

BEST ATTAINABLE RATES OF CONVERGENCE FOR ESTIMATES OF PARAMETERS OF REGULAR VARIATION

BY PETER HALL AND A. H. WELSH

Australian National University

We derive lower bounds to rates of convergence for estimators of shape and scale parameters in distributions with regularly varying tails. We exhibit simple estimators which attain these rates.

1. Introduction. The problem of estimating parameters of a distribution F with regularly varying tails but otherwise arbitrary, has received considerable attention in recent years; see for example de Haan and Resnick (1980), Teugels (1981), Hall (1982) and Welsh (1983). In this paper we shall determine optimum asymptotic rates of convergence for sequences of estimators. Specifically, let α_n be an estimator of the exponent α , constructed out of a random n -sample X_1, \dots, X_n . We seek the fastest rate at which a real sequence $\{a_n\}$ can tend to zero and yet satisfy

$$\liminf_{n \rightarrow \infty} \inf_{\mathcal{D}} P(|\alpha_n - \alpha| \leq a_n) = 1,$$

for some class of distributions \mathcal{D} . We shall prove that this fastest rate is attained by a simple class of estimators.

We now describe the class \mathcal{D} of distributions. Given positive constants α_0 , C_0 , ε , ρ and A , let $\mathcal{D} = \mathcal{D}(\alpha_0, C_0, \varepsilon, \rho, A)$ denote the set of densities f on the positive half-line which satisfy

$$f(x) = C\alpha x^{\alpha-1}\{1 + r(x)\} \quad \text{and} \quad |r(x)| \leq Ax^\beta,$$

for all $x > 0$, where $|\alpha - \alpha_0| \leq \varepsilon$, $|C - C_0| \leq \varepsilon$ and $\beta = \rho\alpha$. The distribution function F associated with f satisfies

$$F(x) = Cx^\alpha\{1 + R(x)\},$$

where $|R(x)| \leq Ax^\beta$ for $x > 0$. Note that the union over A of all sets \mathcal{D} , consists of densities f which satisfy

$$f(x) = C\alpha x^{\alpha-1}\{1 + O(x^\beta)\}$$

as $x \rightarrow 0$. The latter condition often provides a satisfactory description of limited knowledge about the behaviour of f near the origin. We shall only consider explicitly those distributions which are regularly varying at the origin, although our results obviously extend to distributions which are regularly varying at infinity.

It turns out that the optimum rate of convergence depends on α and β through their ratio. This is the reason for keeping $\rho = \beta/\alpha$ fixed in the definition of \mathcal{D} .

Received March 1983; revised January 1984.

AMS 1980 subject classifications. Primary 62G05; secondary 62G30.

Key words and phrases. Estimating parameters of regular variation.

Hill (1975) and Hall (1982) proposed particular estimates $\hat{\alpha}_n$ and \hat{C}_n based on order statistics. We shall demonstrate in Section 3 that these estimators achieve optimum rates of convergence. The estimators proposed by de Haan and Resnick (1980), Teugels (1981) and Welsh (1983) do not achieve these optimum rates unless certain restrictive conditions are imposed.

2. Optimum rates of convergence. The theorems below describe optimum rates for estimates α_n and C_n of α and C , respectively. Let $\beta_0 = \rho\alpha_0$.

THEOREM 1. *Suppose that for some $\alpha_0, C_0, \varepsilon$ and ρ , we have*

$$\liminf_{n \rightarrow \infty} \inf_{f \in \mathcal{L}} P_f(|\alpha_n - \alpha| \leq a_n) = 1$$

for all $A > 0$. Then

$$\liminf_{n \rightarrow \infty} n^{\beta_0/(2\beta_0 + \alpha_0)} a_n = \infty.$$

THEOREM 2. *Suppose that for some $\alpha_0, C_0, \varepsilon$ and ρ , we have*

$$\liminf_{n \rightarrow \infty} \inf_{f \in \mathcal{L}} P_f(|C_n - C| \leq a_n) = 1$$

for all $A > 0$. Then

$$\liminf_{n \rightarrow \infty} n^{\beta_0/(2\beta_0 + \alpha_0)} (\log n)^{-1} a_n = \infty.$$

PROOF OF THEOREM 1. We construct two densities f_0 and f_1 , the first governed by fixed parameters α_0, C_0 and the second by varying parameters α_1, C_1, C_2 , where $\alpha_1 = \alpha_0 + \gamma, \gamma = \lambda n^{-\beta_1/(2\beta_1 + \alpha_1)}, \lambda > 0, \beta_1 = \rho\alpha_1$ and both $C_1, C_2 \rightarrow C_0$ as $n \rightarrow \infty$. Specifically, we define

$$f_0(x) = C_0 \alpha_0 x^{\alpha_0 - 1}, \quad 0 \leq x \leq C_0^{-1/\alpha_0},$$

and

$$f_1(x) = \begin{cases} C_1 \alpha_1 x^{\alpha_1 - 1} + \Delta(x), & 0 \leq x \leq \delta \\ C_2 \alpha_0 x^{\alpha_0 - 1}, & \delta < x \leq C_0^{-1/\alpha_0}, \end{cases}$$

where $\delta = n^{-1/(2\beta_1 + \alpha_1)}, k = \alpha_1 + \beta_1 - 1$ and

$$\Delta(x) = \begin{cases} x^k, & 0 < x \leq \delta/4 \\ (\delta/2 - x)^k, & \delta/4 < x \leq \delta/2 \\ -(x - \delta/2)^k, & \delta/2 < x \leq 3\delta/4 \\ -(\delta - x)^k, & 3\delta/4 < x \leq \delta. \end{cases}$$

Note that Δ is continuous on $[0, \delta]$, that $\Delta(0) = \Delta(\delta) = 0$ and

$$\int_0^\delta \Delta(x) dx = 0.$$

The constants C_1, C_2 are chosen so that for large n, f_1 is a proper, continuous density on $[0, C_0^{-1/\alpha_0}]$; that is,

$$(2.1) \quad C_1 \alpha_1 \delta^{\alpha_1} = C_2 \alpha_0 \delta^{\alpha_0}$$

and

$$(2.2) \quad C_1 \delta^{\alpha_1} + C_2(C_0^{-1} - \delta^{\alpha_0}) = 1.$$

Our proof consists initially of showing that

$$(2.3) \quad \int_0^{C_0^{-1/\alpha_0}} \{f_0(x) - f_1(x)\}^{-2} \{f_0(x)\}^{-1} dx = O(n^{-1})$$

as $n \rightarrow \infty$, and for all large n ,

$$(2.4) \quad f_1 \in \mathcal{D}(\alpha_0, C_0, \varepsilon, \rho, A)$$

for each $\varepsilon > 0$ and some $A > 0$. (It is obvious that $f_0 \in \mathcal{D}$.) The symbol K denotes a positive generic constant.

By (2.1) and (2.2),

$$(2.5) \quad \begin{aligned} 1/C_1 &= \delta^{\alpha_1}(1 - \alpha_1/\alpha_0) + (\alpha_1/\alpha_0)\delta^{\alpha_1 - \alpha_0}C_0^{-1} \\ &= C_0^{-1}\delta^\gamma(1 + \gamma/\alpha_0) + O(\gamma\delta^{\alpha_0}) \\ &= C_0^{-1}\{1 + \gamma \log \delta + O(\gamma)\}. \end{aligned}$$

Therefore

$$(2.6) \quad C_1 = C_0\{1 - \gamma \log \delta + O(\gamma)\}.$$

Also, by (2.2) and (2.5),

$$(C_2 - C_0)(C_0^{-1} - \delta^{\alpha_0}) = \delta^{\alpha_0}(\alpha_1/\alpha_0 - 1)(1 - C_0\delta^{\alpha_0}) \\ \times \{\delta^{\alpha_0}(1 - \alpha_1/\alpha_0) + (\alpha_1/\alpha_0)C_0^{-1}\} \sim C_0\alpha_1^{-1}\gamma\delta^{\alpha_0}$$

as $n \rightarrow \infty$. Therefore

$$(2.7) \quad |C_2 - C_0| = O(\gamma\delta^{\alpha_1}).$$

Next observe that

$$(2.8) \quad \begin{aligned} &\int_0^{C_0^{-1/\alpha_0}} \{f_0(x) - f_1(x)\}^2 \{f_0(x)\}^{-1} dx \\ &\leq 2 \int_0^\delta (C_0\alpha_0x^{\alpha_0-1} - C_1\alpha_1x^{\alpha_1-1})^2 (C_0\alpha_0x^{\alpha_0-1})^{-1} dx \\ &\quad + 2 \int_0^\delta \Delta^2(x)(C_0\alpha_0x^{\alpha_0-1})^{-1} dx + (C_0 - C_2)^2 \int_\delta^{C_0^{-1/\alpha_0}} C_0\alpha_0x^{\alpha_0-1} dx; \\ &(2\alpha_1 - \alpha_0) \int_0^\delta (C_0\alpha_0x^{\alpha_0-1} - C_1\alpha_1x^{\alpha_1-1})^2 x^{1-\alpha_0} dx \\ &= \delta^{\alpha_0}\{\alpha_0^2 + 2\alpha_0\gamma\}(C_0 - C_1\delta^\gamma)^2 + C_1^2\gamma^2\delta^{2\gamma} \\ &= O\{\delta^{\alpha_1}(C_0 - C_1\delta^\gamma)^2 + \gamma^2\delta^{\alpha_1}\}; \end{aligned}$$

$$C_0 - C_1\delta^\gamma = C_0 - C_0\{1 - \gamma \log \delta + O(\gamma)\}\{1 + \gamma \log \delta + O(\gamma)\} = O(\gamma),$$

using (2.6); $(C_0 - C_2)^2 = O(\gamma^2 \delta^{\alpha_1})$, by (2.7); and

$$\int_0^\delta \Delta^2(x)x^{1-\alpha_0} dx \leq K \int_0^\delta x^{2k+1-\alpha_0} dx = O(\delta^{2\beta_1+\alpha_1}).$$

Combining the estimates from (2.8) down, we see that the left-hand side of (2.3) equals

$$O(\gamma^2 \delta^{\alpha_1} + \delta^{2\beta_1+\alpha_1}) = O(n^{-1}),$$

which proves (2.3).

The result (2.4) will follow if we prove that

$$(2.9) \quad |C_2 \alpha_0 x^{\alpha_0-1} - C_1 \alpha_1 x^{\alpha_1-1}| \leq K x^{\alpha_1+\beta_1-1}$$

uniformly in $\delta < x \leq C_0^{-1/\alpha_0}$ and large n . By (2.7),

$$|C_2 \alpha_0 x^{\alpha_0-1} - C_0 \alpha_0 x^{\alpha_0-1}| \leq K n^{-(\beta_1+\alpha_1)/(2\beta_1+\alpha_1)} x^{\alpha_1-1} \leq K n^{-\alpha_1/(2\beta_1+\alpha_1)} x^{\alpha_1+\beta_1-1},$$

and so (2.9) will follow if we show that for $\delta < x \leq C_0^{-1/\alpha_0}$,

$$(2.10) \quad |C_0 \alpha_0 x^{\alpha_0-1} - C_1 \alpha_1 x^{\alpha_1-1}| \leq K x^{\alpha_1+\beta_1-1}.$$

But by (2.5),

$$(2.11) \quad \begin{aligned} &|C_0 \alpha_0 x^{\alpha_0-1} - C_1 \alpha_1 x^{\alpha_1-1}| \\ &\leq K_1 x^{\alpha_1-1} |C_0 \delta^{\alpha_1} x^{-\gamma} (\alpha_0 - \alpha_1) + \alpha_1 (\delta^\gamma x^{-\gamma} - 1)| \\ &\leq K_2 \delta^{\alpha_1} x^{\alpha_1-1} \gamma + K_2 x^{\alpha_1-1} |1 - (\delta/x)^\gamma| \\ &\leq K_3 \delta^{\alpha_1} x^{\alpha_1+\beta_1-1} + K_3 x^{\alpha_1-1} \gamma \log(x/\delta). \end{aligned}$$

Now, $x^{-\beta_1} \gamma \log(x/\delta) = (\delta/x)^{\beta_1} \log(x/\delta)$, and is maximised by taking $x/\delta = e^{1/\beta_1}$. Therefore by (2.11),

$$|C_0 \alpha_0 x^{\alpha_0-1} - C_1 \alpha_1 x^{\alpha_1-1}| \leq K_3 \delta^{\alpha_1} x^{\alpha_1+\beta_1-1} + K_4 x^{\alpha_1+\beta_1-1} \leq K_5 x^{\alpha_1+\beta_1-1}$$

uniformly in $\delta < x \leq C_0^{-1/\alpha_0}$. This proves (2.10), and completes the proof of (2.4).

From this point, our proof is inspired by Farrell (1972). Observe that

$$(2.12) \quad \begin{aligned} &P_{f_1} \{ |\alpha_n(X_1, \dots, X_n) - \alpha_1| \leq a_n \} \\ &= E_{f_0} [I \{ |\alpha_n(X_1, \dots, X_n) - \alpha_1| \leq a_n \} \prod_1^n \{ f_1(X_i)/f_0(X_i) \}] \\ &\leq [P_{f_0} \{ |\alpha_n(X_1, \dots, X_n) - \alpha_1| \leq a_n \}]^{1/2} \\ &\quad \cdot (E_{f_0} [\prod_1^n \{ f_1(X_i)/f_0(X_i) \}^2])^{1/2}, \end{aligned}$$

and

$$\begin{aligned} \left(E_{f_0} \left[\prod_1^n \left\{ \frac{f_1(X_i)}{f_0(X_i)} \right\}^2 \right] \right)^{1/n} &= \int_0^{C_0^{-1/\alpha_0}} \left\{ \frac{f_1^2(x)}{f_0(x)} \right\} dx \\ &= 1 + \int_0^{C_0^{-1/\alpha_0}} \{f_1(x) - f_0(x)\}^2 \{f_0(x)\}^{-1} dx \\ &= 1 + O(n^{-1}), \end{aligned}$$

using (2.3). Hence

$$(2.13) \quad \begin{aligned} P_{f_1}\{|\alpha_n(X_1, \dots, X_n) - \alpha_1| \leq a_n\} \\ \leq K[P_{f_0}\{|\alpha_n(X_1, \dots, X_n) - \alpha_1| \leq a_n\}]^{1/2}. \end{aligned}$$

By hypothesis and by the result (2.4), the left-hand side of (2.13) tends to 1 as $n \rightarrow \infty$. Therefore $P_{f_0}\{|\alpha_n(X_1, \dots, X_n) - \alpha_1| \leq a_n\}$ is bounded away from zero as $n \rightarrow \infty$. Also by hypothesis, $P_{f_0}\{|\alpha_n(X_1, \dots, X_n) - \alpha_0| \leq a_n\}$ tends to 1 as $n \rightarrow \infty$, and so

$$P_{f_0}[\{|\alpha_n(X_1, \dots, X_n) - \alpha_1| \leq a_n\} \cap \{|\alpha_n(X_1, \dots, X_n) - \alpha_0| \leq a_n\}]$$

is bounded away from zero. Consequently, for large n ,

$$|\alpha_1 - \alpha_0| \leq 2a_n;$$

that is, $\lambda n^{-\beta_1/(2\beta_1+\alpha_1)} \leq 2a_n$, and so

$$\liminf_{n \rightarrow \infty} n^{\beta_0/(2\beta_0+\alpha_0)} a_n \geq \lambda/2.$$

Since this is true for each $\lambda > 0$, Theorem 1 is proved.

PROOF OF THEOREM 2. The proof of Theorem 2 is very similar to that of Theorem 1, and uses the same density functions f_0, f_1 . Replace the left-hand side of (2.12) by

$$P_{f_1}\{|C_n(X_1, \dots, X_n) - C_1| \leq a_n\},$$

with similar changes at other places. Following the argument of the previous paragraph we see that for large n ,

$$|C_1 - C_0| \leq 2a_n.$$

But by (2.6),

$$C_1 - C_0 \sim -C_0\gamma \log \delta = C_0(2\beta_1 + \alpha_1)^{-1} \lambda n^{-\beta_0/(2\beta_0+\alpha_0)} \log n,$$

and so

$$\liminf_{n \rightarrow \infty} n^{\beta_0/(2\beta_0+\alpha_0)} (\log n)^{-1} a_n \geq C_0(2\beta_0 + \alpha_0)^{-1} \lambda/2,$$

for all $\lambda > 0$.

3. Achieving optimum rates. Let $X_{n1} < \dots < X_{nn}$ denote the ordered n -sample, and define r to be the integer part of $n^{2\beta_0/(2\beta_0+\alpha_0)}$,

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

These estimators are in fact conditional maximum likelihood estimators under a restricted model. They approximate Bayes estimators; see Hill (1975). The set \mathcal{D} has the same meaning as before, except that we assume $\varepsilon < \min(\alpha_0, C_0)$, to preclude the possibility that the parameters α and C governing densities in \mathcal{D} can take the value zero.

THEOREM 3. For each sequence $\{a_n\}$ satisfying

$$n^{\beta_0/(2\beta_0+\alpha_0)} a_n \rightarrow \infty$$

as $n \rightarrow \infty$, we have

$$\liminf_{n \rightarrow \infty} \inf_{f \in \mathcal{L}} P_f(|\alpha_n - \alpha| \leq a_n) = 1$$

whenever $\varepsilon < \min(\alpha_0, C_0)$ and $A > 0$.

THEOREM 4. For each sequence $\{a_n\}$ satisfying

$$n^{\beta_0/(2\beta_0+\alpha_0)}(\log n)^{-1}a_n \rightarrow \infty$$

as $n \rightarrow \infty$, we have

$$\liminf_{n \rightarrow \infty} \inf_{f \in \mathcal{D}} P_f(|C_n - C| \leq a_n) = 1$$

whenever $\varepsilon < \min(\alpha_0, C_0)$ and $A > 0$.

Theorems 3 and 4 may be proved by modifying arguments in Hall (1982), and so are not derived here.

We should note that Hill (1975) suggests r be chosen on an adaptive, data-analytic basis, while Hall (1982) considers deterministic values of r . In the latter case, optimal selection of r depends on the ratio β_0/α_0 . The problem of adaptive, "asymptotically optimal" approaches to choosing r will be the subject of a forthcoming paper.

REFERENCES

- DE HAAN, L. and RESNICK, S. J. (1980). A simple asymptotic estimate for the index of a stable distribution. *J. Roy. Statist. Soc. Ser. B* **42** 83–88.
- FARRELL, R. H. (1972). On the best obtainable asymptotic rates of convergence in estimation of a density function at a point. *Ann. Math. Statist.* **43** 170–180.
- HALL, P. (1982). On some simple estimates of an exponent of regular variation. *J. Roy. Statist. Soc. Ser. B* **44** 37–42.
- HILL, B. M. (1975). A simple approach to inference about the tail of a distribution. *Ann. Statist.* **3** 1163–1174.
- TEUGELS, J. L. (1981). Limit theorems on order statistics. *Ann. Probab.* **9** 868–880.
- WELSH, A. H. (1983). On the use of the empirical distribution and characteristic function to estimate parameters of regular variation. Unpublished manuscript.

DEPARTMENT OF STATISTICS
THE FACULTY OF ECONOMICS
THE AUSTRALIAN NATIONAL UNIVERSITY
G.P.O. BOX 4, CANBERRA A.C.T. 2601
AUSTRALIA