

Asymptotic normality of least-squares estimators of tail indices

SÁNDOR CSÖRGŐ¹ and LÁSZLÓ VIHAROS^{2*}

¹*Department of Statistics, University of Michigan, Ann Arbor, MI 48109-1027, USA*

²*Bolyai Institute, University of Szeged, Aradi vértanúk tere 1, 6720 Szeged, Hungary*

Based on least-squares considerations, Schultze and Steinebach proposed three new estimators for the tail index of a regularly varying distribution function and proved their consistency. We show that, unlike the Hill estimator, all three least-squares estimators can be centred to have normal asymptotic distributions universally over the whole model, and for two of these estimators this in fact happens at the desirable order of the norming sequence. We analyse the conditions under which asymptotic confidence intervals become possible. In a submodel, we compare the asymptotic mean square errors of optimal versions of these and earlier estimators. The choice of the number of extreme order statistics to be used is also discussed through the investigation of the asymptotic mean square error for a comprehensive set of examples of a general kind.

Keywords: asymptotic confidence intervals; asymptotic mean square errors; least-squares estimators; tail index; universal asymptotic normality

1. Introduction and main results

Let X, X_1, X_2, \dots be independent random variables with a common distribution function $F(x) = P\{X \leq x\}$, $x \in \mathbb{R}$, and for each integer $n \geq 1$, let $X_{1,n} \leq \dots \leq X_{n,n}$ denote the order statistics based on the sample X_1, \dots, X_n . Let \mathcal{R}_α be the class of all distribution functions F such that $1 - F$ is regularly varying at infinity with index $-1/\alpha$, i.e.

$$1 - F(x) = x^{-1/\alpha} l(x), \quad 0 < x < \infty, \quad (1.1)$$

where $l(\cdot)$ is an unknown nuisance function slowly varying at infinity and $\alpha > 0$ is a fixed unknown parameter to be estimated. Introducing $Q(s) := \inf \{x: F(x) \geq s\}$, $0 < s \leq 1$, $Q(0) := Q(0+)$, the inverse or quantile function of F and letting $Q(1-s-)$ denote the left-continuous version of the right-continuous function $Q(1-s)$, $0 < s < 1$, it is well known that $F \in \mathcal{R}_\alpha$ if and only if, for some function $L(\cdot)$ slowly varying at zero,

$$Q(1-s-) = s^{-\alpha} L(s), \quad 0 < s < 1. \quad (1.2)$$

Several estimators have been proposed, but a considerable part of the large literature is

*To whom correspondence should be addressed. e-mail: viharos@math.u-szeged.hu.

centred around the asymptotic properties of Hill's (1975) estimator $\hat{\alpha}_n^{(0)} := \hat{\alpha}_n^{(0)}(k_n) := k_n^{-1} \sum_{i=1}^{k_n} \log^* X_{n+1-i,n} - \log^* X_{n-k_n,n}$, where the k_n are some integers satisfying

$$1 \leq k_n < n, \quad k_n \rightarrow \infty, \quad k_n/n \rightarrow 0 \text{ as } n \rightarrow \infty, \quad (1.3)$$

and $\log^* x = \log \max(x, 1)$, $x \in \mathbb{R}$. In particular, the basic weak- and strong-consistency results are due to Mason (1982) and Deheuvels *et al.* (1988), while the difficult problem of the asymptotic normality of $\hat{\alpha}_n^{(0)}$ has been investigated by Hall (1982), Hall and Welsh (1985), Haeusler and Teugels (1985), Csörgő and Mason (1985), Beirlant and Teugels (1989) and their references, and by Csörgő *et al.* (1985) for a whole class of kernel estimators containing $\hat{\alpha}_n^{(0)}$. Asymptotic normality here refers to that of $k_n^{1/2} \{\hat{\alpha}_n^{(0)} - \mu_n^{(0)}\}$, where $\mu_n^{(0)} = \mu_n^{(0)}(k_n)$ is a numerical centring sequence such that $\mu_n^{(0)} \rightarrow \alpha$ as $n \rightarrow \infty$. In some restrictions of \mathcal{R}_α one has $k_n^{1/2} \{\mu_n^{(0)} - \alpha\} \rightarrow 0$, so that it is possible to centre $\hat{\alpha}_n^{(0)}$ at α in these statements. However, even with varying $\mu_n^{(0)}$, these results all use restrictive conditions on \mathcal{R}_α itself, and the various sets of conditions are difficult to compare. Bringing the approach of Csörgő and Mason (1985) to what appears to be its ultimate limit, Csörgő and Viharos (1995) proved even more general results, at the price of changing even the norming sequence $k_n^{1/2}$ to sequences depending also on the unknown L in (1.2) in complicated ways. However, we also constructed there distribution functions $F \in \mathcal{R}_\alpha$ such that $\hat{\alpha}_n^{(0)}(\lfloor n^{2/3} \rfloor)$, where $\lfloor \cdot \rfloor$ denotes integer part, does not have a non-degenerate asymptotic distribution for *any* centring and norming sequences. So, while the weak consistency of $\hat{\alpha}_n^{(0)}(k_n)$ for all $\{k_n\}$ satisfying (1.3) in fact characterizes the classes \mathcal{R}_α by Mason's (1982) theorem, the Hill estimator is *not* universally asymptotically normal over \mathcal{R}_α .

From analytic considerations about L , Viharos (1997) has derived two new classes of estimators such that, denoting a member of any one of them by $\hat{\alpha}_n^\dagger(k_n)$, for suitable $\mu_n^\dagger(k_n) \rightarrow \alpha$ the sequence $k_n^{1/2} \log(n/k_n) \{\hat{\alpha}_n^\dagger(k_n) - \mu_n^\dagger(k_n)\}$ is universally asymptotically normal over \mathcal{R}_α for all $\{k_n\}$ satisfying (1.3) for one of the classes and whenever, additionally, $k_n/\log^2 n \rightarrow \infty$ for the other. Thus the price of universal asymptotic normality for both classes is the presence of the factor $\log(n/k_n)$, indicating a possibly greater order of bias than that of $\hat{\alpha}_n^{(0)}$ (at least when a measure for the latter is available).

For simplicity of notation, we assume without loss of generality from now on that $F(0) = 0$ for all $F \in \mathcal{R}_\alpha$, i.e., X is positive; otherwise one only has to replace \log by \log^* in what follows with some trivial extra reasonings in the proofs. Approximating a Bayes estimator, Hill's (1975) estimator $\hat{\alpha}_n^{(0)}$ is in fact a conditional maximum likelihood estimator under the restricted model in which $l(x) = e^c$ in (1.1), for all x beyond a threshold, for some constant c . Recently, Schultze and Steinebach (S&S) (1996) proposed three new estimators of α , which are based on least-squares considerations in the same restricted model. (In a mathematically equivalent fashion, they in fact do this in a corresponding exponential model briefly touched upon in the next section.) Taking the logarithm of (1.1) in the restricted case, substituting $x = X_{n+1-i,n}$ into $-\log\{1 - F(x)\} = -c + \alpha^{-1} \log x$ and approximating the left-hand sides by $-\log\{1 - \hat{F}_n(X_{n+1-i,n})\} = \log(n/i)$, where $\hat{F}_n(\cdot)$ is the sample distribution function, for some $1 \leq k_n < n$ we have either $\log X_{n+1-i,n} \approx d + \alpha \log(n/i)$, $i = 1, \dots, k_n$, where $d = \alpha c$, or $\alpha^{-1} \log X_{n+1-i,n} \approx c + \log(n/i)$, $i = 1, \dots, k_n$. A least-squares fit based on the first set of approximative equations gives the first estimator $\hat{\alpha}_n^{(1)} = \hat{\alpha}_n^{(1)}(k_n)$, which thus results from minimizing $\sum_{i=1}^{k_n} \{\log X_{n+1-i,n} - d -$

$\alpha \log(n/i)\}^2$ in the two variables α and d . The univariate minimization problem with $d = c = 0$ yields the second estimator $\hat{\alpha}_n^{(2)} = \hat{\alpha}_n^{(2)}(k_n)$; so this belongs to $I(\cdot) \equiv 1$. Finally, the second set of approximative equations gives the third estimator $\hat{\alpha}_n^{(3)} = \hat{\alpha}_n^{(3)}(k_n)$; this results from minimizing $\sum_{i=1}^{k_n} \{\alpha^{-1} \log X_{n+1-i,n} - c - \log(n/i)\}^2$ in the two variables α^{-1} and c . Therefore, the S&S estimators are as follows:

$$\begin{aligned} \hat{\alpha}_n^{(1)}(k_n) &:= \left[\frac{1}{k_n} \sum_{i=1}^{k_n} \left\{ \log \left(\frac{n}{i} \right) \right\} (\log X_{n+1-i,n}) - \frac{1}{k_n^2} \left(\sum_{i=1}^{k_n} \log X_{n+1-i,n} \right) \left\{ \sum_{i=1}^{k_n} \log \left(\frac{n}{i} \right) \right\} \right] \\ &\quad \times \left[\frac{1}{k_n} \sum_{i=1}^{k_n} \log^2 \left(\frac{n}{i} \right) - \left\{ \frac{1}{k_n} \sum_{i=1}^{k_n} \log \left(\frac{n}{i} \right) \right\}^2 \right]^{-1}, \\ \hat{\alpha}_n^{(2)}(k_n) &:= \left[\sum_{i=1}^{k_n} \left\{ \log \left(\frac{n}{i} \right) \right\} (\log X_{n+1-i,n}) \right] \left[\sum_{i=1}^{k_n} \log^2 \left(\frac{n}{i} \right) \right]^{-1}, \\ \hat{\alpha}_n^{(3)}(k_n) &:= \left\{ \frac{1}{k_n} \sum_{i=1}^{k_n} \log^2 X_{n+1-i,n} - \left(\frac{1}{k_n} \sum_{i=1}^{k_n} \log X_{n+1-i,n} \right)^2 \right\} \\ &\quad \times \left[\frac{1}{k_n} \sum_{i=1}^{k_n} \left\{ \log \left(\frac{n}{i} \right) \right\} (\log X_{n+1-i,n}) - \frac{1}{k_n^2} \left(\sum_{i=1}^{k_n} \log X_{n+1-i,n} \right) \left\{ \sum_{i=1}^{k_n} \log \left(\frac{n}{i} \right) \right\} \right]^{-1}. \end{aligned}$$

As is usual for $\hat{\alpha}_n^{(0)}$, we now drop the restricted model and thus return to the whole \mathcal{R}_α . Using results of Csörgő *et al.* (1985), Deheuvels *et al.* (1988), Csörgő *et al.* (1991), Lo (1989) and Viharos (1993), S&S proved the consistency of these, for all $\{k_n\}$ satisfying (1.3) for $\hat{\alpha}_n^{(2)}$, and under the additional assumption that $k_n/\log^2 n \rightarrow \infty$ for $\hat{\alpha}_n^{(1)}$ and $\hat{\alpha}_n^{(3)}$, supposing also the continuity of Q near 1 for the result on $\hat{\alpha}_n^{(3)}$.

Let \rightarrow denote convergence in distribution and let $N(\mu, \sigma^2)$ be the normal distribution with mean $\mu \in \mathbb{R}$ and variance $\sigma^2 > 0$. Understanding limiting and order relations and asymptotic equalities \sim as $n \rightarrow \infty$ throughout if not specified otherwise, our main results in this paper are contained in the following theorems.

Theorem 1.1. *If k_n is any sequence of positive integers such that (1.3) holds and $k_n/\log^4 n \rightarrow \infty$, then, whenever $F \in \mathcal{R}_\alpha$ for some $\alpha \in (0, \infty)$,*

$$k_n^{1/2} \{ \hat{\alpha}_n^{(1)}(k_n) - \mu_n^{(1)}(k_n) \} \xrightarrow{\mathcal{L}} N(0, 2\alpha^2),$$

where $\mu_n^{(1)}(k_n) = \mu_n^{(1)} := -(n/k_n) \int_0^{k_n/n} \log Q(1-t) \{1 + \log(nt/k_n)\} dt \rightarrow \alpha$.

Theorem 1.2. *If k_n is any sequence of positive integers such that (1.3) holds and $\{k_n \log^2(n/k_n)\}/\log^4 n \rightarrow \infty$, then, whenever $F \in \mathcal{R}_\alpha$ for some $\alpha \in (0, \infty)$,*

$$k_n^{1/2} \log \left(\frac{n}{k_n} \right) \{ \hat{\alpha}_n^{(2)}(k_n) - \mu_n^{(2)}(k_n) \} \xrightarrow{\mathcal{L}} N(0, 2\alpha^2),$$

where $\mu_n^{(2)}(k_n) = \mu_n^{(2)} := \mu_{2,n}/j_n$, with $\mu_{2,n} = \mu_{2,n}(k_n) := - \int_0^{k_n/n} \log Q(1-t) \log t \, dt$ and $j_n = j_n(k_n) := \int_0^{k_n/n} \log^2 t \, dt = (k_n/n) \{ \log^2(n/k_n) + 2 \log(n/k_n) + 2 \}$, and $\mu_n^{(2)}(k_n) \rightarrow \alpha$.

Theorem 1.3. *If k_n is any sequence of positive integers such that (1.3) holds and $k_n/\log^4 n \rightarrow \infty$, then, whenever $F \in \mathcal{R}_\alpha$ for some $\alpha \in (0, \infty)$,*

$$k_n^{1/2} \{ \hat{\alpha}_n^{(3)}(k_n) - \mu_n^{(3)}(k_n) \} \xrightarrow{\mathcal{L}} N(0, 2\alpha^2),$$

where $\mu_n^{(3)}(k_n) = \mu_n^{(3)} := \{ \mu_{4,n} - \mu_{5,n}^2 \} / \mu_n^{(1)}$, with $\mu_{5,n} = \mu_{5,n}(k_n) := (n/k_n) \int_0^{k_n/n} \log Q(1-t) \, dt$ and $\mu_{4,n} = \mu_{4,n}(k_n) := (n/k_n) \int_0^{k_n/n} \log^2 Q(1-t) \, dt$ and $\mu_n^{(3)}(k_n) \rightarrow \alpha$.

Thus all three S&S estimators are universally asymptotically normal over the whole model \mathcal{R}_α , provided that $k_n/\log^4 n \rightarrow \infty$. The behaviour of $\hat{\alpha}_n^{(2)}$ is analogous to that of Viharos’s (1996) estimators. We do not know whether the theoretically weak size requirements on k_n are necessary. The universal asymptotic normality of $\hat{\alpha}_n^{(1)}$ and $\hat{\alpha}_n^{(3)}$ with the desired norming factor $k_n^{1/2}$ appears to be the first such results for the problem.

The basic practical problem is of course the choice of k_n for a given n . While we do not believe that this problem can be solved ‘universally’, it is part of the statistical nature of the S&S least-squares estimators that heuristically appealing methods for the data-driven choice of k_n are readily available, at least for $\hat{\alpha}_n^{(1)}(k_n)$ and $\hat{\alpha}_n^{(3)}(k_n)$. Indeed, S&S suggested two such methods: one is based on minimizing a sum of squared residuals, and the other on maximizing a corresponding sample correlation coefficient. As a result of their careful simulation study, they conclude that the resulting $\hat{\alpha}_n^{(1)}$ and $\hat{\alpha}_n^{(3)}$ compare fairly well with $\hat{\alpha}_n^{(0)}$. The theorems above may provide the theoretical framework to investigate optimality properties of these methods. A theoretical aspect of the problem of choosing k_n is considered in Section 4, where we suggest a heuristic rule of thumb, while the conditions under which asymptotic confidence intervals for α become possible are discussed in Section 3. The proofs are in Section 5.

2. The exponential model

For a given $\alpha \in (0, \infty)$, let \mathcal{E}_α denote the class of all distribution functions G on \mathbb{R} such that $1 - G(x) = e^{-x/\alpha} r(x)$, $0 < x < \infty$, where $r(\cdot)$ is a function regularly varying at infinity. Let Z, Z_1, Z_2, \dots be independent random variables with distribution function G and quantile function $S(s) = \inf \{ x: G(x) \geq s \}$, $0 < s \leq 1$, and let $Z_{1,n} \leq \dots \leq Z_{n,n}$ be the order statistics pertaining to Z_1, \dots, Z_n . Then $G \in \mathcal{E}_\alpha$ if and only if $F \in \mathcal{R}_\alpha$, where $F(x) = P \{ e^Z \leq x \}$, $x \in \mathbb{R}$, so that $F(x) = G(\log x)$, $x > 0$ (Csörgő and Mason 1985). Hence, whenever $G \in \mathcal{E}_\alpha$, *Theorems 1.1, 1.2 and 1.3 all remain valid if we replace $\log X_{n+1-i,n}$ by $Z_{n+1-i,n}$, $i = 1, \dots, k_n$, and $\log Q(\cdot)$ by $S(\cdot)$.* S&S discussed interesting examples from insurance and risk theory which give rise to \mathcal{E}_α .

3. The centering problem: asymptotic confidence intervals

Having Theorems 1.1 and 1.3, a natural question concerns the conditions under which

$$k_n^{1/2} \left(\frac{\mu_n^{(j)}(k_n)}{\alpha} - 1 \right) \rightarrow b, \tag{3.1}$$

and so

$$k_n^{1/2} \{ \hat{\alpha}_n^{(j)}(k_n) - \alpha \} \xrightarrow{\mathcal{D}} N(b\alpha, 2\alpha^2), \quad j = 1, 3,$$

provided that $k_n/\log^4 n \rightarrow \infty$, for some limiting bias $b \in \mathbb{R}$ scaled by α . If Φ denotes the standard normal distribution function and $z_\varepsilon(b) > 0$ is the unique value of z for which $\Phi((z - b)/2^{1/2}) - \Phi(-(z + b)/2^{1/2}) = 1 - \varepsilon$ for a given $\varepsilon \in (0, 1)$, then of course (3.1) implies that

$$P \left(\hat{\alpha}_n^{(j)}(k_n) - \frac{z_\varepsilon(b)\hat{\alpha}_n(k_n)}{k_n^{1/2}} \leq \alpha \leq \hat{\alpha}_n^{(j)}(k_n) + \frac{z_\varepsilon(b)\hat{\alpha}_n(k_n)}{k_n^{1/2}} \right) \rightarrow 1 - \varepsilon, \quad j = 1, 3, \tag{3.2}$$

for any consistent estimator $\hat{\alpha}_n(k_n)$ of α . The corresponding question for Theorem 1.2 concerns the conditions for $k_n^{1/2} \log(n/k_n) \{ \mu_n^{(2)}(k_n) - \alpha \} \rightarrow b\alpha$ for some $b \in \mathbb{R}$.

Since $\mu_n^{(2)}(k_n) - \alpha = \{ - \int_0^{k_n/n} \log L(t) \log t \, dt \} / \{ \int_0^{k_n/n} \log^2 t \, dt \}$ by calculation from (1.2), we see by Lemma 5.1 below that $\mu_n^{(2)}(k_n) \rightarrow \alpha$ in Theorem 1.2, as claimed. Also, if $L(\cdot) \equiv 1$ in a right neighbourhood of 0, then $\mu_n^{(2)} = \alpha$ for all n sufficiently large. However, even if $L(\cdot) = e^{c_*}$ for a constant $c_* \neq 0$ in that neighbourhood, which certainly is a best possible distributional assumption for the problem at hand, we have $\mu_n^{(2)}(k_n) - \alpha \sim c_*/\log(n/k_n)$. Thus, in general, the asymptotic normality of $k_n^{1/2} \log(n/k_n) \{ \hat{\alpha}_n^{(2)}(k_n) - \alpha \}$ is *not* feasible. Since $\hat{\alpha}_n^{(2)}$ is derived exactly under the side condition that $L(\cdot) \equiv 1$, this is not surprising.

That $\mu_n^{(1)}(k_n) \rightarrow \alpha$ in Theorem 1.1 can be seen by putting $\rho(v) := - \int_0^v (1 + \log s) \, ds = -v \log v$ and noting that

$$\begin{aligned} \mu_n^{(1)}(k_n) &= \int_0^{k_n/n} \log Q(1 - t-) \, d\rho \left(\frac{nt}{k_n} \right) \\ &= - \int_0^{k_n/n} \rho \left(\frac{nt}{k_n} \right) \, d\{ \log Q(1 - t-) \} \\ &= - \int_0^1 \rho(v) \, df_{k_n/n}(v) \\ &\rightarrow - \int_0^1 \rho(v) \, d(\log v^{-\alpha}) \end{aligned}$$

$$= -\alpha \int_0^1 \log v \, dv = \alpha, \tag{3.3}$$

using the notation and the statement of Lemma 5.4 (iv) below. However, this approach does not provide a handle on the rate of convergence. Another approach rests on introducing the functionals $A_t(f) := -\int_0^1 (1 + \log s) \log f(st) \, ds$, $0 \leq t \leq 1$, on functions f . Then $\mu_n^{(1)}(k_n) - \alpha = -\alpha\{1 + A_{k_n/n}(t)\} + A_{k_n/n}(L)$ by (1.2), where $t(u) \equiv u$, and it is easy to see that $A_{k_n/n}(t) = -1$. Hence, in fact, $\mu_n^{(1)}(k_n) - \alpha = A_{k_n/n}(L)$. Thus, under $k_n/\log^4 n \rightarrow \infty$, (3.1) and (3.2) hold for $j = 1$ if and only if $k_n^{1/2} A_{k_n/n}(L) \rightarrow b\alpha$.

Sufficient conditions for the most useful case $b = 0$ may be obtained in terms of the functions a_L and b_L in the Karamata representation $L(s) = a_L(s) \exp\{\int_s^1 [b_L(u)/u] \, du\}$, $0 < s < 1$, where $\lim_{s \downarrow 0} a_L(s) = a_0$ for some $a_0 \in (0, \infty)$ and $\lim_{s \downarrow 0} b_L(s) = 0$. Then

$$A_{k_n/n}(L) = -\int_0^1 (1 + \log s) \log a_L\left(\frac{sk_n}{n}\right) \, ds - \int_0^1 (1 + \log s) \left(\int_{sk_n/n}^1 \frac{b_L(u)}{u} \, du \right) \, ds,$$

and using $\int_0^1 (1 + \log s) \, ds = 0$, one concludes that, if besides $k_n/\log^4 n \rightarrow \infty$ we have

$$k_n^{1/2} \sup_{0 < t \leq k_n/n} |\log a_L(t) - \log a_0| \rightarrow 0, \quad k_n^{1/2} \sup_{0 < t \leq k_n/n} |b_L(t)| \rightarrow 0,$$

then (3.1) and (3.2) hold for $j = 1$ and $b = 0$.

For example, if $L(s) = D_1(1 + D_2s^\beta)$, $0 < s \leq \delta$, for some $\delta \in (0, 1]$, as a special case of (4.1) below, then $a_L(s) = D_1$ and $b_L(s) = -\beta D_2 s^\beta / (1 + D_2 s^\beta)$, $0 < s \leq \delta$, so that $k_n^{1/2} \sup_{0 \leq u \leq k_n/n} |b_L(u)| \leq 2\beta |D_2| k_n^{(2\beta+1)/2} / n^\beta$ for all n large enough. Hence (3.1) and (3.2) hold for $j = 1$ and $b = 0$ whenever $k_n = o(n^{2\beta/(2\beta+1)})$. Also, it follows from the proof of Theorem 4.1 below that, if $k_n \sim \lambda n^{2\beta/(2\beta+1)}$, then (3.1) holds for a b such that $bD_2 < 0$, once the constant $\lambda > 0$ is given by $\lambda = \{-b(\beta + 1)^2 \alpha \beta^{-1} D_2^{-1}\}^{2/(2\beta+1)}$.

Finally, for $\mu_n = \mu_n(k_n) := \mu_{4,n}(k_n) - \mu_{5,n}^2(k_n)$ integrations by part and some algebra gives $\mu_n(k_n) = -2 \int_0^1 v f_{k_n/n}(v) \, df_{k_n/n}(v) - \left\{ \int_0^1 v \, df_{k_n/n}(v) \right\}^2$, so that $\mu_n(k_n) \rightarrow -2\alpha^2 \int_0^1 \log v \, dv - (-\alpha \int_0^1 dv)^2 = \alpha^2$ similarly as in (3.3). Hence $\mu_n^{(3)}(k_n) \rightarrow \alpha$ in Theorem 1.3. The problem of the rate in the latter convergence, and hence the centring problem in (3.1) for $j = 3$, is more complicated than for $j = 1$ and is not pursued here.

Monte Carlo experiments to assess actual coverage probabilities in (3.2) in some standard situations, such as the latest example above, would be useful. We refer to Novak and Utev (1990) for a nice mathematical treatment of a version of the problem in (3.1) for their estimator, which is based on the whole sample X_1, \dots, X_n .

4. The choice of k_n : asymptotic mean squared errors

The ‘ideal’ submodel of (1.2) is when $L(\cdot)$ is constant in $[0, \delta]$ for some $\delta \in (0, 1]$ (equivalently, $l(\cdot)$ in (1.1) is constant beyond some threshold), in which case $\mu_n^{(j)}(k_n) = \alpha$, $j = 1, 3$, by an extra calculation for $j = 3$, for all n for which $k_n/n \in [0, \delta]$ (and hence (3.1) holds with $b = 0$). In this ideal case, one would choose k_n quite large in practice to make the asymptotic variance $2\alpha^2/k_n$ of $\hat{\alpha}_n^{(1)}$ and $\hat{\alpha}_n^{(3)}$ small. In order to get a sense of the general

situation, where for $\hat{\alpha}_n^{(1)}$ the quantity $A_{k_n/n}^2(L)$ may be viewed as a measure of deviation of $L(\cdot)$ from a constant near 0, first we determine the theoretically optimal k_n which minimizes the asymptotic mean square errors of the estimators in another submodel of (1.2). The complexity of this submodel, considered by Hall (1982) in the same context, is the next step after that of the ‘ideal’ above; it is defined by assuming that

$$Q(1 - s-) = s^{-\alpha} D_1 [1 + D_2 s^\beta \{1 + o(1)\}] \quad \text{as } s \rightarrow 0, \tag{4.1}$$

where $D_1 > 0$, $D_2 \neq 0$ and $\beta > 0$ are constants. The theorems above suggest defining the asymptotic mean square error of $\hat{\alpha}_n^{(j)}$ as $M_n^{(j)} = M_n^{(j)}(k_n) := b_{j,n}^2(k_n) + 2\alpha^2 k_n^{-1}$ for $j = 1, 3$, and as $M_n^{(2)} = M_n^{(2)}(k_n) := b_{2,n}^2(k_n) + 2\alpha^2 k_n^{-1} \log^{-2}(n/k_n)$ for $j = 2$, for some $b_{j,n}^2(k_n) \sim \{\mu_n^{(j)}(k_n) - \alpha\}^2$, $j = 1, 2, 3$; the sum of the asymptotic squared bias and variance (cf. Theorem 5 of Csörgő *et al.* (1985), with several examples, and Theorem 2.2 of Viharos (1995); they do not formally include the factor α^2 in their variance terms).

Theorem 4.1. *If (4.1) holds with some constants $D_1 > 0$, $D_2 \neq 0$ and $\beta > 0$, then for both $\hat{\alpha}_n^{(1)}(k_n)$ and $\hat{\alpha}_n^{(3)}(k_n)$ the optimal choice is*

$$k_n = k_n^* := \left\lfloor \left(\frac{\alpha^2(\beta + 1)^4}{\beta^3 D_2^2} \right)^{1/(2\beta+1)} n^{2\beta/(2\beta+1)} \right\rfloor \tag{4.2}$$

in which case, for both $j = 1$ and $j = 3$,

$$M_n^{(j)}(k_n^*) \sim \alpha^{4\beta/(2\beta+1)} D_2^{2/(2\beta+1)} 2 \left(1 + \frac{1}{2\beta} \right) \left(\frac{\beta^3}{(\beta + 1)^4} \right)^{1/(2\beta+1)} n^{-2\beta/(2\beta+1)}. \tag{4.3}$$

Furthermore, if $D_1 \neq 1$, then $M_n^{(2)}(k_n) \sim b_{2,n}^2(k_n) = (\log D_1)^2 / \{\log(n/k_n)\}^2$.

The behaviour of $\hat{\alpha}_n^{(2)}$ is again completely analogous to those of Viharos’s (1996) estimates. Its large bias makes $\hat{\alpha}_n^{(2)}$ worse than the other two estimators. (Of course, $M_n^{(2)}(k_n)$ is small if D_1 is close to 1.) Under (4.1), there is no difference between the ideal asymptotic performances of $\hat{\alpha}_n^{(1)}(k_n^*)$ and $\hat{\alpha}_n^{(3)}(k_n^*)$ with mean square error as criterion.

Using Theorem 4.1, it is possible to compare $\hat{\alpha}_n^{(1)}(k_n^*)$ with the Hill (1975) estimator, the optimal kernel estimator of Csörgő *et al.* (1985), and the optimal linear combination of Viharos (1995). For each of these, the corresponding smallest possible asymptotic mean square error M_n under (4.1) is of the same order with the same two leading constants as in (4.3); so the comparison can be made by means of the corresponding four functions

$$m(\beta) := \lim_{n \rightarrow \infty} \frac{M_n n^{2\beta/(2\beta+1)}}{\alpha^{4\beta/(2\beta+1)} D_2^{2/(2\beta+1)}}, \quad \beta > 0.$$

Figure 1, drawn by Maple V, depicts the $m(\beta)$ of the optimal Hill estimator $\hat{\alpha}_{H,n} = \hat{\alpha}_n^{(0)}$, of Viharos’s (1995) optimal estimator $\hat{\alpha}_{V,n}$, of the optimal kernel estimator $\hat{\alpha}_{K,n}$ and of $\hat{\alpha}_{*,n} = \hat{\alpha}_n^{(j)}(k_n^*)$, $j = 1, 3$. We have $\lim_{\beta \rightarrow \infty} m(\beta) = 2$ for $\hat{\alpha}_{*,n}$, and Viharos (1995) has shown that $\lim_{\beta \rightarrow \infty} m(\beta) = 1$ for the other three estimators.

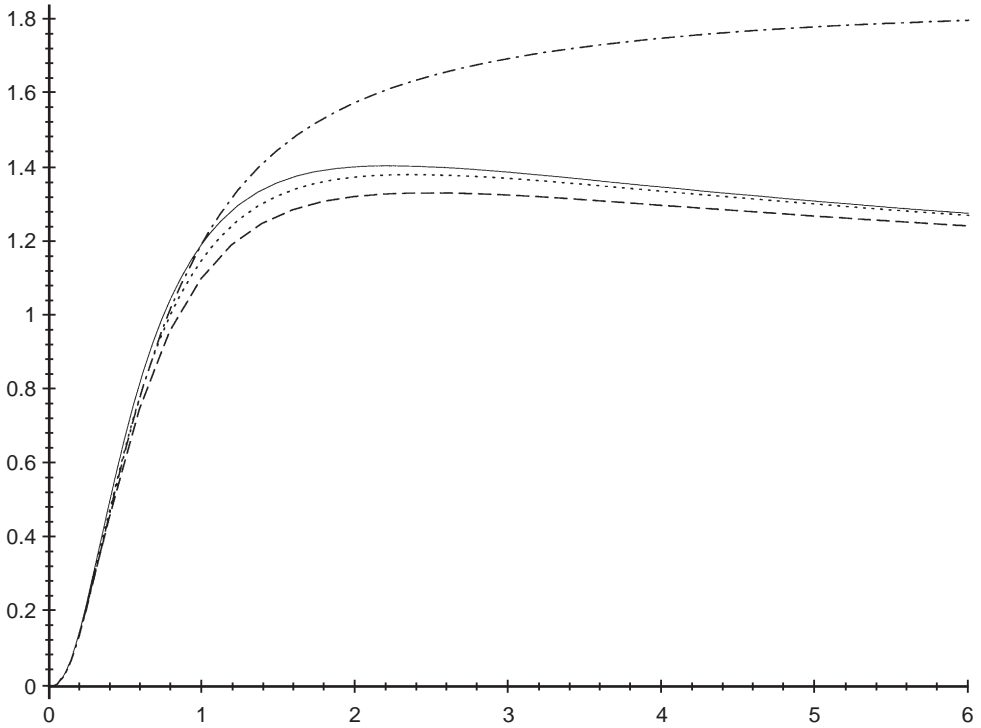


Figure 1. The curves of $m(\beta)$ for $\hat{\alpha}_{H,n}$ (—), $\hat{\alpha}_{V,n}$ (⋯), $\hat{\alpha}_{K,n}$ (---) and $\hat{\alpha}_{*,n}$ (-·-).

The estimators $\hat{\alpha}_{K,n}$ and $\hat{\alpha}_{V,n}$ are better uniformly than $\hat{\alpha}_{H,n}$ because optimal choices of a kernel and an extra tuning parameter are made besides that of k_n and both of the original classes contain the Hill estimator for a particular choice of the kernel and the tuning parameter, respectively. They are also uniformly better than $\hat{\alpha}_{*,n}$. While $\hat{\alpha}_n^{(j)}(k_n^*)$, $j = 1, 3$, are a little better than the Hill estimator for $\beta < 1$, they are relatively poor for large β ; here $\beta \rightarrow \infty$ means approaching the ‘ideal’ submodel, where L is constant near zero. Even though the relative gains quickly become negligible as n increases, the results suggest using the other three estimators with a somewhat large k_n , rather than $\hat{\alpha}_n^{(j)}(k_n)$, $j = 1, 3$, if there is sufficient reason to believe that $l(x)$ in (1.1) is nearly constant for large x . Plots of the graph of $x^{1/\hat{\alpha}_n}\{1 - \hat{F}_n(x)\}$ for large x , with preliminary estimates $\hat{\alpha}_n$ may be helpful in exploring this. What should be done if a substantial deviation of $l(\cdot)$ from a constant is suspected far out so that the validity of Hall’s very restrictive submodel becomes questionable?

From now on we concentrate primarily on $\hat{\alpha}_n^{(1)}$. In a general situation, its asymptotic mean square error is $M_n^{(1)}(k_n; L, \alpha) = A_{k_n/n}^2(L) + 2\alpha^2 k_n^{-1}$ for any L in (1.2), where the first term, reflecting the asymptotic squared bias, is a nonlinear measure of deviation of L from a constant near zero. The universal asymptotic normality of $\hat{\alpha}_n^{(j)}(k_n)$, $j = 1, 3$, implies that

their asymptotic variance is $2\alpha^2/k_n$ for any L , which may *not* be true for the other estimators. Thus, while the other estimators may be slightly better for a near-constant L , as discussed for the Hall model above, the least-squares estimators $\hat{\alpha}_n^{(j)}$, $j = 1, 3$, are more ‘robust’ against deviations of L from a constant. To investigate the mean square error, introduce now the function $M_n^*(t; L, \alpha) := A_t^2(L) + 2\alpha^2 t^{-1} n^{-1}$, $0 < t < 1$, and note that $M_n^{(1)}(k_n; L, \alpha) = M_n^*(k_n/n; L, \alpha)$ for any L in (1.2) and any k_n as in (1.3). We considered a variety of slowly varying functions L , with particular interest in unbounded functions. Those that we eventually kept for illustration are defined by

$$\begin{aligned}
 L_{10}(s) &:= \log \log \left(\frac{1}{s} \right), & L_9(s) &:= \exp \left\{ \log^{3/4} \left(\frac{1}{s} \right) \right\}, \\
 L_8(s) &:= \exp \left\{ \log^{2/3} \left(\frac{1}{s} \right) \right\}, & L_7(s) &:= \log \left(\frac{1}{s} \right), \\
 L_6(s) &:= \exp \left\{ \log^{1/2} \left(\frac{1}{s} \right) \right\}, & L_5(s) &:= \exp \left[\log^{1/3} \left(\frac{1}{s} \right) \cos \left\{ \log^{1/3} \left(\frac{1}{s} \right) \right\} \right], \\
 L_4(s) &:= \exp \left\{ \log^{1/4} \left(\frac{1}{s} \right) \right\}, & L_3(s) &:= 1 + s, \\
 L_2(s) &:= 1 + s \sin \left(\frac{1}{s} \right), & L_1(s) &:= \exp \left(\frac{\log(1/s)}{\log \log(1/s)} \right),
 \end{aligned}$$

with L_1, L_4, \dots, L_{10} taken from the list of Bingham *et al.* (1987, p. 16). Here $L_k(s)$ is as above for $s \in (0, s_k(\alpha))$, where $s_k(\alpha) \in (0, 1]$ is such that $Q_k(1 - s) = L_k(s)/s^\alpha$ in (1.2) is decreasing on $(0, s_k(\alpha))$; if $s_k(\alpha) < 1$, we set $L_k(s) := L_k(s_k(\alpha))$ for all $s \in [s_k(\alpha), 1)$, $k = 1, \dots, 10$, to preserve this property on the whole $(0, 1)$. We have $s_1(\alpha) = \exp(-\exp[\{(1 + 4\alpha)^{1/2} - 1\}/2\alpha])$ and $s_k(\alpha) = 1$, $k = 4, 6, 7, 8, 9$, for all $\alpha > 0$, $s_{10}(\alpha) = 1$ for $\alpha \leq e$ and $s_5(\alpha) = 1$ for $\alpha \geq 2/3\pi \approx 0.2122$. Also, $s_k(\alpha) = 1$, $k = 2, 3$, if $\alpha \geq 1$. However, if $\alpha < 1$, then $s_3(\alpha) = \alpha/(1 - \alpha)$ and L_2 must be deleted from the list. Also, L_5 will be dropped in Figure 4 below since $s_5(1/5) \approx 0.00177 < 0.01$.

Figure 2 fully contains the graphs of $M_{500}^*(t; L_k, 1)$, $t \in (0.01, 0.36]$, for $k = 1, \dots, 10$ as solid curves, labelled by the corresponding indices of the ten slowly varying functions considered, as they leave the Figure. It also fully contains the five graphs of $M_{1000}^*(t; L_k, 1)$, $0.01 < t \leq 0.36$, for $k = 2, 3, 8, 9, 10$ as dotted curves, pulled down by the doubled sample size in a clear fashion, and also those parts of the dotted graphs of $M_{1000}^*(\cdot; L_k, 1)$ for $k = 1, 4, 5, 6, 7$ which do not block the view of the interesting crossing pattern of the solid curves. (These are for t values before they reach a solid line and for $t \in [0.269, 0.36]$; the missing parts are easily visualized by the overall shape and the pertaining solid curves.)

On the basis of Figure 2 the present discussion refers to the estimation of $\alpha = 1$, but we shall see below that it is of much wider applicability.

Although the oscillating $M_n^*(\cdot; L_2, 1)$ becomes worse beyond e^{-1} , particularly for $n \ll 500$, the good behaviour of $\hat{\alpha}_n^{(1)}$ on L_2 , deviating from L_3 of (4.1), is surprising in view

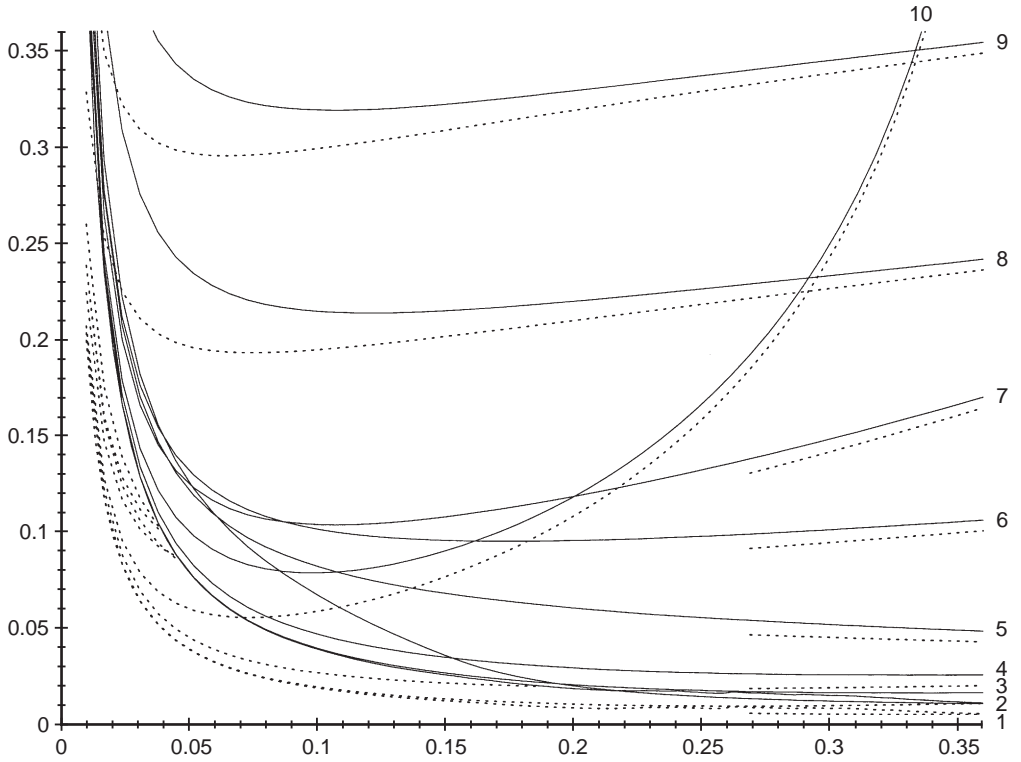


Figure 2. The curves of $M_{500}^*(t; L_k, 1)$ (—) and $M_{1000}^*(t; L_k, 1)$ (⋯), $k = 1, \dots, 10$.

of how poor L_2 is for the Hill estimator (Csörgő and Viharos 1995). Similarly, the ‘infinite oscillation’ of L_5 is harmless for $\hat{\alpha}_n^{(1)}$. The slopes of $M_n^*(\cdot; L_k, 1)$, $k = 4, 6, 8, 9$, change beyond e^{-1} at the rate seen near e^{-1} , while $M_n^*(\cdot; L_7, 1)$ becomes as bad near $t = 1$ as $M_n^*(\cdot; L_{10}, 1)$ near $t = e^{-1}$, $n = 500, 1000$. Writing $L \prec L^*$ if $\lim_{s \downarrow 0} |L^*(s)|/|L(s)| = \infty$, the ordering $L_2 \sim L_3 \prec L_{10} \prec L_7 \prec L_4 \prec L_6 \prec L_8 \prec L_9 \prec L_1$ is not preserved globally for $M_n^*(\cdot; L, 1)$, except for the subseries $L_3 \prec L_4 \prec L_6 \prec L_8 \prec L_9$. Estimation will be poor with L_9 even for very large n . However, it may be surprisingly good with the ‘largest’ function L_1 , for the lucky choice of $k_n \approx n/4$, $n \in [100, 3000]$.

For sample sizes $n = 400, 300, 200, 100$ the curves shift more and more upwards and become ‘less and less convex’, but the whole global picture remains the same; for the fully decreasing curves the gain beyond $t \approx 0.27$, say, is negligible, but the harm of using a large $t = k_n/n$ for all the others may be great. The harm of using a very small $t = k_n/n$ is clear for all the curves; the price for a small bias is a huge variance. Although the typical ten curves here, and the others that we looked at, do not of course ‘grade’ the whole ocean of L functions substantially different from a constant near zero, the following pragmatic rule of thumb appears reasonable for estimating α near 1: For $100 \leq n < 300$ use $\hat{\alpha}_n^{(1)}(k_n)$ with

$9n/100 \leq k_n \leq 28n/100$, and for $300 \leq n \leq 3000$ use $\hat{\alpha}_n^{(1)}(k_n)$ with $8n/100 \leq k_n \leq 27n/100$; the factors of n decrease very slowly as n increases. Since the asymptotic behaviour of $\hat{\alpha}_n^{(3)}$ is analogous, within these intervals choose (adaptively) that k_n for which $\hat{\alpha}_n^{(1)}(k_n)$ and $\hat{\alpha}_n^{(3)}(k_n)$ are the closest to each other; this is part of the suggested rule and it uniquely determines the choice of a data-driven k_n for $n \geq 100$.

The question now is whether the fact that α is near 1 is unimportant in these considerations. Do these findings remain applicable for a whole range of the estimated parameter?

With a decreased unit on the vertical axis, Figure 3 is the version of Figure 2 for $\alpha = 2$. While the bias term remains the same for a given L , the greater α (or the smaller the exponent $1/\alpha$ in (1.1)), the greater is the asymptotic variance of course. However, not only does the picture remain qualitatively the same, but also we do not hesitate to claim that our rule of thumb extends to the range $1 \leq \alpha \leq 2$. One may argue that the case $\alpha > 2$, when $E(X^{1/2}) = \infty$, is not practical any longer. (For example, there is no instance in the large literature when typical estimates would suggest $\alpha > 2$ if the controversial Mandelbrot hypothesis for stock price changes and related quantities is valid in a certain situation, and

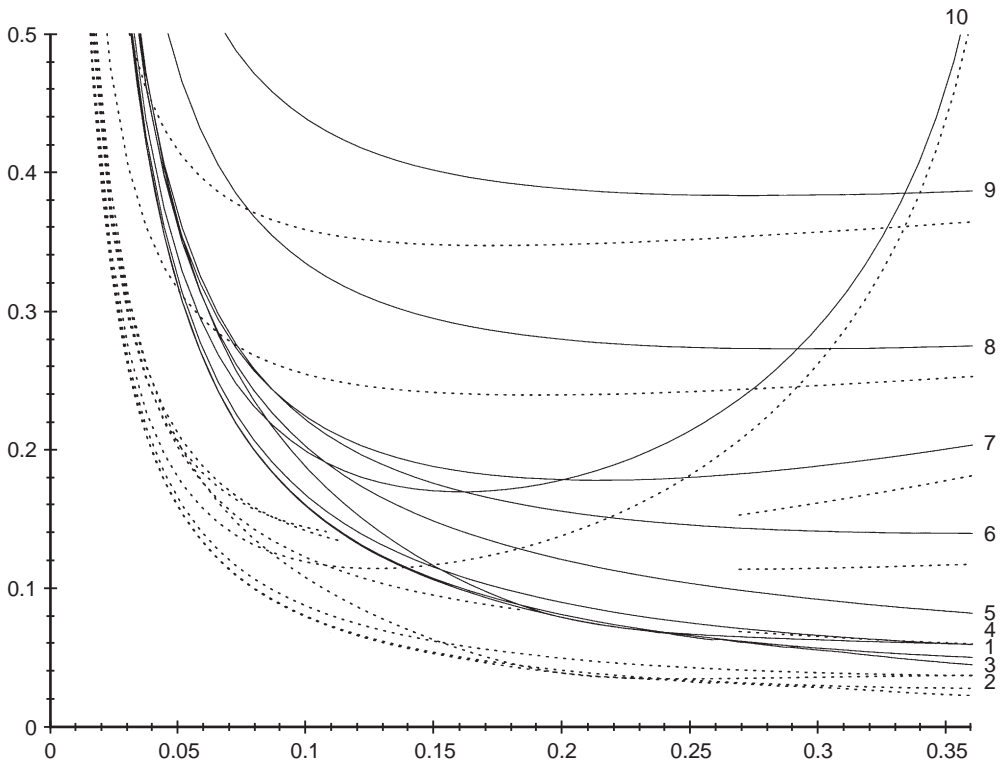


Figure 3. The curves of $M_{500}^*(t; L_k, 2)$ (—) and $M_{1000}^*(t; L_k, 2)$ (⋯), $k = 1, \dots, 10$.

hence the underlying distribution is approximately stable with exponent $1/\alpha \in (0, 2)$; see Csörgő (1984) for many of the early references.) Clearly, the case $\alpha < 1$ is more important for possible applications. We have plotted the cases $\alpha = 1/m$, $m = 2, \dots, 10$, for which Figure 4 gives the case $\alpha = 1/5$ with increased units on the vertical axis and no need to delete portions of the dotted curves. The transition through the values $m = 2, 3, 4$ from Figure 2 to Figure 4 is ‘smooth’ and continues to be so for $m = 6, 7, \dots$. The smaller α , the smaller is the mean square error for each L . While the relative error for estimating a small α may become huge for a large L , our rule of thumb for the choice of k_n remains pragmatic; it guards against worst cases when $\alpha \in [1/2, 2]$, while for $\alpha < 1/2$ the loss from the exclusion of a smaller k_n for some L , such as L_{10}, L_9, L_8 and L_7 , is negligible. So, we recommend the rule for all $\alpha \leq 2$.

We do not believe that asymptotic considerations are very helpful for this estimation problem when $n \ll 100$. We refer to S&S for simulation results with $n = 50$. The results raise the question of *weighted* least-squares versions.

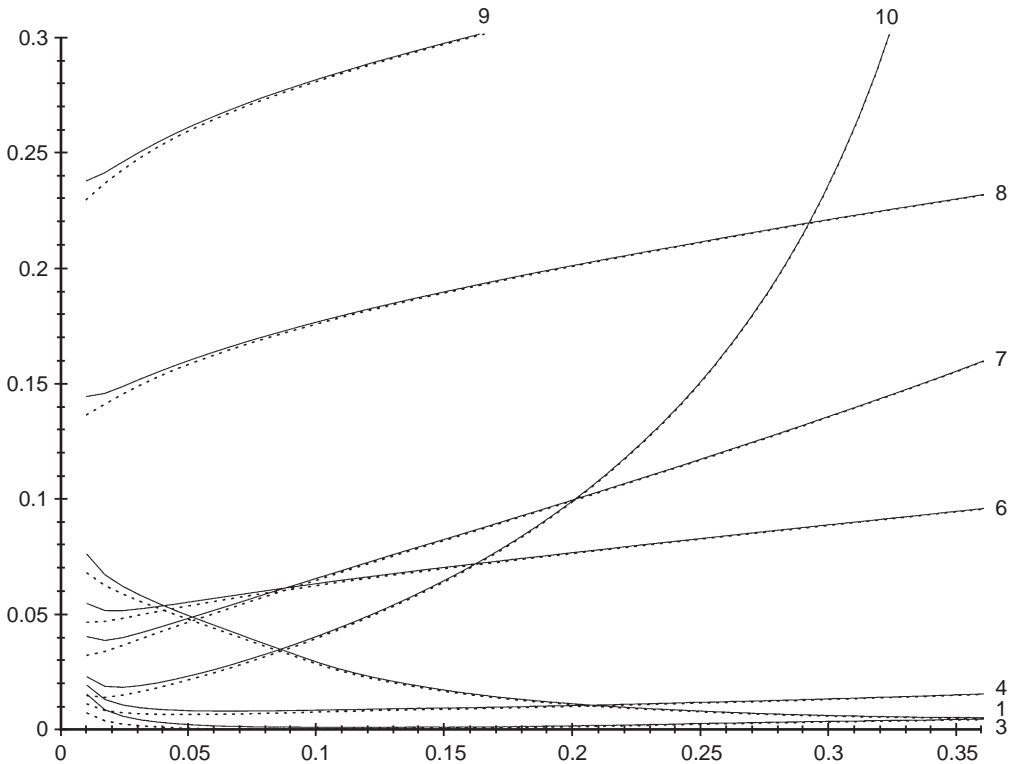


Figure 4. The curves of $M^*_{500}(t; L_k, 1/5)$ (—) and $M^*_{1000}(t; L_k, 1/5)$ (⋯), $k = 1, 3, 4, 6, 7, 8, 9, 10$.

5. Proofs

Throughout, (1.3) is assumed to hold, $\stackrel{\mathcal{D}}{=}$ denotes distributional equality, $u \wedge v = \min(u, v)$ and $u \vee v = \max(u, v)$, $u, v \in \mathbb{R}$. The claims that $\mu_n^{(j)} \rightarrow \alpha$, $j = 1, 2, 3$, in Theorems 1.1–1.3 have been established above. The first lemma is for general use while the next two aim at Theorem 1.2.

Lemma 5.1 (de Haan 1970, Corollary 1.2.1.1; Bingham *et al.* 1987, Proposition 1.3.6). *If L is a function slowly varying at zero, then $\lim_{s \downarrow 0} \log L(s)/\log s = 0$.*

Lemma 5.2. *Put $d_{n+1-i,n} := -n \int_{(i-1)/n}^{i/n} \log t \, dt$. If $\{k_n \log^2(n/k_n)\}/\log^2 n \rightarrow \infty$, then*

$$\frac{1}{2^{1/2} \alpha k_n^{1/2} \log(n/k_n)} \left(\sum_{i=1}^{k_n} d_{n+1-i,n} \log X_{n+1-i,n} - n\mu_{2,n} \right) \stackrel{\mathcal{D}}{\rightarrow} N(0, 1). \tag{5.1}$$

Proof. A straightforward extension of Corollary 1.1 of Viharos (1995) gives (5.1) with $\mu'_n = \mu'_n(k_n) := -n \int_{1/n}^{k_n/n} \log Q(1-t) \log t \, dt + d_{n,n} \log Q(1-n^{-1})$ replacing $n\mu_{2,n}$, where now $d_{n,n} = 1 + \log n$. Integrating by parts,

$$\begin{aligned} \mu'_n &= n\mu_{2,n} + n \int_0^{1/n} \log Q(1-t) \log t \, dt + (1 + \log n) \log Q(1-n^{-1}) \\ &= n\mu_{2,n} - n \int_0^{1/n} t(\log t - 1) \, d\{\log Q(1-t)\}. \end{aligned}$$

Furthermore, as $x \downarrow 0$,

$$\begin{aligned} \int_0^x t(\log t - 1) \, d\{\log Q(1-t)\} &= - \int_{Q(1-x)}^\infty \{1 - F(v)\} [\log \{1 - F(v)\} - 1] \frac{1}{v} \, dv \\ &\sim -\alpha [1 - F\{Q(1-x)\}] (\log [1 - F\{Q(1-x)\}] - 1) \\ &\sim -\alpha x \log x, \end{aligned}$$

so that $\mu'_n = n\mu_{2,n} + O(\log n)$, implying the lemma by the condition on $\{k_n\}$. □

Lemma 5.3. *Put $c_{n+1-i,n} := d_{n+1-i,n} - \log(n/i) = 1 - (i-1) \log \{i/(i-1)\}$, where $c_{n,n}$ is understood as 1. Then $\sum_{i=1}^{k_n} c_{n+1-i,n} \log X_{n+1-i,n} = O_P(\log^2 n)$.*

Proof. The function $q(x) := 1 - x \log \{(x+1)/x\}$, $x > 0$, is strictly decreasing, with $q(0+) = 1$ and $q(\infty) = 0$, and $q(x) \leq C \log \{(x+1)/x\}$, $x > 0$, for some constant $C > 0$. This implies that $0 < c_{i,n} \leq 1$ and $c_{n+1-i,n} \leq C e_{n+1-i,n}/n$, $i = 2, 3, \dots, n$, where $e_{n+1-i,n} := n \int_{(i-1)/n}^{i/n} t^{-1} \, dt = n \log \{i/(i-1)\}$. Noting that $\log X_{n,n} = O_P(\log n)$, it suffices to show that $S_n := \sum_{i=2}^{k_n} e_{n+1-i,n} \log X_{n+1-i,n} = O_P(n \log^2 n)$.

Recalling (1.2) and using Lemma 5.1 for the L there, we fix a $\delta \in (0, 1)$ such that $\log Q(1-s)/\log s^{-\alpha} \leq 2$ for all $s \in (0, \delta)$. Let $U_{1,n} \leq \dots \leq U_{n,n}$ denote the order

statistics based on a sample of size n from the uniform(0,1) distribution, so that $S_n \stackrel{\mathcal{D}}{=} \sum_{i=2}^{k_n} e_{n+1-i,n} \log Q(1 - U_{i,n}) =: \Sigma_n$. Then, on the event $E_n := \{U_{k_n,n} < \delta\}$, we have $\Sigma_n \leq 2 \sum_{i=2}^{k_n} e_{n+1-i,n} \log U_{i,n}^{-\alpha} =: 2\Sigma_n^*$. Hence, with $I\{\cdot\}$ denoting the indicator function, $\Sigma_n \leq 2I\{E_n\}\Sigma_n^* + O_P(n \log^2 n)$, since $P\{E_n\} \rightarrow 1$. Therefore, it now suffices to show that $\Sigma_n^* \stackrel{\mathcal{D}}{=} \tilde{\Sigma}_n := \sum_{i=2}^{k_n} e_{n+1-i,n} \log \tilde{Q}(U_{n+1-i,n}) = O_P(n \log^2 n)$, where $\tilde{Q}(1-s) := s^{-\alpha}$, $s \in (0, 1)$. Introduce $\tilde{\mu}_n := n \int_{1/n}^{k_n/n} t^{-1} \log \tilde{Q}(1-t) dt + e_{n-1,n} \log \tilde{Q}(1-n^{-1})$ and $a_n^2 := \int_{1/n}^{k_n/n} \int_{1/n}^{k_n/n} (u \wedge v - uv) dK(u) dK(v)$, where $K(t) := -\int_{1/2}^t u^{-1} d\{\log \tilde{Q}(1-u)\}$. Then by Theorem 1.2 of Viharos (1995), $\{\tilde{\Sigma}_n - \tilde{\mu}_n\}/n^{1/2} a_n = O_P(1)$, and easy calculation shows that $\tilde{\mu}_n = \mathcal{O}(n \log^2 n)$ and $a_n^2 \sim 2\alpha^2 n$. Hence we finally obtain $\tilde{\Sigma}_n = n^{1/2} a_n \{(n^{1/2} a_n)^{-1}(\tilde{\Sigma}_n - \tilde{\mu}_n)\} + \tilde{\mu}_n = O_P(n) + O(n \log^2 n) = O_P(n \log^2 n)$. \square

Proof of Theorem 1.2. For the sequence $r_n := \sum_{i=1}^{k_n} \log^2(n/i)$, figuring in the denominator of $\hat{\alpha}_n^{(2)}$, we have $n \int_{1/n}^{(k_n+1)/n} \log^2 t dt \leq r_n \leq n \int_0^{k_n/n} \log^2 t dt = nj_n$, where $nj_n \sim k_n \log^2(n/k_n)$, so that $r_n/nj_n = 1 + O[\log^2 n / \{k_n \log^2(n/k_n)\}]$. These relations and the growth condition on $\{k_n\}$ imply the theorem through Lemmas 5.2 and 5.3. \square

Consider next $\hat{\alpha}_n^{(1)}$. Set $i_n = i_n(k_n) := k_n^{-1} \sum_{i=1}^{k_n} \log^2(n/i) - \{k_n^{-1} \sum_{i=1}^{k_n} \log(n/i)\}^2$ for its denominator. Since $i_n \rightarrow 1$ if $k_n/\log^2 n \rightarrow \infty$ (S&S, Lemma 2.1), first we examine $i_n(k_n)\hat{\alpha}_n^{(1)}(k_n) \stackrel{\mathcal{D}}{=} \tilde{\alpha}_n(k_n) = \bar{\alpha}_n(k_n) + R_{1,n}(k_n) + R_{2,n}(k_n)$, where $\tilde{\alpha}_n = \tilde{\alpha}_n(k_n)$, $\bar{\alpha}_n = \bar{\alpha}_n(k_n)$, $R_{1,n} = R_{1,n}(k_n)$ and $R_{2,n} = R_{2,n}(k_n)$ are given by

$$\tilde{\alpha}_n := \frac{1}{k_n} \sum_{i=1}^{k_n} \log\left(\frac{n}{i}\right) \log Q(1 - U_{i,n}) - \frac{1}{k_n^2} \left(\sum_{i=1}^{k_n} \log Q(1 - U_{i,n}) \right) \left\{ \sum_{i=1}^{k_n} \log\left(\frac{n}{i}\right) \right\}, \tag{5.2}$$

$$\bar{\alpha}_n := \frac{1}{k_n} \sum_{i=1}^{k_n} g_{n+1-i,n} \log Q(1 - U_{i,n}) \text{ with } g_{n+1-i,n} := -k_n \int_{(i-1)/k_n}^{i/k_n} (1 + \log x) dx,$$

$$R_{1,n} := -\frac{1}{k_n} \sum_{i=1}^{k_n} c_{n+1-i,n} \log Q(1 - U_{i,n}) \text{ with } c_{n+1-i,n} = 1 - (i-1) \log\left(\frac{i}{i-1}\right), c_{n,n} = 1,$$

and

$$R_{2,n} := -\frac{1}{k_n} \left\{ \frac{1}{k_n} \sum_{i=1}^{k_n} \log\left(\frac{n}{i}\right) - \log\left(\frac{n}{k_n}\right) - 1 \right\} \sum_{i=1}^{k_n} \log Q(1 - U_{i,n}).$$

Assuming henceforth that $k_n/\log^4 n \rightarrow \infty$, the next lemma says that $\bar{\alpha}_n$ dominates in $\tilde{\alpha}_n$.

Lemma 5.4. *The following four relations hold: (i) $R_{1,n}(k_n) = O_P\{(\log^2 n)/k_n\}$, (ii) $R_{2,n}(k_n) = O_P[(\log n)\{\log(n/k_n)\}/k_n]$, (iii) $i_n(k_n) = 1 + O\{(\log^2 n)/k_n\}$ and (iv) $f_s(v) := \log Q(1 - sv) - \log Q(1 - s) \rightarrow \log v^{-\alpha}$ as $s \downarrow 0$ for every $v > 0$.*

Proof. Here (i) follows from Lemma 5.3. Further, S&S showed while proving their Lemmas 2.5 and 2.3 that $|k_n^{-1} \sum_{i=1}^{k_n} \log(n/i) - \log(n/k_n) - 1| \leq (\log n + 1)/k_n$ and

$\{k_n \log(n/k_n)\}^{-1} \sum_{i=1}^{k_n} \log Q(1 - U_{i,n}) \rightarrow a$ in probability. These imply (ii), (iii) follows by the technique in the proof of Theorem 1.2, and (iv) is evident from (1.2). \square

To prove Theorems 1.1 and 1.3, we use the probability space constructed by Mason and van Zwet (1987). It carries a sequence $\{U_n\}_{n=1}^\infty$ of independent random variables uniformly distributed on $(0, 1)$, with $U_{1,n} \leq \dots \leq U_{n,n}$ as the order statistics of U_1, \dots, U_n again, and a sequence $\{B_n(t) : 0 \leq t \leq 1\}_{n=1}^\infty$ of Brownian bridges such that, letting $G_n(s) := n^{-1} \sum_{k=1}^n I\{U_k \leq s\}$, $0 < s < 1$, and $U_n(s) := U_{k,n}$, $(k-1)/n < s \leq k/n$, $k = 1, \dots, n$, and putting $\beta_n(s) := n^{1/2}\{G_n(s) - s\}$ and $\gamma_n(s) := n^{1/2}\{s - U_n(s)\}$ for the respective uniform empirical and quantile processes, we have

$$\sup_{\lambda/n \leq s \leq 1 - (\lambda/n)} \frac{|\beta_n(s) - B_n(s)|}{\{s(1-s)\}^{(1/2) - \nu}} = O_P\left(\frac{1}{n^\nu}\right), \quad \sup_{\lambda/n \leq s \leq 1 - (\lambda/n)} \frac{|\gamma_n(s) - B_n(s)|}{\{s(1-s)\}^{(1/2) - \kappa}} = O_P\left(\frac{1}{n^\kappa}\right) \tag{5.3}$$

for any fixed $\nu \in [0, \frac{1}{2})$, $\kappa \in [0, \frac{1}{4})$ and $\lambda > 0$, and

$$\sup_{0 < s < 1} |\beta_n(s) - B_n(s)| = O\left(\frac{\log n}{n^{1/2}}\right) \text{ almost surely.} \tag{5.4}$$

(This probability space carries the construction ‘dual’ to that in Csörgő *et al.* (1986). In fact, through Viharos (1995), the proof of Theorem 1.2 was also on one of the two spaces.)

Proof of Theorem 1.1. The general outline of the proof of Theorem 1 in Viharos (1993) is followed; so we set $J(t) := -1 - \log t$ and $g(t) := \log Q(1 - t)$ to accommodate the notation there. Introducing

$$N_n = N_n(l_n, k_n) := -\left(\frac{n}{k_n}\right)^{1/2} \int_{l_n/n}^{k_n/n} B_n(t) J\left(\frac{n}{k_n} t\right) dg(t),$$

note first that by Lemma 5.4 (i)–(iii) it is enough to prove that

$$k_n^{1/2}\{\bar{\alpha}_n(k_n) - \mu_n^{(1)}(k_n)\} = N_n(l_n, k_n) + o_P(1), \quad N_n(l_n, k_n) \xrightarrow{\mathcal{D}} N(0, 2\alpha^2) \tag{5.5}$$

for some sequence l_n such that $l_n \rightarrow \infty$ and $k_n/l_n \rightarrow \infty$.

We need $\bar{\mu}_n = \bar{\mu}_n(k_n) := nk_n^{-1} \int_{1/n}^{k_n/n} g(t) J(nt/k_n) dt + g_{n,n} g(1/n) k_n^{-1}$ as a modified centring sequence. Also, let $G_n^*(s) := \{G_n(s) \wedge (1 - n^{-1})\} \vee n^{-1}$ and, for $1 \leq m \leq r \leq n$,

$$\Theta_n(m, r) := -\int_{m/n}^{r/n} \int_{nt/k_n}^{nG_n^*(t)} J(s) ds dg(t), \quad V_n(m) := \int_{U_{m,n}}^{m/n} \int_{m/k_n}^{nG_n^*(t)/k_n} J(s) ds dg(t).$$

Fix any integers m and l such that $1 \leq m \leq l \leq k_n$. Then we obtain

$$\begin{aligned} \bar{\alpha}_n - \bar{\mu}_n &= \Theta_n(1, m) - V_n(1) + R_n + \Theta_n(m, l) + V_n(k_n) + \Theta_n(l, k_n) \\ &=: \Delta_n(m, l) + V_n(k_n) + \Theta_n(l, k_n) \end{aligned} \tag{5.6}$$

as an analogue of (2.3) in Viharos (1993), where $R_n := g_{n,n}\{g(U_{1,n}) - g(1/n)\}/k_n = O_P\{(\log k_n)/k_n\}$ by an application of Lemma 5.4 (iv), because $g_{n,n} = \log k_n$.

Consider first $V_n(k_n) = -n \int_{U_{k_n,n}}^{k_n/n} G_n^*(t) \log\{nG_n^*(t)/k_n\} dg(t)/k_n$. On the event $E_{1,n} :=$

$\{U_{1,n} < k_n/n < U_{n,n}\}$ we have $G_n^*(t) = G_n(t)$ in the integral, so that $V_n(k_n) = I\{E_{1,n}\}V_n(k_n) + o_P(k_n^{-1/2})$ since $P\{E_{1,n}\} \rightarrow 1$. Hence $|V_n(k_n)| \leq I\{E_{1,n}\}\Gamma_{1,n}\Gamma_{2,n} + o_P(k_n^{-1/2})$, where $\Gamma_{1,n} := n \sup_t |G_n(t)| \log \{nG_n(t)/k_n\}/k_n$, with the supremum taken for all t in between k_n/n and $U_{k_n,n}$, and $\Gamma_{2,n} := |g(k_n/n) - g(U_{k_n,n})| = o_P(1)$ by Lemma 5.4 (iv). The analysis of the function $\rho^*(s) := s|\log s|$ and the fact that $\rho^*(nG_n(U_{k_n,n})/k_n) = 0$ shows that on the event $E_{2,n} := \{nG_n(k_n/n)/k_n > 1/e\}$ we have $\Gamma_{1,n} = nG_n(k_n/n) \log \{nG_n(k_n/n)/k_n\}/k_n =: \Lambda_n$. Since $P(E_{2,n}) \rightarrow 1$, we get $|V_n(k_n)| \leq I\{E_{2,n}\}\Lambda_n o_P(1) + o_P(k_n^{-1/2})$. An application of (5.3) gives $nG_n(k_n/n)/k_n = 1 + O_P(k_n^{-1/2})$, and hence $\Lambda_n = O_P(k_n^{-1/2})$. Therefore, $V_n(k_n) = o_P(k_n^{-1/2})$.

Next, on the event $E_{3,n} := \{nG_n^*(l/n)/k_n < 1/e\} \cap \{l/n < U_{n,n}\}$, for some $\xi_n(v)$ in between $nG_n^*(v/n)/k_n$ and v/k_n , for which $J(\xi_n(v)) > 0$, we have

$$\begin{aligned} |\Theta_n(m, l)| &= \left| \int_m^l \int_{v/k_n}^{nG_n^*(v/n)/k_n} J(s) ds dg\left(\frac{v}{n}\right) \right| \\ &\leq - \int_m^l \left| \frac{n}{k_n} G_n^*\left(\frac{v}{n}\right) - \frac{v}{k_n} \right| J(\xi_n(v)) dg\left(\frac{v}{n}\right) \\ &\leq -J\left(\frac{1}{k_n}\right) \left\{ \frac{n^{1/2}}{k_n} \int_m^l \left| \beta_n\left(\frac{v}{n}\right) \right| df_{1/n}(v) + \frac{1}{k_n} \int_m^l df_{1/n}(v) \right\} \\ &= -J\left(\frac{1}{k_n}\right) \left\{ \frac{n^{1/2}}{k_n} \int_m^l \left| B_n\left(\frac{v}{n}\right) \right| df_{1/n}(v) + \frac{n^{1/2}}{k_n} O\left(\frac{\log n}{n^{1/2}}\right) + O\left(\frac{1}{k_n}\right) \right\} \\ &= (-1 + \log k_n) \left\{ O_P\left(\frac{1}{k_n}\right) + O\left(\frac{\log n}{k_n}\right) + O\left(\frac{1}{k_n}\right) \right\} \\ &= O_P\left(\frac{(\log k_n)(\log n)}{k_n}\right), \end{aligned}$$

using (5.4), a simple argument for the integral involving the Brownian bridge B_n and Lemma 5.4 (iv). Since $P\{E_{3,n}\} \rightarrow 1$, this implies that $k_n^{1/2}\Theta_n(m, l) = o_P\{(\log k_n)/k_n^{1/4}\}$ by the growth condition on the sequence $\{k_n\}$. A similar argument yields $k_n^{1/2}V_n(1) = o_P\{(\log k_n)/k_n^{1/4}\}$.

Thus, $k_n^{1/2}\Delta_n(m, l) = o_P\{(\log k_n)/k_n^{1/4}\}$, for all fixed $1 \leq m < l$, for the first term in (5.6). Then by a diagonal selection procedure, similar to that used in the proof of Lemma 1 by Viharos (1993), we can construct sequences $1 \leq m_n \leq l_n \leq k_n$ such that $m_n \rightarrow \infty$, $l_n/m_n \rightarrow \infty$, $k_n/l_n \rightarrow \infty$ and $k_n^{1/2}\Delta_n(m_n, l_n) = o_P(1)$. It follows that $k_n^{1/2}(\bar{\alpha}_n - \bar{\mu}_n) = k_n^{1/2}\Theta_n(l_n, k_n) + o_P(1)$, and an argument as in the proof of Lemma 2 in Viharos (1993) and the approximations (5.3) give that $k_n^{1/2}\Theta_n(l_n, k_n) = N_n + o_P(1)$.

Next we show that $k_n^{1/2}(\bar{\mu}_n - \mu_n^{(1)}) \rightarrow 0$. Integrating by parts, we have

$$\bar{\mu}_n = \mu_n^{(1)} - \frac{n}{k_n} \int_0^{1/n} g(t) J\left(\frac{nt}{k_n}\right) dt + \frac{g_{n,n}g(1/n)}{k_n} = \mu_n^{(1)} - \int_0^{1/n} \frac{nt}{k_n} \log\left(\frac{nt}{k_n}\right) dg(t).$$

Hence, for all n large enough,

$$k_n^{1/2} |\bar{\mu}_n - \mu_n^{(1)}| \leq -k_n^{1/2} \int_0^{1/n} \left(\frac{nt}{k_n}\right)^{3/4} dg(t) = O(k_n^{-1/4}), \tag{5.7}$$

using Lemma 3.2 of Csörgő and Mason (1985) in the last step.

Thus we established the first statement in (5.5). Finally, to prove the second, note that N_n is a normal random variable with mean zero and variance

$$\begin{aligned} E(N_n^2) &= \frac{n}{k_n} \int_{l_n/n}^{k_n/n} \int_{l_n/n}^{k_n/n} (s \wedge t - st) J\left(\frac{ns}{k_n}\right) J\left(\frac{nt}{k_n}\right) dg(s) dg(t) \\ &= \int_{l_n/k_n}^1 \int_{l_n/k_n}^1 (u \wedge v) J(u) J(v) df_{k_n/n}(u) df_{k_n/n}(v) - \frac{k_n}{n} \left(\int_{l_n/k_n}^1 u J(u) df_{k_n/n}(u) \right)^2 \\ &=: v_n - w_n. \end{aligned} \tag{5.8}$$

Using Lemma 5.4 (iv), we see that $v_n \rightarrow \int_0^1 \int_0^1 (u \wedge v) J(u) J(v) d(\log u^{-\alpha}) d(\log v^{-\alpha}) = 2\alpha^2$ and $nw_n/k_n \rightarrow \left\{ \int_0^1 u J(u) d(\log u^{-\alpha}) \right\}^2 = 0$. Hence $w_n \rightarrow 0$, and so $E(N_n^2) \rightarrow 2\alpha^2$. \square

Theorem 1.3 requires two more lemmas, the first of which is analogous to Lemma 5.4 (iv) while the one after corresponds to Lemma 2.6 of S&S, without their regularity condition. For the rest we use the notation $g(s) = \log Q(1 - s-)$, $0 < s < 1$, as above.

Lemma 5.5. Put $h_n(s) := g(s)\{g(s) - 2\mu_{5,n}^\diamond\}$ and $f_{n,s}^*(v) := h_n(sv) - h_n(s)$, where $\mu_{5,n}^\diamond = \mu_{5,n}^\diamond(k_n) := nk_n^{-1} \int_{1/n}^{k_n/n} g(t) dt + g(1/n)k_n^{-1}$. Then $f_{n,1/n}^*(v) = \alpha^2(2 + \log v) \log v + O(\log n)$ and $f_{n,k_n/n}^*(v) \rightarrow \alpha^2(2 + \log v) \log v =: f_\alpha^*(v)$ for every fixed $v > 0$.

Proof. Integrating by parts,

$$\mu_{5,n}^\diamond = g\left(\frac{k_n}{n}\right) - \frac{n}{k_n} \int_0^{k_n/n} t dg(t) + \frac{n}{k_n} \int_0^{1/n} t dg(t).$$

By Lemma 3.1 of Csörgő and Mason (1985), $\mu_{5,n}^\diamond = g(k_n/n) + \alpha + o(1)$, and so $f_{n,s}^*(v) = \{g(sv) - g(s)\}\{g(sv) - g(s) - 2\alpha + o(1)\} + 2\{g(sv) - g(s)\}\{g(s) - g(k_n/n)\}$. Since $g(s) = O(\log s)$ as $s \downarrow 0$, by Lemma 5.1, both statements follow from Lemma 5.4 (iv). \square

Lemma 5.6. Let $W_n = W_n(k_n) := k_n^{-1} \sum_{i=1}^{k_n} g^2(U_{i,n}) - \{k_n^{-1} \sum_{i=1}^{k_n} g(U_{i,n})\}^2$ and recall that $\mu_n = \mu_n(k_n) = \mu_{4,n}(k_n) - \mu_{5,n}^2(k_n) = \mu_n^{(1)}(k_n)\mu_n^{(3)}(k_n)$. Then

$$k_n^{1/2}(W_n - \mu_n) = N_n^* + o_P(1),$$

where

$$N_n^* = N_n^*(l_n, k_n) := -\left(\frac{n}{k_n}\right)^{1/2} \int_{l_n/n}^{k_n/n} B_n(t) dh_n(t)$$

and $N_n^* \xrightarrow{\mathcal{D}} N(0, 8\alpha^4)$, for the same sequence $\{l_n\}$ for which (5.5) holds.

Proof. Set $\mu_n^\diamond = \mu_n^\diamond(k_n) := \mu_{4,n}^\diamond(k_n) - \{\mu_{5,n}^\diamond(k_n)\}^2$, where $\mu_{4,n}^\diamond(k_n) := nk_n^{-1} \int_{1/n}^{k_n/n} g^2(t) dt + g^2(1/n)k_n^{-1}$ and $\mu_{5,n}^\diamond(k_n)$ is as in Lemma 5.5. Then, with $h_n(\cdot)$ from the same lemma, $W_n - \mu_n^\diamond = k_n^{-1} \sum_{i=1}^{k_n} h_n(U_{i,n}) - \mu_n^* - R_n^\diamond$, where

$$\mu_n^* = \mu_n^*(k_n) := \frac{n}{k_n} \int_{1/n}^{k_n} h_n(s) ds + \frac{h_n(1/n)}{k_n}, \quad R_n^\diamond = R_n^\diamond(k_n) := \left(\frac{1}{k_n} \sum_{i=1}^{k_n} g(U_{i,n}) - \mu_{5,n}^\diamond \right)^2.$$

Here $R_n^\diamond = O_P(1/k_n)$ by Corollary 1.1 of Viharos (1995), and $k_n^{1/2}(\mu_n - \mu_n^\diamond) \rightarrow 0$ as in (5.7). Similarly as in (5.6), we now obtain

$$\frac{1}{k_n} \sum_{i=1}^{k_n} h_n(U_{i,n}) - \mu_n^* = \Theta_n^*(1, m) - V_n^*(1) + R_n^* + \Theta_n^*(m, l) + V_n^*(k_n) + \Theta_n^*(l, k_n)$$

for fixed integers m and l such that $1 \leq m \leq l \leq k_n$, where, for $1 \leq m \leq r \leq n$,

$$\Theta_n^*(m, r) := -\frac{n}{k_n} \int_{m/n}^{r/n} \{G_n(t) - t\} dh_n(t), \quad V_n^*(m) := \frac{n}{k_n} \int_{U_{m,n}}^{m/n} \left(G_n(t) - \frac{m}{n}\right) dh_n(t)$$

and $R_n^* := \{h_n(U_{1,n}) - h_n(1/n)\}/k_n$. Using now Lemma 5.5 instead of Lemma 5.4 (iv), an analogue of the proof of (5.5) yields the first statement of Lemma 5.6, where, executing the two proofs jointly, the sequence $\{l_n\}$ can be and is chosen as the $\{l_n\}$ in (5.5).

Also, $E\{(N_n^*)^2\} \sim v_n^* := \int_{l_n/k_n}^1 \int_{l_n/k_n}^1 (u \wedge v) df_{n,k_n/n}^*(u) df_{n,k_n/n}^*(v)$, similarly as in (5.8), where, using Lemma 5.5, $v_n^* \rightarrow \int_0^1 \int_0^1 (u \wedge v) df_\alpha^*(u) df_\alpha^*(v) = 8\alpha^4$. □

Proof of Theorem 1.3. Recall $\tilde{\alpha}_n$ from (5.2). Since $\mu_n^{(1)} \rightarrow \alpha$ and $\tilde{\alpha}_n \rightarrow \alpha$ in probability by Theorem 1.1 and Lemma 5.4, and $\mu_n \rightarrow \alpha^2$ as noted in Section 3, by another application of Lemma 5.4 and the facts that $k_n^{1/2}(\tilde{\alpha}_n - \mu_n^{(1)}) = O_P(1)$ in (5.5) and $k_n^{1/2}(W_n - \mu_n) = O_P(1)$ in Lemma 5.6 we see that

$$\begin{aligned} k_n^{1/2}(\hat{\alpha}_n^{(3)} - \mu_n^{(3)}) &\stackrel{\mathcal{O}}{=} \frac{1}{\tilde{\alpha}_n} k_n^{1/2}(W_n - \mu_n) - \frac{\mu_n}{\tilde{\alpha}_n \mu_n^{(1)}} k_n^{1/2}(\tilde{\alpha}_n - \mu_n^{(1)}) \\ &= \frac{1}{\alpha} k_n^{1/2}(W_n - \mu_n) - k_n^{1/2}(\tilde{\alpha}_n - \mu_n^{(1)}) + o_P(1). \end{aligned}$$

Hence $k_n^{1/2}(\hat{\alpha}_n^{(3)} - \mu_n^{(3)}) \stackrel{\mathcal{O}}{=} \tilde{N}_n + o_P(1)$, where $\tilde{N}_n := \alpha^{-1}N_n^* - N_n$ by (5.5) and Lemma 5.6 again. Thus, setting $c_n := \int_{l_n/k_n}^1 \int_{l_n/k_n}^1 (u \wedge v) J(u) df_{k_n/n}(u) df_{n,k_n/n}^*(v)$, we now obtain $E(\tilde{N}_n^2) \sim \alpha^{-2}v_n^* + v_n - 2\alpha^{-1}c_n$ by (5.8), the end of the previous proof, and by similar considerations showing $E(N_n N_n^*) \sim c_n$. Here, $c_n \rightarrow \int_0^1 \int_0^1 (u \wedge v) J(u) d(\log u^{-\alpha}) df_\alpha^*(v) = 4\alpha^3$, using Lemmas 5.4 and 5.5, and so $E(\tilde{N}_n^2) \rightarrow 8\alpha^2 + 2\alpha^2 - 8\alpha^2 = 2\alpha^2$. □

Proof of Theorem 4.1. First we prove the theorem for $\hat{\alpha}_n^{(1)}$. Under (4.1) we obtain $S(1 - s-) = \log Q(1 - s-) = -\alpha \log s + \log D_1 + D_2 s^\beta \{1 + \varepsilon(s)\}$, where $\varepsilon(s) \rightarrow 0$ as $s \downarrow 0$. Whence

$$\begin{aligned}\mu_n^{(1)}(k_n) - \alpha &= D_2 \left(\frac{k_n}{n}\right)^\beta \int_0^1 v^\beta \left\{ 1 + \varepsilon \left(\frac{k_n v}{n}\right) \right\} (-1 - \log v) \, dv \\ &= \{1 + o(1)\} D_2 \left(\frac{k_n}{n}\right)^\beta \int_0^1 v^\beta (-1 - \log v) \, dv = \{1 + o(1)\} b_{1,n}(k_n),\end{aligned}$$

where $b_{1,n}(k_n) = -D_2 \beta (k_n/n)^\beta / (\beta + 1)^2$, so that $k_n = k_n^*$ is chosen as the integer sequence for which $M_n^{(1)}(k_n) = b_{1,n}^2(k_n) + 2\alpha^2 k_n^{-1}$ is minimized. Then we get (4.2), from which (4.3) follows for $j = 1$. The proof for $j = 3$ is similar. Finally, under (4.1),

$$\mu_n^{(2)}(k_n) - \alpha = -\frac{1}{j_n(k_n)} \int_0^{k_n/n} \log(D_1[1 + D_2 t^\beta \{1 + o(1)\}]) \log t \, dt \sim \frac{-\log D_1}{\log(k_n/n)},$$

which implies the last statement. □

Acknowledgements

S. Csörgő was supported in part by the US National Science Foundation Grant DMS-9208067 and Grant DMS-9625732, and L. Viharos was supported in part by the Hungarian National Science Foundation Grant F016226. We thank Josef Steinebach for the preprint of S&S and two referees for their remarks.

References

- Beirlant, J. and Teugels, J.L. (1989) Asymptotic normality of Hill's estimator. *Extreme Value Theory, Oberwolfach, 1987*, pp. 148–155. Lecture Notes Statist. 51. New York: Springer-Verlag.
- Bingham, N.H., Goldie, C.M. and Teugels, J.L. (1987) *Regular Variation*. Cambridge: Cambridge University Press.
- Csörgő, M., Csörgő, S., Horváth, L. and Mason, D.M. (1986) Weighted empirical and quantile processes. *Ann. Probab.*, **14**, 31–85.
- Csörgő, S. (1984) Adaptive estimation of the parameters of stable laws. In P. Révész (ed.) *Limit Theorems in Probability and Statistics, 1982*, pp. 305–367. Colloq. Math. Soc. János Bolyai 36. Amsterdam: North-Holland.
- Csörgő, S., Deheuvels, P. and Mason, D.M. (1985) Kernel estimates of the tail index of a distribution. *Ann. Statist.*, **13**, 1050–1077.
- Csörgő, S., Haeusler, E. and Mason, D.M. (1991) The asymptotic distribution of extreme sums. *Ann. Probab.*, **19**, 783–811.
- Csörgő, S. and Mason, D.M. (1985) Central limit theorems for sums of extreme values. *Math. Proc. Cambridge Philos. Soc.*, **98**, 547–558.
- Csörgő, S. and Viharos, L. (1995) On the asymptotic normality of Hill's estimator. *Math. Proc. Cambridge Philos. Soc.*, **118**, 375–382.
- Deheuvels, P., Haeusler, E. and Mason, D.M. (1988) Almost sure convergence of the Hill estimator. *Math. Proc. Cambridge Philos. Soc.*, **104**, 371–381.
- de Haan, L. (1970) *On Regular Variation and its Application to the Weak Convergence of Sample*

- Extremes*. Math. Centre Tract 32. Amsterdam: Mathematisch Centrum.
- Haeusler, E. and Teugels, J.L. (1985) On the asymptotic normality of Hill's estimator for the exponent of regular variation. *Ann. Statist.*, **13**, 743–756.
- Hall, P. (1982) On some simple estimates of an exponent of regular variation. *J. Roy. Statist. Soc., Ser. B*, **44**, 37–42.
- Hall, P. and Welsh, A.H. (1985) Adaptive estimates of parameters of regular variation. *Ann. Statist.*, **13**, 331–341.
- Hill, B.M. (1975) A simple general approach to inference about the tail of a distribution. *Ann. Statist.*, **3**, 1163–1174.
- Lo, G.S. (1989) A note on the asymptotic normality of sums of extreme values. *J. Stat. Plan. Inference*, **22**, 127–136.
- Mason, D.M. (1982) Laws of large numbers for sums of extreme values. *Ann. Probab.*, **10**, 754–764.
- Mason, D.M. and van Zwet, W.R. (1987) A refinement of the KMT inequality for the uniform empirical process. *Ann. Probab.*, **15**, 871–884.
- Novak, S. Yu. and Utev, S.A. (1990) Asymptotics of the distribution of the ratio of sums of random variables. *Siberian Math. J.*, **31**, 781–788.
- Schultze, J. and Steinebach, J. (1996) On least squares estimates of an exponential tail coefficient. *Statist. Decisions*, **14**, 353–372.
- Viharos, L. (1993) Asymptotic distributions of linear combinations of extreme values. *Acta Sci. Math. (Szeged)*, **58**, 211–231.
- Viharos, L. (1995) Limit theorems for linear combinations of extreme values with applications to inference about the tail of a distribution. *Acta Sci. Math. (Szeged)*, **60**, 761–777.
- Viharos, L. (1997) Estimators of the exponent of regular variation with universally normal asymptotic distributions. *Math. Methods Statist.* To appear.

Received February 1996 and revised August 1996