Centrum voor Wiskunde en Informatica
Centre for Mathematics and Computer Science

L. de Haan, A.L.M. Dekkers

Optimal choice of sample fraction in extreme-value estimation

Department of Operations Research, Statistics, and System Theory Report BS-R8927 November

The Centre for Mathematics and Computer Science is a research institute of the Stichting Mathematisch Centrum, which was founded on February 11, 1946, as a nonprofit institution aiming at the promotion of mathematics, computer science, and their applications. It is sponsored by the Dutch Government through the Netherlands Organization for the Advancement of Research (N.W.O.).

Optimal Choice of Sample Fraction in Extreme-Value Estimation

L. de Haan
Erasmus University Rotterdam
P.O. Box 1738, 3000 DR Rotterdam, The Netherlands

A.L.M. Dekkers
Centre for Mathematics and Computer Science
P.O. Box 4079, 1009 AB Amsterdam, The Netherlands

Estimators of the extreme-value index typically are functions of the k largest order statistics. Here $k=k(n)\to\infty$ and $k(n)/n\to 0$ $(n\to\infty)$. We consider the problem how to choose the number k in an (asymptotically) optimal way.

1980 Mathematics Subject Classification: 62E20, 62F10, 62F12
Key Words & Phrases: extreme-value index, estimation based on number of largest order statistics

1. Introduction

Suppose one is given a sequence $X_1, X_2, \ldots$ of i.i.d. observations from some distribution function F. Suppose for some constants $a_n>0$ and b_n and some $\gamma\in\mathbb{R}$

$$\lim_{n\to\infty} P\{\frac{\max(X_1,X_2,\ldots,X_n)-b_n}{a_n}\leqslant x\} = G_\gamma(x) \tag{1.1}$$

for all x where $G_\gamma(x)$ is one of the extreme-value distributions

$$G_\gamma(x) = \exp - (1+\gamma x)^{-1/\gamma}. \tag{1.2}$$

Here γ is a real parameter (interpret $(1+\gamma x)^{-1/\gamma}$ as e^{-x} for $\gamma=0$) and x such that $1+\gamma x>0$. We consider the estimators (based on the order statistics $\{X_{(i,n)}\}_{i=1}^n$)

$$\hat{\gamma}_n^{(1)} := (\log 2)^{-1}\log\frac{X_{(n-k+1,n)}-X_{(n-2k+1,n)}}{X_{(n-2k+1,n)}-X_{(n-4k+1,n)}} \tag{1.3}$$

and

$$\hat{\gamma}_n^{(2)} := M_n^{(1)} + 1 - 1/2\{1-(M_n^{(1)})^2/M_n^{(2)}\}^{-1}. \tag{1.4}$$

with for $j=1,2$

$$M_n^{(j)} := \frac{1}{k}\sum_{i=0}^{k-1}\{\log X_{(n-i,n)} - \log X_{(n-k,n)}\}^j \tag{1.5}$$

for γ. Dekkers and de Haan (1989) and Einmahl, Dekkers and de Haan (1989) proved that $\sqrt{k}(\hat{\gamma}_n^{(j)}-\gamma)$ has asymptotically a normal distribution (for $j=1,2$) under some conditions on $k=k(n)$ including $k(n)\to\infty, k(n)/n\to 0$ $(n\to\infty)$, provided the second order condition ("second order" relative to the domain of attraction condition)

$$\lim_{t\to\infty}\frac{(tx)^{1-\gamma}U'(tx)-t^{1-\gamma}U'(t)}{a(t)} = \pm\log x \text{ for all } x>0 \tag{1.6}$$

hold. Here $U := \{1/(1-F)\}^{\leftarrow}$ (inverse function).

If $k(n)$ is of smaller order than a given sequence $k_1(n)$ (depending on the distribution), then the

Report BS-R8927
Centre for Mathematics and Computer Science
P.O. Box 4079, 1009 AB Amsterdam, The Netherlands

asymptotic distribution of $\sqrt{k(n)}(\hat{\gamma}_n^{(j)}-\gamma)$ has mean zero. One would like to take $k(n)$ as large as possible since the asymptotic variance of $\hat{\gamma}_n^{(j)}$ decreases with $k(n)$. If $k(n)\sim c.k_1(n)$ $(n\to\infty)$ then there is an asymptotic bias depending on c (DEKKERS and DE HAAN: Remark 2.5 & EINMAHL, DEKKERS and DE HAAN: Remark at the end of section 3).

We consider the asymptotic second moment of $\hat{\gamma}_n^{(j)}-\gamma$ and choose $k(n)$ such that this expression is minimal by balancing the variance and bias parts. This value of k will be denoted by $k_0=k_0(n)$. It will be proved that under a third order condition for this optimal choice of k the bias is of larger order than the variance so that there is no longer asymptotic normality (Section 2). The given third order conditions are in general hard to verify in this case - much harder than the second order conditions since the latter can be given in terms of the distribution function while for the former a formulation in terms of the distribution function seems difficult. The point is that the given third order conditions are natural, so that the described behaviour is "usual".

In DEKKERS and DE HAAN (1989) and EINMAHL, DEKKERS and DE HAAN (1989) also an alternative second order condition is given for asymptotic normality of $\hat{\gamma}_n^{(j)}$ in case $\gamma\neq 0$. Since the full range of γ's is missing here anyway, we simplify the estimator $\hat{\gamma}_n^{(2)}$ under these conditions in two different ways according to $\gamma>0$ or $\gamma<0$. For $\gamma>0$ this leads to a variant of the well known Hill estimator. Here again we consider minimizing $E(\hat{\gamma}_n^{(j)}-\gamma)^2$ and prove that under a third order condition the optimal choice of k leads to an asymptotic normal distribution with known mean $\neq 0$. We show how to estimate the mean in a consistent way (Section 3).

Our results are illustrated by two examples: the normal and Cauchy distributions.

2. OPTIMAL CHOICE OF k
(first case)

2.a. *Pickands estimator*

The following theorem shows that under a natural strengthening of the conditions of th. 2.3 from DEKKERS and DE HAAN (1989), the optimal choice for $k(n)$ does not lead to a limiting distribution for $\hat{\gamma}_n^{(1)}$ that is useful for the construction of a confidence interval.

THEOREM 2.1 *Let* $R(x):=\log F^{\leftarrow}(1-e^{-x})$. *Suppose* (1.6) *holds,* $\lim_{t\to\infty} R''(t)/R'(t)=0$ *and*

$$\pm\frac{R''(\log t)}{R'(\log t)}\in\Pi. \tag{2.1}$$

Let $k_0(n)$ be the value minimizing the asymptotic second moment of $\hat{\gamma}_n^{(1)}-\gamma$ and let $\hat{\gamma}_{n,0}^{(1)}$ be the corresponding estimator. Then

$$\sqrt{k_0(n)}(\hat{\gamma}_{n,0}^{(1)}-\gamma)+b_n \tag{2.2}$$

has asymptotically a normal distribution with mean zero and variance $Z(\gamma):=\gamma^2(2^{2\gamma+1}+1)/\{2(2^\gamma-1)\log 2\}^2$, where b_n is an unbounded sequence of real numbers $(b_n=o(\sqrt{k_0(n)}),\ n\to\infty)$.

REMARK. Write

$$R(t+x)=R(t)+xR'(t)+\frac{x^2}{2}R''(t)+..$$

$R''(t)=o(R'(t))$ $(t\to\infty)$ is a natural condition for

$$\lim_{t\to\infty}\frac{R(t+x)-R(t)-xR'(t)}{R''(t)}=\frac{x^2}{2} \tag{2.3}$$

(DEKKERS and DE HAAN (1989), discussion before theorem A6). Now (2.3) implies $R'(\log t)\in\Pi(R''(\log t))$. Hence $R''(\log t)/R'(\log t)\in RV_0$. The present condition is slightly stronger. We shall show that the condition holds for the normal distribution.

PROOF. We shall give the proof for the upper sign in (2.1) and use the notation of DEKKERS and DE HAAN (1989) freely. The proof of theorem 2.3, DEKKERS and DE HAAN, contains the following statement:

$$2^{\hat{\gamma}_n^{(1)}} - 2^{\gamma} = \frac{2^{\gamma-1/2}Q_n/\sqrt{k} - 2^{-1}R_n/\sqrt{k} + (1+o(1))\frac{2^{\gamma-1}}{\gamma}(\log 2)\frac{\beta(\log\frac{n}{2k})}{V'(\log\frac{n}{2k})}}{(1-2^{-\gamma})/\gamma}.$$

Now $\hat{\gamma}_n^{(1)} - \gamma \sim (2^{\gamma}\log 2)^{-1}(2^{\hat{\gamma}_n^{(1)}} - 2^{\gamma})$, hence

$$\hat{\gamma}_n^{(1)} - \gamma \sim \frac{N_n}{\sqrt{k}} + (1+o(1)).\frac{\beta(\log\frac{n}{2k})}{V'(\log\frac{n}{2k})} \tag{2.4}$$

with N_n asymptotically $N(0, Z(\gamma))$. Hence the asymptotic second moment of $\hat{\gamma}_n^{(1)} - \gamma$ is

$$\frac{Z(\gamma)}{k} + \{\frac{\beta(\log\frac{n}{2k})}{V'(\log\frac{n}{2k})}\}^2 = \frac{r2Z(\gamma)}{n} \mp \tau(r)$$

with $r := n/(2k)$ and $\tau(r) := \{\beta(\log r)/V'(\log r)\}^2$. We consider

$$\tau_c(n) := \inf_r [\frac{r2Z(\gamma)}{n} + \tau(r)]. \tag{2.5}$$

Since $t^{-\gamma}\beta(t)$ is the auxiliary function of $t^{-\gamma}V'(t) \in \Pi$, we have $\lim_{t\to\infty}\beta(t)/V'(t) = 0$. Hence the infimum exists for sufficiently large values of n. The third order condition (2.1) implies that for some function $q > 0$

$$\lim_{t\to\infty}\frac{\tau(tx)-\tau(t)}{q(t)} = -\log x, \tag{2.6}$$

i.e. $-\tau \in \Pi$.

Now we proceed in a way similar as in GELUK and DE HAAN (1987), p. 61 sqq. with respect to the complementary convex function. When disregarding terms of lower order we may (and do) assume (GELUK and DE HAAN (1987) lemma 1.23) that τ has a continuous strictly decreasing derivative $s \in RV_{-1}$. Further it is straightforward to see that, when disregarding terms of lower order, in (2.5) one may as well take the infimum over *all* positive real values of r (instead of $2r/n \in \mathbb{N}$). We thus consider

$$\tau_c(n) := \inf_x \{\frac{x2Z(\gamma)}{n} + \int_x^{\infty} s(u)du\}$$

(cf. lemma 2.7, GELUK and DE HAAN (1987), p. 64). Hence the optimal choice of x is $s^{\leftarrow}(\frac{2Z(\gamma)}{n})$ and

$$\tau_c(n) = \frac{2s^{\leftarrow}(\frac{2Z(\gamma)}{n})Z(\gamma)}{n} + \int_{s^{\leftarrow}(\frac{2Z(\gamma)}{n})}^{\infty} s(u)du = \int_0^{\frac{2Z(\gamma)}{n}} s^{\leftarrow}(v)dv$$

From $s \in RV_{-1}$ it follows that

$$\lim_{t \downarrow 0} \frac{s^{\leftarrow}(tx)}{s^{\leftarrow}(t)} = x^{-1}$$

and hence for all $x > 0$

$$\lim_{t \to \infty} \frac{\tau_c(tx) - \tau_c(t)}{t^{-1} s^{\leftarrow}(t^{-1})} = -\log x.$$

Recall that $k_0(n) \sim n / \{2s^{\leftarrow}(\frac{2Z(\gamma)}{n})\}$ $(n \to \infty)$. Hence

$$\{\sqrt{k_0(n)} \frac{\beta(\log \frac{n}{2k_0(n)})}{V'(\log \frac{n}{2k_0(n)})}\}^2 \sim k_0(n). \int_{s^{\leftarrow}(\frac{2Z(\gamma)}{n})}^{\infty} s(u)du$$

$$= \{ \int_{s^{\leftarrow}(\frac{2Z(\gamma)}{n})}^{\infty} s(u)du\} / \{2s^{\leftarrow}(\frac{2Z(\gamma)}{n})/n\} = Z(\gamma) \frac{\int_y^{\infty} s(u)dy}{y\, s(y)} \text{ with } y := s^{\leftarrow}(\frac{2Z(\gamma)}{n}).$$

Since $s \in RV_{-1}$, the latter expression tends to infinity as $y \to \infty$ i.e. as $n \to \infty$ (cf. GELUK and DE HAAN (1987), remark 1 following Cor. 1.18). The proof is complete.

2.b. Moment estimator

THEOREM 2.2. *Suppose that the conditions of* th. 3.1, EINMAHL, DEKKERS and DE HAAN (1989) *hold and condition* (2.1). *Let* $k_0(n)$ *be the value minimizing the asymptotic second moment of* $\hat{\gamma}_n^{(2)} - \gamma$ *and let* $\hat{\gamma}_{n,0}^{(2)}$ *be the corresponding estimator. Then*

$$\sqrt{k_0(n)}(\hat{\gamma}_{n,0}^{(2)} - \gamma) + d_n$$

has asymptotically a normal distribution with variance

$$W(\gamma) := \begin{cases} 1, & \gamma \geq 0 \\ (1-\gamma)^2(1-2\gamma)\{4 - 8\frac{1-2\gamma}{1-3\gamma} + \frac{(5-11\gamma)(1-2\gamma)}{(1-3\gamma)(1-4\gamma)}\}, & \gamma < 0 \end{cases}$$

where d_n *is an unbounded sequence of real numbers* ($d_n = o(\sqrt{k_0(n)})$, $n \to \infty$).

PROOF. We shall give the proof for $\gamma = 0$ and the upper choice of the two signs in (2.1), DEKKERS and DE HAAN (1989) and (2.1), present paper. The other cases are similar. We shall use the notation of EINMAHL, DEKKERS and DE HAAN (1989) without comment. Since

$$\lim_{t \to \infty} \frac{\log U(tx) - \log U(t) - \dot{a}(t)\log x}{a_1(t)} = \frac{(\log x)^2}{2}, \tag{2.7}$$

we get as in the proof of theorem 3.1 (EINMAHL, DEKKERS and DE HAAN (1989))

$$M_n^{(1)} \overset{d}{=} a(Y_{(n-k,n)})\rho_1(0) + a(Y_{(n-k,n)})\{\frac{1}{k}\sum_{i=i}^{k-1} \log(\frac{Y_{(n-i,n)}}{Y_{(n-k,n)}}) - \rho_1(0)\} \tag{2.8}$$

$$+ a_1(Y_{(n-k,n)})\frac{\sigma_1(0)}{2} + a_1(\frac{Y_{(n-k,n)}}{2})\{\frac{1}{k}\sum_{i=0}^{k-1} [\log(\frac{Y_{(n-i,n)}}{Y_{(n-k,n)}})]^2 - \sigma_1(0)\}$$

plus a term of lower order where $\{Y_{(i,n)}\}_{i=1}^n$ are n-th order statistics from the distribution function $1 - 1/x$ $(x \geq 1)$, with $\rho_1(0) = 1$ and

$$\sigma_1(0) := \int_1^{\infty} (\log x)^2 \frac{dx}{x^2} = 2.$$

Now by the law of large numbers and $\{Y_{(n-i,n)}/Y_{(n-k,n)}\}_{i=0}^{k-1} \stackrel{d}{=} \{Y_{(k-i,k)}\}_{i=0}^{k-1}$, the last term of the right hand side of (2.8) is of lower order than the previous one. The second term, multiplied by $\sqrt{k}/a(Y_{(n-k,n)})$ is asymptotically normal and the third term represents the bias. Similarly, since by (2.7)

$$\{\log U(tx) - \log U(t)\}^2 \approx a^2(t)(\log x)^2 + a_1^2(t)\frac{(\log x)^4}{4}$$

$$+ 2a(t)a_1(t)\frac{(\log x)^3}{2} \approx a^2(t)(\log x)^2 + a(t)a_1(t)(\log x)^3$$

(disregarding terms of lower order), hence

$$M_n^{(2)} \stackrel{d}{=} \{a(Y_{(n-k,n)})\}^2 \rho_2(0) + \{a(Y_{(n-k,n)})\}^2 \{\frac{1}{k}\sum_{i=0}^{k-1} [\log(\frac{Y_{(n-i,n)}}{Y_{(n-k,n)}})]^2 - \rho_2(0)\}$$

$$+ a(Y_{(n-k,n)}).a_1(Y_{(n-k,n)})\sigma_2(0) + a(Y_{(n-k,n)}).a_1(Y_{(n-k,n)}).$$

$$\cdot\{\frac{1}{k}\sum_{i=0}^{k-1} [\log(\frac{Y_{(n-i,n)}}{Y_{(n-k,n)}})]^3 - \sigma_2(0)\}$$

plus terms of lower order, with $\rho_2(0) = 2$ and

$$\sigma_2(0) = \int_1^{\infty} (\log x)^3 \frac{dx}{x^2} = 3!$$

Again by the law of large numbers the last term is of lower order than the previous one. The second term is again connected with the asymptotic normality and the third represents the bias.

It follows that (disregarding terms of lower order and writing a for $a(Y_{(n-k,n)})$ etc.)

$$\hat{\gamma}_n^{(2)} := M_n^{(1)} + 1 - ½\{1-(M_n^{(1)})^2/M_n^{(2)}\}^{-1} \approx 1 - ½\{1-(M_n^{(1)})^2/M_n^{(2)}\}^{-1}$$

$$= \frac{M_n^{(2)}/2-(M_n^{(1)})^2}{M_n^{(2)}-(M_n^{(1)})^2} \approx \frac{½\{2a^2+a^2\cdot Q_n/\sqrt{k}+a.a_1 3!\}-\{a^2+2a^2P_n/\sqrt{k}+2a.a_1\}}{2a^2+a^2Q_n/\sqrt{k}+a.a_1 3!-a^2-2a^2P_n/\sqrt{k}-2a.a_1}$$

$$= \frac{½Q_n/\sqrt{k}-2P_n/\sqrt{k}+a_1/a}{1+Q_n/\sqrt{k}-2P_n/\sqrt{k}+4a_1/a} \approx Q_n/\{2\sqrt{k}\} - 2P_n/\sqrt{k}+a_1/a$$

$$= (\frac{Q_n}{2} - 2P_n)/\sqrt{k} + \frac{a_1(Y_{(n-k,n)})}{a(Y_{(n-k,n)})} \approx (\frac{Q_n}{2} - 2P_n)/\sqrt{k} + \frac{a_1(\frac{n}{k})}{a(\frac{n}{k})}.$$

The last (approximate) equality is valid since by (2.1) the function $a_1(t)/a(t)$ is the class Π and $\lim_{n\to\infty}\frac{k}{n}Y_{(n-k,n)} = 1$ in probability. Since $a \in \Pi(a_1)$, clearly

$$\lim_{n\to\infty} \frac{a_1(\frac{n}{k})}{a(\frac{n}{k})} = 0.$$

The rest of the proof is more or less the same as the proof of theorem 2.1.

3. Optimal choice of k
(second case)

a. $\gamma>0$: *Hill's estimator*

THEOREM 3.1. *Suppose (1.1) holds with* $\gamma>0$ *and for some* $\rho,c>0$ (cf. DEKKERS and DE HAAN 1989, Th. 2.5)

$$t^{1+1/\gamma}F'(t) \ - \ c\in\pm RV_{-\rho/\gamma}.$$

Determine $k_0(n)$ such that the asymptotic second moment of $M_n^{(1)}-\gamma$ is minimal. The optimal version of Hill's estimator will be called $M_{n,0}^{(1)}$. Then

$$\sqrt{k_0(n)}\,(M_{n,0}^{(1)}-\gamma)$$

is asymptotically normal with mean $\pm\dfrac{\gamma(\rho+1)}{2^{1/2}\rho^{3/2}}$ and variance γ^2.

REMARK. This theorem is related to CSÖRGÖ, DEHEUVELS and MASON (1985) and SMITH (1987).

PROOF. Suppose $t^{1+1/\gamma}\cdot F'(t)-c\in -RV_{-\rho/\gamma}$; the other case is similar. Define $c_1:=c^\gamma\gamma^\gamma$. Since for $x>0$ (cf. proof of Th. 2.5, DEKKERS and DE HAAN (1989))

$$\lim_{t\to\infty}\frac{\log U(tx)-\log U(t)-\gamma\log x}{\log U(t)-\gamma\log t-\log c_1} = 1-x^{-\rho},$$

we have (proceeding as in the proof of theorem 2.2)

$$M_n^{(1)}\stackrel{d}{=}\gamma\frac{1}{k}\sum_{i=0}^{k-1}\log(\frac{Y_{(n-i,n)}}{Y_{(n-k,n)}}) + [\log U(Y_{(n-k,n)}) - \gamma\log Y_{(n-k,n)} - \log c_1].$$

$$\cdot\{1-\frac{1}{k}\sum_{i=0}^{k-1}(\frac{Y_{(n-i,n)}}{Y_{(n-k,n)}})^{-\rho}\}$$

plus terms of smaller order, i.e. by the law of large numbers and

$$\{Y_{(n-i,n)}/Y_{(n-k,n)}\}_{i=0}^{k-1}\stackrel{d}{=}\{Y_{(k-i,k)}\}_{i=0}^{k-1}.$$

$$M_n^{(1)}-\gamma = \gamma\{\frac{1}{k}\sum_{i=0}^{k-1}\log(\frac{Y_{(n-i,n)}}{Y_{(n-k,n)}})-1\} + [\log U(\frac{n}{k})-\gamma\log\frac{n}{k}-\log c_1].$$

$$\cdot\int_1^\infty(1-x^{-\rho})\frac{dx}{x^2}$$

plus terms of smaller order

$$=:\gamma.P_n/\sqrt{k}+a(\frac{n}{k}).\frac{\rho}{\rho+1}$$

where P_n is asymptotically standard normal and $a(t):=\log U(t)-\gamma\log t-\log c\in RV_{-\rho}$. When neglecting terms of lower order we may assume (cf. GELUK and DE HAAN (1987) lemma 1.23) that a^2 has a continuous decreasing derivative $s\in RV_{-2\rho-1}$. We are interested in (with $r:=n/k$)

$$\rho_c(n):=\inf_r\{\frac{r\gamma^2}{n} + a^2(r)\frac{\rho^2}{(\rho+1)^2}\}$$

which is attained for $r=r_0=s^{\leftarrow}(\frac{\gamma^2(\rho+1)^2}{n.\rho^2})$. The infimum itself is

$$\rho_c(n) = \frac{\rho^2}{(\rho+1)^2}\{\frac{\gamma^2(\rho+1)^2}{n\rho^2}s^{\leftarrow}(\frac{\gamma^2(\rho+1)^2}{n\rho^2}) + \int_{s^{\leftarrow}(\frac{\gamma^2(\rho+1)^2}{n\rho^2})}^{\infty} s(u)du\} =$$

$$\frac{\rho^2}{(\rho+1)^2}\cdot \int_0^{\frac{\gamma^2(\rho+1)^2}{n\rho^2}} s^{\leftarrow}(u)du.$$

Further, with $k_0=n/r_0$,

$$\sqrt{k_0}(M_n^{(1)} - \gamma) = \gamma P_n + \tag{3.1}$$

$$\frac{\gamma(\rho+1)}{\rho}.[\{\int_{s^{\leftarrow}(\frac{\gamma^2(\rho+1)^2}{n\rho^2})}^{\infty} s(u)du\}/\{\frac{\gamma^2(\rho+1)^2}{n\rho^2}.s^{\leftarrow}(\frac{\gamma^2(\rho+1)^2}{n\rho^2})\}]^{1/2}.$$

Now $s\in RV_{-2\rho-1}$ hence

$$\lim_{y\to\infty}\frac{\int_y^{\infty} s(u)du}{y.s(y)} = (2\rho)^{-1}$$

i.e. the last term of the righthand side of (3.1) converges to $\frac{\gamma(\rho+1)}{2^{1/2}.\rho^{3/2}}$.
It follows that for the choice $k=k_0$

$$\sqrt{k_0}(M_n^{(1)} - \gamma)$$

is aymptotically normal with mean $\frac{\gamma(\rho+1)}{2^{1/2}.\rho^{3/2}}$ and variance γ^2. Next we show how to estimate the unknown parameter ρ.

THEOREM 3.2. *Suppose that the condition of theorem* 3.1 *are fulfilled. Define*

$$L_n := \frac{\log X_{(n-k,n)} - \log X_{(n-2k,n)}}{\log 2}.$$

Then

$$\lim_{n\to\infty}\frac{(M_n^{(1)})^2 - M_n^{(2)}/2}{M_n^{(1)} - L_n} = \frac{\gamma(3\rho+4)}{\rho+1}(1-\frac{1}{\log 2})^{-1}$$

in probability for sequences $k=k(n)$ satisfying $\sqrt{k}a(\frac{n}{k})\to\infty$ $(n\to\infty)$.

REMARK. Note that in theorem 2.3 of DEKKERS and DE HAAN (1989) the condition $\sqrt{k}a(\frac{n}{k})\to 0$ is used. Hence for estimating the bias many more observations should be used than for estimating γ. A similar situation seems to occur in density estimation.

PROOF. With the notation and convention of the previous proof we have

$$M_n^{(1)} \approx \gamma + \gamma P_n/\sqrt{k} + a(\frac{n}{k})\frac{\rho}{\rho+1},$$

$$(M_n^{(1)})^2 \approx \gamma^2 + 2\gamma^2 P_n/\sqrt{k} + 2.a(\tfrac{n}{k})\frac{\gamma\rho}{\rho+1},$$

$$M_n^{(2)} \approx 2\gamma^2 + \gamma^2 Q_n/\sqrt{k} + 2.a(\tfrac{n}{k}).\gamma\int_1^\infty (\log x)(1-x^{-\rho})\frac{dx}{x^2}.$$

Hence, since $\sqrt{k}.a(\frac{n}{k})\rightarrow\infty$,

$$\frac{(M_n^{(1)})^2 - M_n^{(2)}/2}{a(\frac{n}{k})} \rightarrow \gamma\{2\frac{\rho}{\rho+1} - \frac{1}{(\rho+1)^2}+1\} = \frac{\gamma\rho(3\rho+4)}{(\rho+1)^2}$$

in probability ($n\rightarrow\infty$). Moreover

$$L_n = \{\log U(\frac{Y_{(n-k,n)}}{Y_{(n-2k,n)}}.Y_{(n-2k,n)}) - \log U(Y_{(n-2k,n)})\}/(\log 2)$$

$$\approx \{\gamma\log\frac{Y_{(n-k,n)}}{Y_{(n-2k,n)}} + a(\tfrac{n}{k}).\int_1^\infty (1-x^{-\rho})\frac{dx}{x^2}\}/(\log 2)$$

$$\approx \gamma + \gamma R_n/\{\sqrt{2k}.\log 2\} + a(\tfrac{n}{k}).\frac{\rho}{\rho+1}/(\log 2)$$

with R_n asymptotically standard normal (DEKKERS and DE HAAN (1989) lemma 2.1). Hence, since $\sqrt{k}.a(\frac{n}{k})\rightarrow\infty$ ($n\rightarrow\infty$),

$$\frac{M_n^{(1)} - L_n}{a(\frac{n}{k})} \rightarrow \frac{\rho}{\rho+1}\{1-\frac{1}{\log 2}\}$$

in probability ($n\rightarrow\infty$).

b. $\gamma<0$: *Moment estimator*

THEOREM 3.3. *Suppose* $\gamma<0$ *and for some* $\rho_0>0$

$$\pm\{t^{-1-1/\gamma}.F'(U(\infty)-t^{-1})-c_0\}\in RV_{-\rho_0}.$$

Determine $k_0(n)$ such that the asymptotic second moment of $\hat{\gamma}_n^{(3)}-\gamma$ is minimal. Here

$$\hat{\gamma}_n^{(3)} := 1-1/2\{1-(H_n^{(1)})^2/H_n^{(2)}\}^{-1}$$

with for $j=1,2$

$$H_n^{(j)} := \frac{1}{k}\sum_{i=0}^{k-1}\{X_{(n-i,n)} - X_{(n-k,n)}\}^j.$$

Denote the optimal version of $\hat{\gamma}_n^{(3)}$ by $\hat{\gamma}_{n,0}^{(3)}$. Then

$$\sqrt{k_0(n)}\,(\hat{\gamma}_{n,0}^{(3)}-\gamma)$$

is asymptotically normal with mean

$$\frac{c_0^{-\gamma}(-\gamma)^{-\gamma+1}}{(-2\rho_0\gamma)^{1/2}}\{\frac{1}{(1-\gamma)\gamma^2}[1-\frac{1}{1-\gamma}-\frac{1}{1-\gamma+\rho}+\frac{1}{1-2\gamma+\rho}] + $$
$$- \frac{2(\rho-\gamma)}{(1-\gamma)(1-2\gamma)(-\gamma)(1-\gamma+\rho)}\}$$

and variance

$$(1-\gamma)^2(1-2\gamma)\{4-8\frac{1-2\gamma}{1-3\gamma} + \frac{(5-11\gamma)(1-2\gamma)}{(1-3\gamma)(1-4\gamma)}\}.$$

PROOF. Suppose $\{t^{-1-1/\gamma}F'(U(\infty)-t^{-1})-c_0\}\in RV_{-\rho_0}$.
Then $t^{1-\gamma}U'(t)-c_0^{-\gamma}(-\gamma)^{-\gamma+1}\in RV_{\rho_0\gamma}$ and hence $t^{-\gamma}\{U(\infty)-U(t)\}-c_0^{-\gamma}(-\gamma)^{-\gamma}\in RV_{\rho_0\gamma}$. It follows that

$$\lim_{t\to\infty}\frac{U(tx)-U(t)-ct^{+\gamma}(1-x^{\gamma})}{t^{\gamma}a(t)} = 1-x^{-\rho}x^{\gamma}$$

for $x>0$ where $c:=c_0^{-\gamma}(-\gamma)^{-\gamma+1}, \rho:=-\rho_0\gamma$ and $a(t):=t^{-\gamma}\{U(\infty)-U(t)\}-c$. Hence

$$\frac{H_n^{(1)}}{(-\gamma)(\frac{n}{k})^{\gamma}} \approx \frac{c}{(-\gamma)}\frac{1}{k}\sum_{i=0}^{k-1}\{1-(\frac{Y_{(n-i,n)}}{Y_{(n-k,n)}})^{\gamma}\}$$

$$+ \frac{a(\frac{n}{k})}{-\gamma}\frac{1}{k}\sum_{i=0}^{k-1}\{1-(\frac{Y_{(n-i,n)}}{Y_{(n-k,n)}})^{\gamma-\rho}\}$$

$$\approx c.\rho_1(\gamma) + \frac{c}{-\gamma}\cdot\frac{P_n}{\sqrt{k}} + a(\frac{n}{k})\sigma_1(\gamma,\rho)$$

with $\rho_1(\gamma):=\frac{1}{1-\gamma}$,

$$\sigma_1(\gamma,\rho):=\frac{\rho-\gamma}{(-\gamma)(1-\gamma+\rho)},$$

$$P_n:=\sqrt{k}\{\frac{1}{k}\sum_{i=0}^{k-1}\{1-(\frac{Y_{(n-i,n)}}{Y(n-k,n)})^{\gamma}\}-\rho_1(\gamma)\}.$$

We also need

$$\frac{\{H_n^{(1)}\}^2}{c^2\{-\gamma(\frac{n}{k})^{\gamma}\}^2} \approx (\rho_1(\gamma))^2 + 2\frac{\rho_1(\gamma)}{-\gamma}.\frac{P_n}{\sqrt{k}} + 2c^{-1}\rho_1(\gamma)\sigma_1(\gamma,\rho)a(\frac{n}{k}). \quad (3.2)$$

Similarly

$$\frac{H_n^{(2)}}{c^2\{-\gamma(\frac{n}{k})^{\gamma}\}^2} \approx \rho_2(\gamma) + \frac{Q_n}{\gamma^2.\sqrt{k}} + 2c^{-1}.a(\frac{n}{k})\sigma_2(\gamma,\rho) \quad (3.3)$$

with $\rho_2(\gamma):=\frac{2}{(1-\gamma)(1-2\gamma)}$,

$$\sigma_2(\gamma,\rho):=\frac{1}{\gamma^2}\{1-\frac{1}{1-\gamma}-\frac{1}{1-\gamma+\rho}+\frac{1}{1-2\gamma+\rho}\}$$

$$Q_n:=\sqrt{k}\{\frac{1}{k}\sum_{i=0}^{k-1}(1-(\frac{Y_{(n-i,n)}}{Y_{(n-k,n)}})^{\gamma})^2 - \rho_2(\gamma)\}.$$

Combination gives as in the proof of EINMAHL, DEKKERS and DE HAAN (1989), Cor. 3.2

$$\hat{\gamma}_n^{(3)} - \gamma \approx \rho_1(\gamma)\{\rho_2(\gamma) - (\rho_1(\gamma))^2\}^{-2}[\frac{\rho_1(\gamma)}{2}\frac{Q_n}{\gamma^2\sqrt{k}} - \rho_2(\gamma)\frac{P_n}{(-\gamma)\sqrt{k}}$$

$$+ c^{-1}a(\frac{n}{k})\{\rho_1(\gamma)\sigma_2(\gamma,\rho) - \rho_2(\gamma)\sigma_1(\gamma,\rho)\}].$$

Hence the asymptotic second moment of $\hat{\gamma}_n - \gamma$ is

$$\frac{c_1}{k} + c_2\tau(\frac{n}{k})$$

with $\tau(t) := \{a(t)\}^2$

$$c_1 := (1-\gamma)^2(1-2\gamma)\{4 - 8\frac{1-2\gamma}{1-3\gamma} + \frac{(5-11\gamma)(1-2\gamma)}{(1-3\gamma)(1-4\gamma)}\}.$$

$$c_2 := c^{-2}\{\rho_1(\gamma)\sigma_2(\gamma,\rho) - \rho_2(\gamma)\sigma_1(\gamma,\rho)\}^2.$$

We are interested in (with $r := n/k$)

$$\tau_c(n) := \inf_r\{\frac{rc_1}{n} + c_2\tau(r)\}.$$

We now proceed as in the proof of theorem 3.1. Since $\tau \in RV_{-2\rho}(\rho>0)$, we can assume (neglecting terms of smaller order) that τ has a continuous derivative $-s \in RV_{-2\rho-1}$ (cf. GELUK and DE HAAN (1987) Lemma 1.23). The optimum is then attained for $r_0 = s^{\leftarrow}(\frac{c_1}{nc_2})$ i.e. $k_0 = n/s^{\leftarrow}(\frac{c_1}{nc_2})$, and

$$\sqrt{k_0}(\hat{\gamma}_n^{(3)} - \gamma) \approx \rho_1(\gamma)\{\rho_2(\gamma) - (\rho_1(\gamma))^2\}^{-2}[\frac{\rho_1(\gamma)}{2\gamma^2}Q_n - \frac{\rho_2(\gamma)}{(-\gamma)}P_n]$$

$$+ \sqrt{c_2}[\int_{s^{\leftarrow}(\frac{c_1}{nc_2})}^{\infty} s(u)du/\{\frac{c_1}{nc_2}.s^{\leftarrow}(\frac{c_1}{nc_2})\}]^{1/2}.$$

Now $s \in RV_{-2\rho-1}$, hence

$$\lim_{y\to\infty}\frac{\int_y^{\infty} s(u)du}{y.s(y)} = \frac{1}{2\rho}$$

i.e. the last term at the righthandside of (3.1) converges to $(\frac{c_2}{2\rho})^{1/2}$. It remains to estimate the parameters c and ρ figuring in the expression for c_2.

THEOREM 3.4. *Under the conditions of theorem* 3.3

$$\lim_{n\to\infty}\frac{H_n^{(1)}}{(\frac{n}{k})^{\gamma}} = c.\frac{-\gamma}{1-\gamma} \tag{3.4}$$

in probability and

$$\lim_{n\to\infty}\frac{(H_n^{(1)})^2 - (\rho_1(\gamma))^2c^2\{-\gamma(\frac{n}{k})^{\gamma}\}^2}{H_n^{(2)} - \rho_2(\gamma)c^2\{-\gamma(\frac{n}{k})^{\gamma}\}^2} = \frac{\sigma_2(\gamma,\rho)}{\rho_1(\gamma)\sigma_1(\gamma,\rho)} \tag{3.5}$$

in probability.

PROOF. The statements follow directly from (3.2) and (3.3).

4. APPLICATIONS

Normal distribution.

Let $F(x):=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-t^2/2}dt$ and write $P:=-\log(1-F)$. Let Q be the inverse function of P. One sees from BALKEMA and DE HAAN (1988), Proposition 4.2 or by direct computation that

$$Q'(t) \sim \frac{1}{Q(t)},$$
$$Q''(t) \sim \frac{-1}{\{Q(t)\}^3},$$
$$Q'''(t) \sim \frac{3}{\{Q(t)\}^5},\ t\to\infty$$

(in general $Q^{(j)}(t)\sim(-1)^{j+1}(2j-3)(2j-5)...1/\{Q(t)\}^{2j-1}$, $t\to\infty$, $j=1,2,...$). Now $R=\log Q$ hence

$$R''(t) = \frac{Q''(t)Q(t)-\{Q'(t)\}^2}{\{Q(t)\}^2} \sim \frac{-2}{\{Q(t)\}^4}\ (t\to\infty)$$

and

$$R'''(t) = \frac{Q'''(t)\{Q(t)\}^3 - 3Q''(t)Q'(t)\{Q(t)\}^2 + 2\{Q'(t)\}^3Q(t)}{\{Q(t)\}^4} \sim \frac{7}{\{Q(t)\}^6}$$

$(t\to\infty)$.
Since clearly

$$\lim_{t\to\infty} \frac{(\frac{R''}{R'})'(t+x)}{(\frac{R''}{R'})'(t)} = 1$$

for $x\in\mathbb{R}$, it follows that $\frac{R''(\log x)}{R'(\log x)} \in\Pi$.
Note that

$$k_0(n)\sim\frac{2}{3\sqrt{3}}\ (\log n)^3$$

for the optimal sequence k_0 in this case.

Cauchy distribution.

For Cauchy's distribution

$$F(x):=1/2 + \frac{1}{\pi}\text{arctg}\ x$$

we have

$$t^2F'(t) = \frac{1}{\pi}\frac{t^2}{1+t^2},$$

i.e.

$$\frac{1}{\pi} - t^2F'(t) \sim \frac{1}{\pi t^2}.$$

Hence the conditions of theorem 3.1 are satisfied with $\rho=2$.

References

A.A. Balkema and L. de Haan (1989). Convergence rates in extreme-value theory. Submitted.

S. Csörgö, P. Deheuvels and D. Mason (1985). Kernel estimates of the tail index of a distribution. *Annals of Statistics* **13**, 1050-1077.

A.L.M. Dekkers and L. de Haan (1989). On the estimation of the extreme-value index and large quantile estimation. Annals of Statistics (to appear).

A.L.M. Dekkers, J.H.J. Einmahl and L. de Haan (1989). A moment estimator for the index of an extreme-value distribution. Annals of Statistics (to appear).

J.L. Geluk and L. de Haan (1987). Regular variation, extension and Tauberian theorems. C.W.I. Tract 40. Centrum voor wiskunde en informatica / Mathematisch Centrum, P.O. Box 4079, 1009 AB Amsterdam.

R.L. Smith (1987). Estimating tails of probability distributions. *Ann. Stat.* 15, 1174-1207.