{-x})^{k\nu} > x^{k\nu}(1 - x)$, whence $x^{k\nu}(1 + xe^x) > x^{k\nu}e^x > (1 - e^{-x})^{k\nu}e^x > x^{k\nu}(1 - x)$. Therefore

$$[1 - \exp\{-\textstyle\sum_{i=1}^{r} Y_i/(n-i+1)\}]^{k\nu} \exp\{\sum_{i=1}^{r} Y_i/(n-i+1)\} = \{\sum_{i=1}^{r} Y_i/(n-i+1)\}^{k\nu} + R_{12}$$

where

$$|R_{12}| \le \{\textstyle\sum_{i=1}^{r} y_i/(n-i+1)\}^{k\nu+1} \exp\{\sum_{i=1}^{r} Y_i/(n-i+1)\} < C(S_r/n)^{2-\nu} \exp(CS_r/n).$$

Combining this with (17) and (18) we deduce that (16) holds with $|R_1| \le C\{r(S_r/n)^{1-\nu} + n(S_r/n)^{2-\nu}\}\exp(CS_r/n)$. Now S_r is Gamma (r), and it follows from elementary calculus that $E\{S_r^{\alpha} \exp(CS_r/n)\} = O(r^{\alpha})$ as $n \to \infty$, for any $\alpha > 0$ and $C > 0$. Therefore $E(R_1^2) = O((r^{2-\nu}/n^{1-\nu})^2)$.

LEMMA 2. *Let* $a_j = \sum_{i=1}^{j} (n-i+1)^{-1}$ *and* $T_j = \sum_{i=1}^{j} (Y_i - 1)/(n-i+1)$. *If* $r/n \to 0$ *and* $\nu < \frac{1}{2}$ *then*

$$\begin{aligned} &k \textstyle\sum_{j=1}^{r} \{\sum_{i=1}^{j} Y_i/(n-i+1)\}^{-\nu} - (n-r)(k+1)\{\sum_{i=1}^{r} Y_i/(n-i+1)\}^{1-\nu} \\ &\qquad = k \textstyle\sum_{j=1}^{r} (a_j^{-\nu} - \nu T_j/a_j^{1+\nu}) - (n-r)(k+1)\{a_r^{1-\nu} + (1-\nu)T_r/a_r^{\nu}\} + R_2 \end{aligned} \tag{19}$$

where $E(R_2^2) = O(n^{2\nu}r^{\varepsilon})$ *for all* $\varepsilon > 0$.

PROOF. Let $\varepsilon > 0$ and s be an integer between 1 and r, and define $E = E(s, \varepsilon) = \{\sup_{j\ge s} |j^{-1} \sum_{i=1}^{j} (Y_i - 1)| \le \varepsilon\}$. Then $|T_j| \le \varepsilon j/(n-r)$ on E if $j \ge s$, and so if ε is chosen sufficiently small, $|T_j/a_j| \le \frac{1}{2}$ on E for $j \ge s$. Therefore on E and for $j \ge s$,

$$\{\textstyle\sum_{i=1}^{j} Y_i/(n-i+1)\}^{-\nu} = a_j^{-\nu}(1 + T_j/a_j)^{-\nu} = a_j^{-\nu}(1 - \nu T_j/a_j) + R_{21},$$

where $|R_{21}| \le CT_j^2/a_j^{2+\nu}$ and C does not depend on j. Also

$$\{\textstyle\sum_{i=1}^{r} Y_i/(n-i+1)\}^{1-\nu} = a_r^{1-\nu}(1 + T_r/a_r)^{1-\nu} = a_r^{1-\nu}\{1 + (1-\nu)T_r/a_r\} + R_{22}$$

where $|R_{22}| < CT_r^2/a_r^{1+\nu}$. Therefore on E,

$$\begin{aligned} &k \textstyle\sum_{j=s}^{r} \{\sum_{i=1}^{j} Y_i/(n-i+1)\}^{-\nu} - (n-r)(k+1)\{\sum_{i=1}^{r} Y_i/(n-i+1)\}^{1-\nu} \\ &\qquad = k \textstyle\sum_{j=s}^{r} (a_j^{-\nu} - \nu T_j/a_j^{1+\nu}) - (n-r)(k+1)\{a_r^{1-\nu} + (1-\nu)T_r/a_r^{\nu}\} + R_{23} \end{aligned}$$

where, since $|T_j| \le (n-r)^{-1}|U_j|$ and $a_j \ge j/n$, $|R_{23}| \le Cn^{\nu}(\sum_{j=1}^{r} U_j^2/j^{2+\nu} + U_r^2/r^{1+\nu})$ and

$$E\{R_{23}^2 I(E)\} \le C_1 n^{2\nu}\{\textstyle\sum_{i=1}^{r}\sum_{j=1}^{r} E(U_i^2 U_j^2)/(ij)^{2+\nu} + E(U_r^4)/r^{2(1+\nu)}\} \le C_2 n^{2\nu}.$$

(Note that $E(U_i^2 U_j^2) \leq (EU_i^4)^{1/2}(EU_j^4)^{1/2} \leq Cij$.) Furthermore,

$$E[\sum_{j=1}^{s-1} \{\sum_{i=1}^{j} Y_i/(n-i+1)\}^{-\nu}]^2 \leq E\{s(n/Y_1)^\nu\}^2 = s^2 n^{2\nu} E(Y_1^{-2\nu})$$

and

$$E\{\sum_{j=1}^{s-1} (a_j^{-\nu} - \nu T_j/a_j^{1+\nu})\}^2 \leq Cs^2 n^{2\nu}.$$

Combining these results and noting that $E(Y_1^{-2\nu}) < \infty$ when $\nu < \frac{1}{2}$ we deduce that (19) holds on E, with $E\{R_2^2 I(E)\} = O(s^2 n^{2\nu})$.

Fix $\delta > 0$ and let s equal the integer part of r^δ. Then $P(\tilde{E}) = O(r^{-\lambda})$ for all $\lambda > 0$ (see Petrov, 1975, Theorem 27, page 283), and so for any variable Z, any $p > 1$ and $q = (1 - p^{-1})^{-1}$,

$$E\{Z^2 I(\tilde{E})\} \leq (E\,|\,Z\,|^{2p})^{1/p}\{P(\tilde{E})\}^{1/q} = (E\,|\,Z\,|^{2p})^{1/p} O(r^{-\lambda}).$$

Now, all terms on the right and left in (19) except R_2 are dominated by $Z = Cn^\nu(rY_1^{-\nu} + r^{1-\nu} + |\,U_r\,|^{1-\nu} + r\,|\,U_r\,|)$, and if p is chosen in the range $1 < p < \frac{1}{2}\nu$ then $E\,|\,Z\,|^{2p} \leq Cn^{2\nu p} r^{3p}$. Therefore (19) holds on $\tilde{E}$, with $E\{R_2^2 I(\tilde{E})\} = O(n^{2\nu} r^{-\lambda})$ for all $\lambda > 0$. This completes the proof of Lemma 2.

LEMMA 3. *If $r = o(n^{2/3})$ and $\nu < \frac{1}{2}$ then $k \sum_{j=1}^{r} a_j^{-\nu} - (n-r)(k+1)a_r^{1-\nu} = o(n^\nu r^{1/2-\nu})$.*

PROOF. It is readily shown that $a_j^{-\nu} = (n/j)^\nu + O\{(j/n)^{1-\nu} + n^\nu/j^{1+\nu}\}$ uniformly in $1 \leq j \leq r$, and $a_r^{1-\nu} = (r/n)^{1-\nu} + O\{(r/n)^{2-\nu} + 1/n^{1-\nu}r^\nu\}$, whence

$$\begin{aligned} k \sum_{j=1}^{r} a_j^{-\nu} - (n-r)(k+1)a_r^{1-\nu} &= k \sum_1^r (n/j)^\nu - (n-r)(k+1)(r/n)^{1-\nu} \\ &\quad + O(r^{2-\nu}/n^{1-\nu} + n^\nu) \\ &= (\nu^{-1} - 1)n^\nu\{r^{1-\nu}/(1-\nu) + O(1)\} \\ &\quad - n^\nu \nu^{-1} r^{1-\nu} + O(r^{2-\nu}/n^{1-\nu} + n^\nu) = o(n^\nu r^{1/2-\nu}). \end{aligned}$$

LEMMA 4. *If $r/n \to 0$ and $\nu < \frac{1}{2}$ then $E\{\nu k \sum_{j=1}^{r} T_j/a_j^{1+\nu} + (n-r)(k+1)(1-\nu)T_r/a_r^\nu\}^2 \sim n^{2\nu} r^{1-2\nu}(1-\nu)^2/\nu^2(1-2\nu)$.*

PROOF. The term within braces equals

$$(1-\nu) \sum_{i=1}^{r} (Y_i - 1)(n-i+1)^{-1}\{\sum_{j=i}^{r} a_j^{-(1+\nu)} + (n-r)/\nu a_r^\nu\},$$

and this variable has mean square equal to

$$\begin{aligned} &(1-\nu)^2 \sum_{i=1}^{r} (n-i+1)^{-2}\{\sum_{j=i}^{r} a_j^{-(1+\nu)} + (n-r)/\nu a_r^\nu\}^2 \\ &\quad \sim (1-\nu)^2 n^{2\nu} \sum_{i=1}^{r} \{\sum_{j=i}^{r} j^{-(1+\nu)} + 1/\nu r^\nu\}^2 \\ &\quad = (1-\nu)^2 n^{2\nu} \{\sum_{i=1}^{r} (\sum_{j=i}^{r} j^{-(1+\nu)})^2 + 2(\nu r^\nu)^{-1} \sum_{j=1}^{r} j^{-\nu} + r^{1-2\nu}/\nu^2\}. \end{aligned}$$

Lemma 4 now follows from (15).

Combining Lemmas 1–4 we deduce that if $\nu < \frac{1}{2}$ and $r = o(n^{2/3})$ then

$$\partial \log L_n/\partial\theta = -\{\nu k \sum_{j=1}^{r} T_j/a_j^{1+\nu} + (n-r)(k+1)(1-\nu)T_r/a_r^\nu\} + R_3, \tag{20}$$

where $E(R_3^2) = o(n^{2\nu} r^{1-2\nu})$, and also

$$E(\partial \log L_n/\partial\theta)^2 \sim n^{2\nu} r^{1-2\nu}(1-\nu)^2/\nu^2(1-2\nu). \tag{21}$$

Very similar techniques can be used to prove that $\partial \log L_n/\partial c = -nT_r/c + R_4$ where $E(R_4^2) = o(r)$, and when this is combined with (20) we may obtain

$$E(\partial \log L_n/\partial c)^2 \sim r \quad \text{and} \quad E\{(\partial \log L_n/\partial\theta)(\partial \log L_n/\partial c)\} \sim n^\nu r^{1-\nu}/\nu. \tag{22}$$

(The Cauchy-Schwartz inequality is used to handle the remainders R_3 and R_4, and we have set $c = 1$ in (20).) The information matrix (see, for example, Kendall and Stewart, 1973, page 28) may be computed using (21) and (22), and an inversion of this matrix yields the desired Cramér-Rao lower bound.

We turn finally to the case $k = 1$. We shall use techniques from Weiss and Wolfowitz (1967, 1973, 1974), and prove that the maximum likelihood estimator $\hat{\theta}$ derived without knowing c is asymptotically equivalent to a maximum probability estimator derived knowing c. It is sufficient to show that there exist constants $O < \varepsilon_n \to 0$ and $a(n) \to \infty$ with $a(n)/(n \log r)^{1/2} \to 0$, such that for each $s > 0$ the inequality

$$\int_{-s}^{s} L_n(\mathbf{X} \mid \hat{\theta} + t/(n \log r)^{1/2}, c)\, dt \geq (1 - \varepsilon_n)\sup_{|u|<a(n)} \int_{-s+u}^{s+u} L_n(\mathbf{X} \mid \hat{\theta} + t/(n \log r)^{1/2}, c)\, dt \tag{23}$$

holds with $P_{\theta,c}$-probability tending to 1 as $n \to \infty$; this follows from the Theorem and the remarks on pages 198–199 of Weiss and Wolfowitz (1967). For (23) it is sufficient to prove that with

$$V_n(t) = L_n(\mathbf{X} \mid \hat{\theta} + t/(n \log r)^{1/2}, c)/L_n(\mathbf{X} \mid \hat{\theta}, c)$$

we have $\sup_{|t|\leq 2a(n)}\{|\log V_n(t) + \tfrac{1}{2}t^2 |/t^2\} \to 0$ in $P_{\theta,c}$-probability, and this last result will follow if we show that

$$\sup_{|t|\leq 3a(n)} |(n \log r)^{-1}\{\partial^2 \log L_n(\mathbf{X} \mid \theta + t/(n \log r)^{1/2}, c)/\partial\theta^2\}c + 1| \to 0 \tag{24}$$

in $P_{\theta,c}$-probability. But

$$-\partial^2 \log L_n(\mathbf{X} \mid \theta, c)/\partial\theta^2 = \sum_1^r (\theta - X_{n,n-j+1})^{-2} + 2c(n-r)\{1 - c(\theta - X_{n,n-r+1})^2\}^{-1} + 4c^2(n-r)(\theta - X_{n,n-r+1})^3\{1 - c(\theta - X_{n,n-r+1})^2\}^{-2},$$

and since $X_{n,n-r+1} \to_{\text{a.s.}} \theta$ then (24) will follow if we prove that

$$\sup_{|t|\leq 3a(n)/(n\log r)^{1/2}} |(n \log r)^{-1} \sum_{j=1}^r (\theta + t - X_{n,n-j+1})^{-2} - 1| \to_p 0. \tag{25}$$

Define $a(n) = (\log r)^{1/4}$. Now,

$$\theta - X_{n,n-j+1} = (\sum_{i=1}^j Y_i/n)^{1/2} + O((j/n)^{3/2}) = (j/n)^{1/2} + O\{(j/n)^{3/2} + n^{-1/2}(\log\log j)^{1/2}\} \quad \text{a.s.},$$

and so

$$\sum_{j=1}^r (\theta + t - X_{n,n-j+1})^{-2} = n \sum_1^r j^{-1} + O(n) \quad \text{a.s.}$$

uniformly in $|t| \leq 3a(n)/(n \log r)^{1/2}$. This proves (25).

Proof of Theorem 4. We may assume that $c = 1$. If $2|Z_r| < Y_1^\nu$ then $\{Z_r + (\sum_{i=1}^j Y_i)^\nu\}^{-1} = (\sum_{i=1}^j Y_i)^{-\nu} + R_{4j}$ where $|R_{4j}| \leq 2|Z_r|/(\sum_{i=1}^j Y_i)^{2\nu}$, and in this case the left hand side of equation (6) may be written in the form

$$k\{(\sum_{j=1}^r Y_i)^\nu \sum_{j=1}^r (\sum_{i=1}^j Y_i)^{-\nu} - r\} + R_5,$$

where $|R_5| \leq 2k|Z_r|(\sum_{i=1}^r Y_i)^\nu \sum_{j=1}^r(\sum_{i=1}^j Y_i)^{-2\nu}$. Since $\nu > \tfrac{1}{2}$ then the last written series is $O(1)$ a.s., and so under the hypothesis that $Z_r \to_p 0$ we may rewrite (6) in the form

$$k\{(\sum_{j=1}^r Y_i)^\nu \sum_{j=1}^r (\sum_{i=1}^j Y_i)^{-\nu} - r\} + o_p(r^\nu) = r.$$

But $(\sum_{j=1}^r Y_i)^\nu = r^\nu + O_p(r^{\nu-1/2})$, and $\sum_{j=1}^r(\sum_{i=1}^j Y_i)^{-\nu} = O_p(r^{1-\nu})$, so that

$$\sum_{j=1}^r (\sum_{i=1}^j Y_i)^{-\nu} = r^{1-\nu}/(1-\nu) + o_p(1).$$

This statement is obviously incorrect; consider for example what happens to the series on the left if Y_1 is perturbed by a nonzero amount. Therefore Theorem 4 is proved by contradiction.

Theorem 5 may be proved using the method of Lagrange multipliers.

PROOF OF THEOREM 6. We adopt notation from the proof of Theorem 2, and take $c = 1$ and $\theta = 0$. Equation (8) can be written as

$$\hat{k} \textstyle\sum_{j=1}^{r-1} \xi_{nj} = r, \tag{26}$$

where $\hat{k}$ is given by (7). Let $\hat{\theta}_n$, $n \geq 1$, be *any* sequence of random variables which satisfies $n^\nu \hat{\theta}_n \to_p 0$, and set $\theta = \hat{\theta}_n$. From estimates obtained during the proof of Theorem 2, in particular the result (14), we may deduce that the left hand side of (26) equals

$$(\hat{k}/k)[r - r^\nu(1-\nu) \textstyle\sum_{j=1}^{r} U_j/j^{1+\nu} + U_r - kn^\nu\hat{\theta}_n a(r)\{1 + o_p(1)\} + O_p(r^{1+m}/n^m + r^\nu(n^\nu\hat{\theta}_n)^2 \sum_{j=1}^{r} j^{-3\nu})] \tag{27}$$

as $n \to \infty$. We now examine the asympototic behaviour of $\hat{k}/k$.

From the estimate (9) we may deduce that

$$\begin{aligned} 1 + \xi_{nj} &= (\hat{\theta}_n - X_{n,n-r+1})/(\hat{\theta}_n - X_{n,n-j+1}) \\ &= (\textstyle\sum_{i=1}^{r} Y_i/\sum_{i=1}^{j} Y_i)^\nu[1 + n^\nu\hat{\theta}_n/(\sum_{i=1}^{r} Y_i)^\nu + O((r/n)^m)] \\ &\quad \times [1 - n^\nu\hat{\theta}_n/(\textstyle\sum_{i=1}^{j} Y_i)^\nu + O((j/n)^m + (n^\nu\hat{\theta}_n)^2 j^{-2\nu})] \\ &= (\textstyle\sum_{i=1}^{r} Y_i/\sum_{i=1}^{j} Y_i)^\nu\{1 - n^\nu\hat{\theta}_n[(\sum_{i=1}^{j} Y_i)^{-\nu} - (\sum_{i=1}^{r} Y_i)^{-\nu}] + O((r/n)^m + (n^\nu\hat{\theta}_n)^2 j^{-2\nu})\} \end{aligned}$$

almost surely. Consequently

$$\begin{aligned} \textstyle\sum_{j=1}^{r-1} \log(1 + \xi_{nj}) &= \nu\{(r-1)\log r - \textstyle\sum_{j=1}^{r-1} \log j\} \\ &\quad + \nu\{(r-1)\log(1 + U_r/r) - \textstyle\sum_{j=1}^{r-1} \log(1 + U_j/j)\} \\ &\quad - n^\nu\hat{\theta}_n\{\textstyle\sum_{j=1}^{r-1} (\sum_{i=1}^{j} Y_i)^{-\nu} - (r-1)(\sum_{i=1}^{r} Y_i)^{-\nu}\} \\ &\quad + O_p(r^{1+m}/n^m + (n^\nu\hat{\theta}_n)^2 \textstyle\sum_{j=1}^{r} j^{-2\nu}). \end{aligned} \tag{28}$$

From Stirling's formula we see that

$$(r-1)\log r - \textstyle\sum_{j=1}^{r-1} \log j = \log\{r^{r-1}/\Gamma(r)\} = r - \frac{1}{2}\log r + O(1),$$

and since

$$\begin{aligned} \textstyle\sum_{j=1}^{r-1} \log(1 + U_j/j) &= \textstyle\sum_{j=1}^{r-1} U_j/j - \frac{1}{2}\sum_{j=1}^{r-1} U_j^2/j^2 + O(\sum_{j=1}^{r-1} j^{-3}(j \log\log j)^{3/2}) \text{ a.s.} \\ &= \textstyle\sum_{j=1}^{r-1} U_j/j + O_p(\log r), \end{aligned}$$

then

$$(r-1)\log(1 + U_r/r) - \textstyle\sum_{j=1}^{r-1} \log(1 + U_j/j) = U_r - \sum_{j=1}^{r-1} U_j/j + O_p(\log r).$$

Estimates derived during the proof of (11) allow us to state that

$$\textstyle\sum_{j=1}^{r-1}(\sum_{i=1}^{j} Y_i)^{-\nu} - (r-1)(\sum_{i=1}^{r} Y_i)^{-\nu} = \nu\{r^{1-\nu}/(1-\nu) + U_r/r^\nu - \sum_{j=1}^{r-1} U_j/j^{1+\nu}\} + O_p(1),$$

and in view of (15), the right hand side here equals $\nu r^{1-\nu}\{1 + o_p(1)\}/(1-\nu)$.

Continuing the results from (28) down we find that

$$\begin{aligned} (1/r\nu) \textstyle\sum_{j=1}^{r-1} \log(1 + \xi_{nj}) &= 1 + r^{-1}(U_r - \textstyle\sum_{j=1}^{r-1} U_j/j) - n^\nu\hat{\theta}_n\{1 + o_p(1)\}/r^\nu(1-\nu) \\ &\quad + O_p((r/n)^m + (n^\nu\hat{\theta}_n)^2 r^{-1} \textstyle\sum_{j=1}^{r} j^{-2\nu} + r^{-1}\log r). \end{aligned}$$

We may now deduce from (7) that

$$\nu\hat{k} = (1-\nu) - r^{-1}(U_r - \textstyle\sum_{j=1}^{r-1} U_j/j) + n^\nu\hat{\theta}_n\{1+o_p(1)\}/r^\nu(1-\nu) + O_p((r/n)^m + (n^\nu\hat{\theta}_n)^2 r^{-1}\sum_{j=1}^r j^{-2\nu} + r^{-1}\log r). \tag{29}$$

This leads to an expression for $\hat{k}/k$, and when that is substituted into (27) and the resulting formula simplified we obtain the following expression for the left hand side of (26):

$$\begin{aligned} &r(1 + r^{-1}[\textstyle\sum_{j=1}^r U_j\{1/j(1-\nu) - r^\nu(1-\nu)/j^{1+\nu}\} - \nu\, U_r/(1-\nu)] \\ &\quad + n^\nu\hat{\theta}_n\{1+o_p(1)\}\{1/r^\nu(1-\nu)^2 - (1-\nu)a(r)/\nu r\} \\ &\quad + O_p((r/n)^m + (n^\nu\hat{\theta}_n)^2 r^{\nu-1}\textstyle\sum_{j=1}^r j^{-3\nu} + r^{-1}\log r)). \end{aligned} \tag{30}$$

Let

$$\begin{aligned} V_r &= \textstyle\sum_{j=1}^r U_j\{1/j(1-\nu) - r^\nu(1-\nu)/j^{1+\nu}\} - \nu\, U_r/(1-\nu) \\ &= \textstyle\sum_{i=1}^r (Y_i - 1)\{\sum_{j=i}^r[1/j(1-\nu) - r^\nu(1-\nu)/j^{1+\nu}] - \nu/(1-\nu)\}. \end{aligned}$$

If $\nu = ½$ then the result (15) may be used to prove that $V_r = -\{1+o_p(1)\}V'_r$, where $V'_r = r^\nu(1-\nu)\sum_{j=1}^r U_j/j^{1+\nu}$ has asymptotic variance equal to $\{1+o(1)\}r\log r$. If $\nu < ½$ then after some algebra and an application of (15) we find that

$$\begin{aligned} E(V_r^2) &= (1-\nu)^{-2}\textstyle\sum_{i=1}^r(\sum_{j=i}^r j^{-1})^2 + r^{2\nu}(1-\nu)^2\sum_{i=1}^r(\sum_{j=i}^r j^{-1-\nu})^2 + 2\nu r^\nu\sum_{j=1}^r j^{-\nu} \\ &\quad - 2r^\nu\textstyle\sum_{i=1}^r(\sum_{j=i}^r j^{-1})(\sum_{j=i}^r j^{-1-\nu}) + r(\nu^2-2\nu)/(1-\nu)^2 \\ &= r\{2/(1-\nu)^2 + 2(1-\nu)/(1-2\nu) + 2\nu/(1-\nu) \\ &\quad + (\nu^2-2\nu)/(1-\nu)^2 + o(1)\} - 2r^\nu\textstyle\sum_{i=1}^r(\sum_{j=i}^r j^{-1})(\sum_{j=i}^r j^{-1-\nu}). \end{aligned}$$

It may be shown that

$$\textstyle\sum_{i=1}^r\sum_{j=i}^r j^{-1}\sum_{k=i}^r k^{-1-\nu} = \sum_{j=1}^r j^{-1}\sum_{k=1}^j k^{-\nu} + \sum_{k=2}^r(k-1)/k^{1+\nu} \sim r^{1-\nu}(2-\nu)/(1-\nu)^2,$$

and therefore $E(V_r^2) \sim r\nu^2/(1-\nu)^2(1-2\nu)$. Lindeberg's theorem can be used to prove that if $\nu \le ½$, V_r is asymptotically normally distributed with the appropriate variance derived above.

To complete the proof of the first parts of Theorem 6 we consider the cases $\nu < ½$ and $\nu = ½$ separately. If $\nu < ½$ then combining results from (30) down we see that with probability tending to 1 as $n \to \infty$ and under the conditions of Theorem 6, (26) admits a solution $\hat{\theta}_n$ satisfying $n^\nu\hat{\theta}_n \to_p O$, and that *any* such sequence of solutions also satisfies

$$n^\nu r^{1/2-\nu}\hat{\theta}_n \to_D N(O, (1-2\nu)(1-\nu)^2/\nu^2). \tag{31}$$

The same applies if $\nu = ½$, except that in this case (31) should be replaced by the result $(n\log r)^{1/2}\hat{\theta}_n \to_{\mathscr{D}} N(0,1)$. This proves that the probability that a solution of (8) exists converges to 1 as $n \to \infty$. Since any such solution exceeds X_{nn}, and $|X_{nn}| = -X_{nn} = O_p(n^{-\nu})$ as $n \to \infty$, then the *smallest* solution, $\hat{\theta}$, must satisfy $P(n^\nu\hat{\theta} > \varepsilon) \to 0$ for each $\varepsilon > 0$, and also

$$\lim_{\lambda\to\infty}\limsup_{n\to\infty} P(n^\nu\hat{\theta} < -\lambda) = 0. \tag{32}$$

Arguments like those leading to (11) may be used to strengthen (32) to

$$\lim_{n\to\infty} P(n^\nu\hat{\theta} < -\varepsilon) = 0 \quad \text{for each} \quad \varepsilon > 0,$$

and the asymptotic normality of $\hat{\theta}$ now follows from the results contained earlier in this paragraph.

Finally we prove the last parts of Theorem 6. Let $\hat{\theta}_n = \hat{\theta}$ (our maximum likelihood estimator) in equation (29). In the case $\nu = ½$ the random variable $n^\nu\hat{\theta}_n/r^\nu$ is negligible in comparison with $r^{-1}(U_r - \sum_{j=1}^{r-1} U_j/j)$, and we may deduce from (29) that

$$\hat{k} = k - 2r^{-1}(U_r - \textstyle\sum_{j=1}^{r-1} U_j/j) + o_p(r^{-1/2}).$$

From this it can be proved that $r^{1/2}(\hat{k} - k) \to_{\mathscr{D}} N(0, 4)$.

Now suppose $\nu < \frac{1}{2}$. Making the substitution

$$n^{\nu}\hat{\theta}_n\{1/r^{\nu}(1-\nu)^2 - (1-\nu)a(r)/\nu r\} = -r^{-1}V_r\{1 + o_p(1)\}$$

in (29), we may deduce that

$$\hat{k} = k + (1/\nu r)\sum_{i=1}^{r}(Y_i - 1)[(1-\nu)^2\nu^{-2}\sum_{j=i}^{r}\{j^{-1} - r^{\nu}(1-2\nu)/j^{1+\nu}\} - (1-\nu)/\nu] + o_p(r^{-1/2}).$$

The random variable

$$\sum_{i=1}^{r}(Y_i - 1)[(1-\nu)^2\nu^{-2}\sum_{j=i}^{r}\{j^{-1} - r^{\nu}(1-2\nu)/j^{1+\nu}\} - (1-\nu)/\nu]$$

is asymptotically normally distributed, with variance given by

$$\sum_{i=1}^{r}[(1-\nu)^2\nu^{-2}\sum_{j=i}^{r}\{j^{-1} - r^{\nu}(1-2\nu)/j^{1+\nu}\} - (1-\nu)/\nu]^2 \sim r(1-\nu)^2/\nu^2,$$

using techniques developed earlier. Therefore $r^{1/2}(\hat{k} - k) \to_{\mathscr{D}} N(0, (1-\nu)^2/\nu^4)$, completing the proof of Theorem 6.

Acknowledgements. The constructive suggestions of the Editor, Associate Editor and referee have led to major improvements over the original manuscript. I am grateful for their advice.

REFERENCES

AKAHIRA, M. (1975a). Asymptotic theory for estimation of location in nonregular cases. I. Order of convergence of consistent estimators. *Rep. Statist. Appl. Res.* **22** No. 1 8–26.

AKAHIRA, M. (1975b). Asymptotic theory for estimation of location in nonregular cases. II. Bounds of asymptotic distribution of consistent estimators. *Rep. Statist. Appl. Res.* **23** No. 3 3–19.

AKAHIRA, M. and TAKEUCHI, K. (1979). Discretized likelihood methods—asymptotic properties of discretized likelihood estimators (DLE's). *Ann. Inst. Statist. Math.* **31** 39–56.

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.

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

KENDALL, M. G. and STUART, A. (1973). *The Advanced Theory of Statistics*, Vol 2, third ed. Griffin, London.

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

POLFELDT, T. (1970a). The order of the minimum variance in a non-regular case. *Ann. Math. Statist.* **41** 667–672.

POLFELDT, T. (1970b). Asymptotic results in non-regular estimation. *Skand. Aktuarietidskr.* Supp 1–2.

RÉNYI, A. (1953). On the theory of order statistics. *Acta. Math. Acad. Sci. Hung.* **4** 191–231.

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

ROSENBLATT, M. (1971). Curve estimates. *Ann. Math. Statist.* **42** 1815–1842.

WEISS, L. and WOLFOWITZ, J. (1967). Maximum probability estimators. *Ann. Inst. Statist. Math.* **19** 193–206.

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

WEISS, L. and WOLFOWITZ, J. (1974). *Maximum Probability Estimators and Related Topics.* Lecture Notes in Math. **424** Springer, Berlin.

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 a translation parameter of a truncated distribution II. *Ann. Statist.* **3** 474–488.

DEPARTMENT OF STATISTICS
THE AUSTRALIAN NATIONAL UNIVERSITY
P.O. BOX 4
CANBERRA ACT 2600
AUSTRALIA