Journal of the Korean Statistical Society (1999), **28**: *1, pp 57–72*

On Efficient Estimation of the Extreme Value Index with Good Finite-Sample Performance

Seokhoon Yun[1]

ABSTRACT

Falk (1994) showed that the asymptotic efficiency of the Pickands estimator of the extreme value index β can considerably be improved by a simple convex combination. In this paper we propose an alternative estimator of β which is as asymptotically efficient as the optimal convex combination of the Pickands estimators but has a better finite-sample performance. We prove consistency and asymptotic normality of the proposed estimator. Monte Carlo simulations are conducted to compare the finite-sample performances of the proposed estimator and the optimal convex combination estimator.

Keywords: Extreme value index; Pickands estimator; δ-neighborhood; generalized Pareto distribution; consistency; asymptotic normality

1. INTRODUCTION

Let $X_1, ..., X_n$ be an i.i.d. sample from a distribution function F. Suppose that F belongs to the domain of attraction of an extreme value distribution G_β for some $\beta \in \Re$ (in short, $F \in \mathcal{D}(G_\beta)$), where

$$G_\beta(x) := \exp\{-(1+\beta x)^{-1/\beta}\},\ 1+\beta x > 0,$$

that is, for some constants $a_n > 0$ and $b_n \in \Re$,

$$a_n^{-1}(\max\{X_1, ..., X_n\} - b_n) \xrightarrow{d} G_\beta \text{ as } n \to \infty. \tag{1.1}$$

By $\xrightarrow{d}$ we denote convergence in distribution. The case $\beta = 0$ is always interpreted as the limit $\beta \to 0$ throughout the paper, i.e., $G_0(x) = \exp(-e^{-x})$, $x \in \Re$.

The β is called the extreme value index and estimation of the parameter β based on the sample $X_1, ..., X_n$ has been extensively studied in the literature (see, e.g., Pickands (1975), Hill (1975), Smith (1987), Dekkers and de Haan (1989),

[1]Department of Applied Statistics, University of Suwon, Suwon, Kyonggi-do, 445-743, Korea.

Falk (1994), and Drees (1995)). If one knows that $\beta > 0$, one can use the well-known Hill estimator (see Hill (1975)). Otherwise, one can use the Pickands estimator (see Pickands (1975)) defined by

$$\hat{\beta}_n(m) := \frac{1}{\log 2} \log \left(\frac{X_{n-m+1:n} - X_{n-2m+1:n}}{X_{n-2m+1:n} - X_{n-4m+1:n}} \right),$$

where $1 \leq m \leq n/4$ and $X_{1:n} \leq \cdots \leq X_{n:n}$ denote the order statistics of $X_1, ..., X_n$. Pickands proved weak consistency of this estimator for any $\beta \in \Re$ and any sequence of integers $m = m(n) \to \infty$ such that $m/n \to 0$ as $n \to \infty$, and Dekkers and de Haan (1989) proved strong consistency for any sequence $m = m(n)$ such that $m/n \to 0$ and $m/\log\log n \to \infty$ as $n \to \infty$. The Pickands estimator is moreover invariant under the choice of a_n and b_n used in (1.1).

Asymptotic normality of the Pickands estimator also holds under additional conditions on F (see, e.g., Dekkers and de Haan (1989) and Falk (1994)). However it has a rather poor asymptotic efficiency. Falk (1994) showed that the asymptotic variance of the Pickands estimator can considerably be reduced by a simple convex combination as

$$\begin{aligned} \hat{\beta}_n^{(F)}(m,p) &:= p \cdot \hat{\beta}_n([m/2]) + (1-p) \cdot \hat{\beta}_n(m) \\ &= \frac{1}{\log 2} \log \left\{ \left(\frac{X_{n-[m/2]+1:n} - X_{n-2[m/2]+1:n}}{X_{n-2[m/2]+1:n} - X_{n-4[m/2]+1:n}} \right)^p \right. \\ &\qquad \left. \times \left(\frac{X_{n-m+1:n} - X_{n-2m+1:n}}{X_{n-2m+1:n} - X_{n-4m+1:n}} \right)^{1-p} \right\}, \end{aligned}$$

where $p \in [0,1]$, $2 \leq m \leq n/4$, and $[x]$ denotes the integer part of $x \in \Re$. Specifically, assuming that F is in a δ-neighborhood of a generalized Pareto distribution (GPD) H_β (see Section 2 for the definition), where

$$H_\beta(x) := 1 - (1+\beta x)^{-1/\beta}, \ x \geq 0, \ 1 + \beta x > 0,$$

which is a stronger assumption than $F \in \mathcal{D}(G_\beta)$, he showed that, for any sequence $m = m(n) \to \infty$ such that $(m/n)^\delta \sqrt{m} \to 0$ as $n \to \infty$,

$$\sqrt{m}(\hat{\beta}_n^{(F)}(m,p) - \beta) \xrightarrow{d} N(0, \sigma_\beta^2 \cdot (\nu_\beta^{(F)}(p))^2) \text{ as } n \to \infty, \tag{1.2}$$

where

$$\sigma_\beta^2 := \frac{1 + 2^{-2\beta-1}}{2\log^2 2} \left(\frac{\beta}{1 - 2^{-\beta}} \right)^2$$

and

$$(\nu_\beta^{(F)}(p))^2 := 1 + p^2\left(3 + \frac{4\cdot 2^{-\beta}}{2^{-2\beta}+2}\right) - p\left(2 + \frac{4\cdot 2^{-\beta}}{2^{-2\beta}+2}\right).$$

Interpret $\sigma_0^2 = \lim_{\beta\to 0}\sigma_\beta^2 = 3/(4\log^4 2)$. Note here that σ_β^2 is the asymptotic variance of $\sqrt{m}(\hat\beta_n(m)-\beta)$. Since $\hat\beta_n^{(F)}(m,0) = \hat\beta_n(m)$, $\hat\beta_n^{(F)}(m,p)$ is obviously a great improvement on $\hat\beta_n(m)$ if p is chosen appropriately.

In this paper we consider a different mixture of the form

$$\hat\beta_n(m,a) := \frac{1}{\log 2}\log\left\{\frac{a(X_{n-[m/2]+1:n} - X_{n-2[m/2]+1:n}) + (X_{n-m+1:n} - X_{n-2m+1:n})}{a(X_{n-2[m/2]+1:n} - X_{n-4[m/2]+1:n}) + (X_{n-2m+1:n} - X_{n-4m+1:n})}\right\}, \quad (1.3)$$

where $a \geq 0$ and $2 \leq m \leq n/4$, which uses the same order statistics as the Falk estimator $\hat\beta_n^{(F)}(m,p)$. Since $\hat\beta_n(m,0) = \hat\beta_n(m)$, $\hat\beta_n(m,a)$ is also an extension of the Pickands estimator. We prove weak and strong consistency of $\hat\beta_n(m,a)$ for any $\beta \in \Re$ under the sole condition $F \in \mathcal{D}(G_\beta)$. Assuming that F is in a δ-neighborhood of a GPD H_β, we also prove that $\sqrt{m}(\hat\beta_n(m,a)-\beta)$ is asymptotically normal for any sequence $m = m(n) \to \infty$ such that $(m/n)^\delta\sqrt{m} \to 0$ as $n \to \infty$. It turns out that the estimator $\hat\beta_n(m,a)$ has the same asymptotic performance as the estimator $\hat\beta_n^{(F)}(m,p)$ if a and p are chosen in optimal ways, respectively. Moreover, the estimator $\hat\beta_n(m,a)$ with optimal a turns out to have a better finite-sample performance than the estimator $\hat\beta_n^{(F)}(m,p)$ with optimal p.

The rest of the paper is organized as follows. In Section 2 we establish (weak and strong) consistency and asymptotic normality of $\hat\beta_n(m,a)$. Further we determine the optimal $a^*(\beta)$ which minimizes the asymptotic variance of $\sqrt{m}(\hat\beta_n(m,a)-\beta)$ and investigate the asymptotic behavior of the data-driven optimal estimator $\hat\beta_n(m,a^*(\tilde\beta_n))$, where $\tilde\beta_n$ is a weakly consistent estimator of β. In Section 3 we compare the finite-sample performance of $\hat\beta_n(m,a^*(\tilde\beta_n))$ with that of $\hat\beta_n^{(F)}(m,p^*(\tilde\beta_n))$ by various Monte Carlo simulations, where $p^*(\beta)$ is the optimal p minimizing the asymptotic variance of $\sqrt{m}(\hat\beta_n^{(F)}(m,p)-\beta)$. In Section 4 we briefly mention the possible extension of $\hat\beta_n(m,a)$ to a higher mixture form. All proofs are collected in the Appendix.

2. CONSISTENCY AND ASYMPTOTIC NORMALITY

First, we establish weak consistency of $\hat\beta_n(m,a)$ under the sole condition $F \in \mathcal{D}(G_\beta)$. By $\xrightarrow{p}$ we denote convergence in probability.

Theorem 2.1. (Weak consistency). *Suppose that $F \in \mathcal{D}(G_\beta)$ for some $\beta \in \Re$. Then, for $a \geq 0$ and any sequence $m = m(n) \to \infty$ such that $m/n \to 0$ as $n \to \infty$,*

$$\hat{\beta}_n(m,a) \xrightarrow{P} \beta \text{ as } n \to \infty.$$

If the sequence $m = m(n)$ increases suitably rapidly, then strong consistency of $\hat{\beta}_n(m,a)$ also holds.

Theorem 2.2. (Strong consistency). *Suppose that $F \in \mathcal{D}(G_\beta)$ for some $\beta \in \Re$. Then, for $a \geq 0$ and any sequence $m = m(n)$ such that $m/n \to 0$ and $m/\log\log n \to \infty$ as $n \to \infty$,*

$$\hat{\beta}_n(m,a) \to \beta \text{ a.s. as } n \to \infty.$$

For asymptotic normality of $\hat{\beta}_n(m,a)$, we assume a δ-neighborhood of a GPD as Falk (1994) did in his paper. For $\delta > 0$, F is said to be in a δ-neighborhood of a GPD H_β (in short, $F \in Q(\delta; H_\beta)$) if $x_F = x_{H_\beta}$ and F has a density f on $[x_0, x_F)$ for some $x_0 < x_F$ such that

$$f(x) = h_\beta(x)(1 + O((1 - H_\beta(x))^\delta)), \ x \in [x_0, x_F),$$

where h_β denotes the density of H_β and $x_F := \sup\{x : F(x) < 1\}$, the right endpoint of F. The connection of δ-neighborhoods of GPD's with rates of convergence of extremes is well described in Falk, Hüsler and Reiss (1994). We now establish asymptotic normality of $\hat{\beta}_n(m,a)$ in terms of variational distance, which then gives the rate of convergence to the normal distribution.

Lemma 2.1. *Suppose that $F \in Q(\delta; H_\beta)$ for some $\delta > 0$ and $\beta \in \Re$. Then, for $a \geq 0$, $m \in \{2, ..., n/4\}$ and $n \in \{8, 9, ...\}$,*

$$\sup_{B \in \mathcal{B}} |P\{\sqrt{m}(\hat{\beta}_n(m,a) - \beta) \in B\} - P\{\sigma_\beta \nu_\beta(a) Z + O_P(1/\sqrt{m}) \in B\}|$$
$$= O((m/n)^\delta \sqrt{m} + m/n + 1/\sqrt{m}),$$

where

$$\nu_\beta^2(a) := 1 + \frac{a^2}{(a + 2^{-\beta})^2} - \frac{a \cdot 2^{-\beta}}{(a + 2^{-\beta})^2}\left(2 + \frac{4 \cdot 2^{-\beta}}{2^{-2\beta} + 2}\right),$$

Z is a standard normal random variable and $\mathcal{B}$ denotes the Borel σ-field in $\Re$.

The following result is an immediate consequence of Lemma 2.1.

Theorem 2.3. (Asymptotic normality). *Suppose that $F \in Q(\delta; H_\beta)$ for some $\delta > 0$ and $\beta \in \Re$. Then, for $a \geq 0$ and any sequence $m = m(n) \to \infty$ such that $(m/n)^\delta \sqrt{m} \to 0$ as $n \to \infty$,*

$$\sqrt{m}(\hat{\beta}_n(m,a) - \beta) \xrightarrow{d} N(0, \sigma_\beta^2 \cdot \nu_\beta^2(a)) \text{ as } n \to \infty. \tag{2.1}$$

Note here that (2.1) includes the asymptotic normality of the Pickands estimator $\hat{\beta}_n(m) = \hat{\beta}_n(m,0)$. From (1.2), $(\nu_\beta^{(F)}(p))^2$ is the asymptotic relative efficiency (ARE) of $\hat{\beta}_n(m)$ with respect to $\hat{\beta}_n^{(F)}(m,p)$, which is defined by the ratio of the asymptotic variances of $\sqrt{m}(\hat{\beta}_n^{(F)}(m,p) - \beta)$ and $\sqrt{m}(\hat{\beta}_n(m) - \beta)$. Also, from (2.1), $\nu_\beta^2(a)$ is the ARE of $\hat{\beta}_n(m)$ with respect to $\hat{\beta}_n(m,a)$.

Now the optimal choice of p minimizing $(\nu_\beta^{(F)}(p))^2$ is

$$p^*(\beta) := \frac{(2^{-2\beta} + 2) + 2 \cdot 2^{-\beta}}{3(2^{-2\beta} + 2) + 4 \cdot 2^{-\beta}},$$

in which case $(\nu_\beta^{(F)}(p))^2$ becomes

$$(\nu_\beta^{(F)}(p^*(\beta)))^2 = 1 - p^*(\beta)\left(1 + \frac{2 \cdot 2^{-\beta}}{2^{-2\beta} + 2}\right).$$

Similarly, the optimal choice of a minimizing $\nu_\beta^2(a)$ is

$$a^*(\beta) := \frac{2^{-\beta}(2^{-2\beta} + 2 \cdot 2^{-\beta} + 2)}{2(2^{-2\beta} + 2^{-\beta} + 2)},$$

in which case $\nu_\beta^2(a)$ becomes

$$\nu_\beta^2(a^*(\beta)) = 1 - \frac{(2^{-2\beta} + 2 \cdot 2^{-\beta} + 2)^2}{(2^{-2\beta} + 2)(3 \cdot 2^{-2\beta} + 4 \cdot 2^{-\beta} + 6)}.$$

It is interesting to observe that $(\nu_\beta^{(F)}(p^*(\beta)))^2 = \nu_\beta^2(a^*(\beta)) < 1$ for all $\beta \in \Re$. In fact, $\min_{\beta \in \Re} \nu_\beta^2(a^*(\beta)) = 0.34$ (approximately) and $\sup_{\beta \in \Re} \nu_\beta^2(a^*(\beta)) = 2/3$. These imply that $\hat{\beta}_n^{(F)}(m, p^*(\beta))$ and $\hat{\beta}_n(m, a^*(\beta))$ are obviously superior to the Pickands estimator $\hat{\beta}_n(m)$ and that they have exactly the same asymptotic performance.

However the optimal p and optimal a depend on the unknown parameter β which is to be estimated. This suggests utilizing an adaptive estimator. Falk (1994) gave the following result which says that the estimator $\hat{\beta}_n^{(F)}(m, p^*(\tilde{\beta}_n))$ has the same asymptotic performance as $\hat{\beta}_n^{(F)}(m, p^*(\beta))$ if $\tilde{\beta}_n$ is a weakly consistent estimator of β.

Theorem 2.4. *Suppose that $F \in Q(\delta; H_\beta)$ for some $\delta > 0$ and $\beta \in \Re$. Then, for any sequence $m = m(n) \to \infty$ such that $(m/n)^\delta \sqrt{m} \to 0$ as $n \to \infty$,*

$$\sqrt{m}(\hat{\beta}_n^{(F)}(m, p^*(\tilde{\beta}_n)) - \beta) \xrightarrow{d} N(0, \sigma_\beta^2 \cdot (\nu_\beta^{(F)}(p^*(\beta)))^2) \text{ as } n \to \infty$$

if $\tilde{\beta}_n$ is a weakly consistent estimator of β.

We now show that the estimator $\hat{\beta}_n(m, a^*(\tilde{\beta}_n))$ also has the same asymptotic performance as $\hat{\beta}_n(m, a^*(\beta))$ and thus that $\hat{\beta}_n(m, a^*(\tilde{\beta}_n))$ and $\hat{\beta}_n^{(F)}(m, p^*(\tilde{\beta}_n))$ have equal asymptotic performance if $\tilde{\beta}_n$ is a weakly consistent estimator of β. For this we need the following lemma.

Lemma 2.2. *Suppose that $F \in Q(\delta; H_\beta)$ for some $\delta > 0$ and $\beta \in \Re$. Then, for any sequence $m = m(n) \to \infty$ such that $(m/n)^\delta \sqrt{m} \to 0$ as $n \to \infty$,*

$$\sqrt{m}(\hat{\beta}_n(m, a^*(\tilde{\beta}_n)) - \hat{\beta}_n(m, a^*(\beta))) = o_P(1) \text{ as } n \to \infty$$

if $\tilde{\beta}_n$ is a weakly consistent estimator of β.

The following result is an easy consequence of Lemma 2.2 and (2.1).

Theorem 2.5. *Suppose that $F \in Q(\delta; H_\beta)$ for some $\delta > 0$ and $\beta \in \Re$. Then, for any sequence $m = m(n) \to \infty$ such that $(m/n)^\delta \sqrt{m} \to 0$ as $n \to \infty$,*

$$\sqrt{m}(\hat{\beta}_n(m, a^*(\tilde{\beta}_n)) - \beta) \xrightarrow{d} N(0, \sigma_\beta^2 \cdot \nu_\beta^2(a^*(\beta))) \text{ as } n \to \infty$$

if $\tilde{\beta}_n$ is a weakly consistent estimator of β.

3. FINITE-SAMPLE PERFORMANCE

In this section the finite-sample performance of the new estimator $\hat{\beta}_n(m, a^*(\tilde{\beta}_n))$ is compared with that of the Falk estimator $\hat{\beta}_n^{(F)}(m, p^*(\tilde{\beta}_n))$ by various Monte Carlo simulations.

Falk (1994) proposed $\tilde{\beta}_n = \hat{\beta}_n^{(F)}(m, p^*(0)) = \hat{\beta}_n^{(F)}(m, 5/13)$ as an initial estimator. This is quite reasonable since the parameter $\beta = 0$ is crucial as it is some kind of change point: if $\beta < 0$, then the right endpoint x_F of F is finite, while in case $\beta > 0$ x_F is infinite. By the same reason we propose $\tilde{\beta}_n = \hat{\beta}_n(m, a^*(0)) = \hat{\beta}_n(m, 5/8)$ as an initial estimator when we deal with $\hat{\beta}_n(m, a^*(\tilde{\beta}_n))$. As a consequence, we compare $\hat{\beta}_n^{(F)}(m, \hat{p}^*)$ with $\hat{\beta}_n(m, \hat{a}^*)$, where $\hat{p}^* = p^*(\hat{\beta}_n^{(F)}(m, 5/13))$ and $\hat{a}^* = a^*(\hat{\beta}_n(m, 5/8))$.

The study is based on $k = 1000$ Monte Carlo simulations. In each simulation $t = 1, ..., k$, we generated $n = 50, 100, 200, 400$ replicates $X_1, ...X_n$ of a (pseudo-) random variable X with different distribution function F in each case; we then computed the three estimators $\hat{\beta}_n(m)$, $\hat{\beta}_n^{(F)}(m, \hat{p}^*)$, and $\hat{\beta}_n(m, \hat{a}^*)$ of the pertaining values of β with $m = 8, 10, 14, 16$, and stored by

$$B_t := |\hat{\beta}_n(m) - \beta|, \ C_t := |\hat{\beta}_n^{(F)}(m, \hat{p}^*) - \beta|, \ D_t := |\hat{\beta}_n(m, \hat{a}^*) - \beta|$$

their corresponding absolute errors. By $B_{1:k} \le \cdots \le B_{k:k}$, $C_{1:k} \le \cdots \le C_{k:k}$, and $D_{1:k} \le \cdots \le D_{k:k}$ we denote the ordered values of $(B_t)_{t=1}^k$, $(C_t)_{t=1}^k$, and $(D_t)_{t=1}^k$, respectively. Figures 1∼4 display the corresponding sample quantile functions

$$(t/(k+1), B_{t:k}), \ (t/(k+1), C_{t:k}), \ (t/(k+1), D_{t:k}), \ t = 1, ..., k,$$

which now visualize the concentration of the three estimators around β.

In Figure 3.1, F is the triangular distribution, that is, X is the sum of two independent $\mathcal{U}(0,1)$-distributed random variables ($\beta = -0.5$ and $F \in Q(\delta; H_{-0.5})$ for any $\delta > 0$); in Figure 3.2, F is the standard Gumbel distribution, i.e., $F = G_0$ ($\beta = 0$ and $F \in Q(1; H_0)$); in Figure 3.3, F is the Cauchy distribution ($\beta = 1$ and $F \in Q(2; H_1)$). In Figure 3.4, F is the standard normal distribution, in which case $\beta = 0$ and F does not belong to any $Q(\delta; H_0)$, but it can be seen that the asymptotic normality in Theorem 2.5 is true for any sequence $m = m(n) \to \infty$ such that $\sqrt{m}\log^2(m+1)/\log n \to 0$ as $n \to \infty$ (see Example 2.33 of Falk (1986)).

The figures clearly show that the new estimator $\hat{\beta}_n(m, \hat{a}^*)$ has the best finite-sample performance. The improvement by using $\hat{\beta}_n(m, \hat{a}^*)$ against $\hat{\beta}_n^{(F)}(m, \hat{p}^*)$ is good particularly when $\beta \le 0$. We have done extensive simulations covering a wide range of distributions for F whose plots are omitted here. According to these, gamma and logistic distributions for instance have shown very similar performance to that of Figure 3.2, whereas the performances of $F = G_{-1}$ and $F = G_1$ are similar to those of Figure 3.1 and Figure 3.3, respectively.

4. CONCLUDING REMARKS

Recently, Drees (1995) considered a higher linear combination of the Pickands estimators like

$$\sum_{i=1}^{k} p_i \hat{\beta}_n([m/2^{i-1}]), \ p_i \ge 0, \ \sum_{i=1}^{k} p_i = 1,$$

and showed that the combination estimator with optimal p_i has a better asymptotic performance than the Falk estimator $\hat{\beta}_n^{(F)}(m, p^*(\tilde{\beta}_n))$. This is not surprising

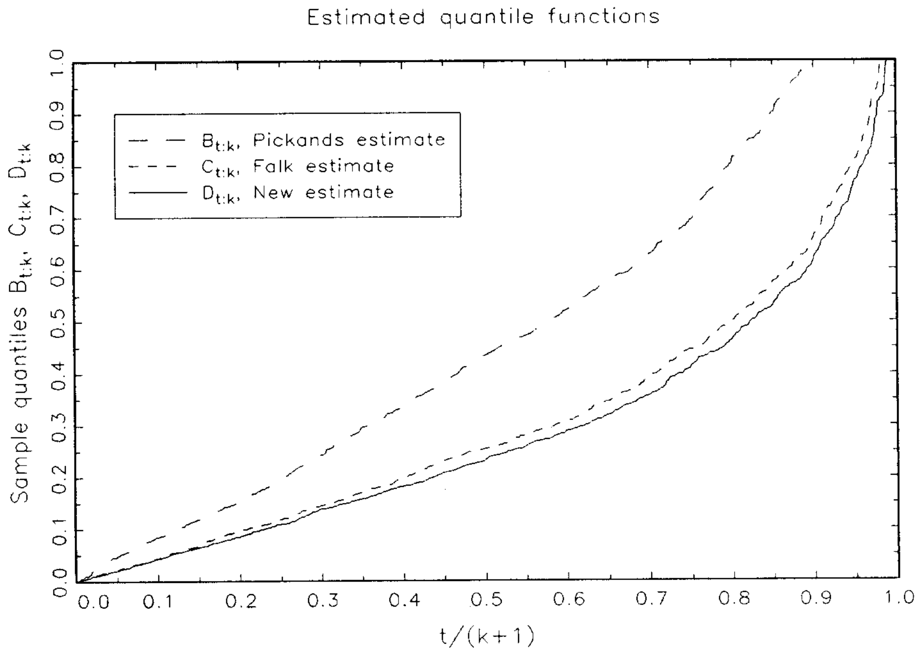

Figure 3.1: F = triangular distribution ($\beta = -0.5$), $n = 50$, $m = 8$

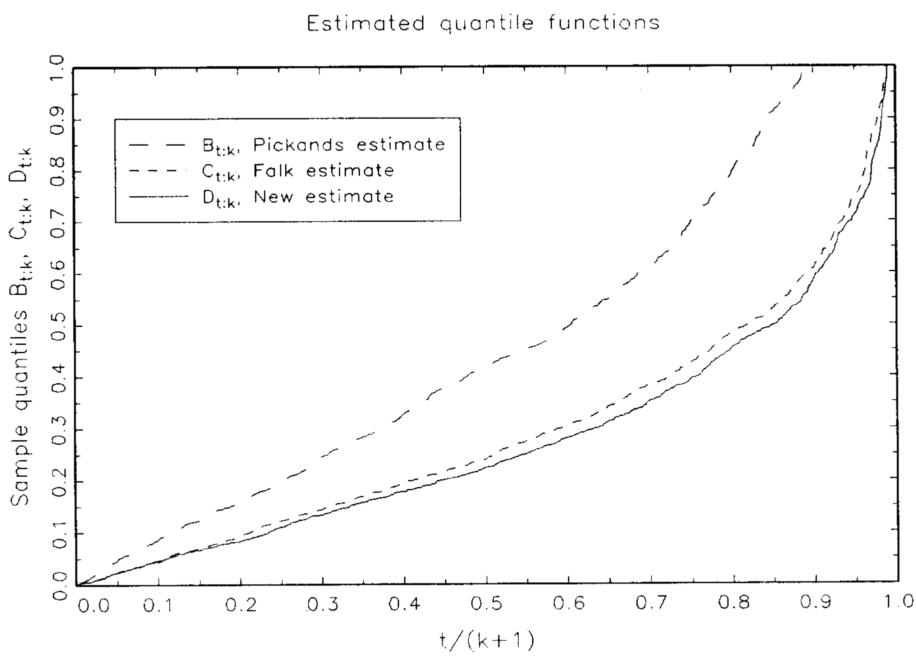

Figure 3.2: F = Gumbel distribution ($\beta = 0$), $n = 100$, $m = 10$

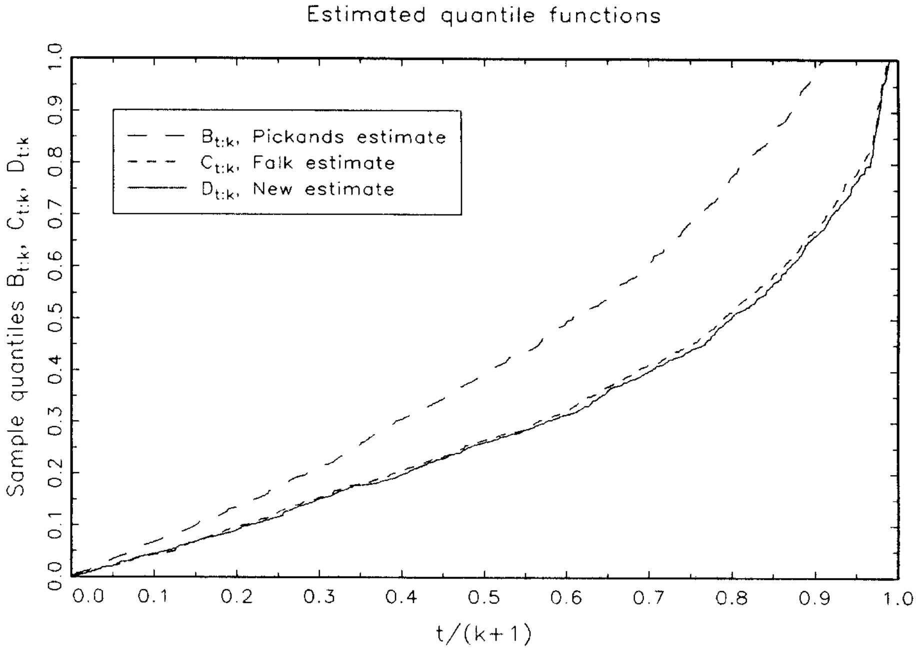

Figure 3.3: F = Cauchy distribution ($\beta = 1$), $n = 200$, $m = 14$

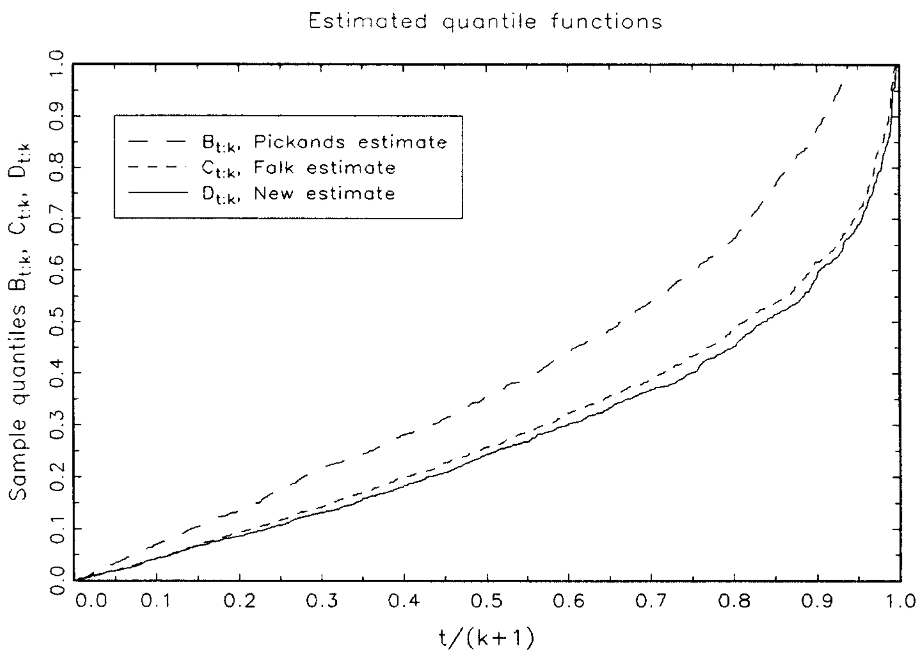

Figure 3.4: F = normal distribution ($\beta = 0$), $n = 400$, $m = 16$

since the higher linear combination estimator uses more number of observations than the Falk estimator.

Likewise, one may extend (1.3) to a higher mixture of the form

$$\frac{1}{\log 2}\log\left\{\frac{\sum_{i=1}^{k} a_i(X_{n-[m/2^{i-1}]+1:n} - X_{n-2[m/2^{i-1}]+1:n})}{\sum_{i=1}^{k} a_i(X_{n-2[m/2^{i-1}]+1:n} - X_{n-4[m/2^{i-1}]+1:n})}\right\},$$

where $a_i \geq 0$ with $a_1 = 1$. However, it is clear that the theoretical details will be much more complicated.

APPENDIX: PROOFS

We need the following well-known result (see, e.g., de Haan (1984)) to prove Theorem 2.1.

Lemma A.1. *For some $\beta \in \Re$, $F \in \mathcal{D}(G_\beta)$ if and only if*

$$\lim_{t\downarrow 0}\frac{F^{-1}(1-tx) - F^{-1}(1-t)}{F^{-1}(1-ty) - F^{-1}(1-t)} = \frac{x^{-\beta}-1}{y^{-\beta}-1} \text{ locally uniformly}$$

for $x, y > 0$ with $y \neq 1$, where F^{-1} denotes the quantile function of F.

Proof of Theorem 2.1

Writing $V_n(m) := X_{n-m+1:n} - X_{n-2m+1:n}$, we have

$$\hat{\beta}_n(m,a) = \frac{1}{\log 2}\log\left(\frac{a(V_n(2[m/2])/V_n(m))2^{\hat{\beta}_n([m/2])} + 1}{aV_n(2[m/2])/V_n(m) + 2^{-\hat{\beta}_n(m)}}\right),$$

which converges in probability to β as $n \to \infty$ if we show that $V_n(m)/V_n(2[m/2]) \xrightarrow{p} 1$ as $n \to \infty$. For this it is enough to show that, for $m = 2k+1$ with $k = k(n) \to \infty$ and $k/n \to 0$ as $n \to \infty$,

$$V_n(m)/V_n(2[m/2]) = V_n(2k+1)/V_n(2k) \xrightarrow{p} 1 \text{ as } n \to \infty.$$

Let $\xi_1, \xi_2, \ldots$ be i.i.d. standard exponential random variables and let $\xi_{1:n} \leq \cdots \leq \xi_{n:n}$ be the order statistics of $\xi_1, \ldots, \xi_n$. Then $(X_{n-j+1:n})_{j=1}^n \overset{d}{=} (F^{-1}(1-e^{-\xi_{n-j+1:n}}))_{j=1}^n$, and further there exist i.i.d. standard exponential random variables $\tilde{\xi}_1, \ldots, \tilde{\xi}_n$ such that $(\xi_{n-j+1:n} - \xi_{n-j:n})_{j=1}^n \overset{d}{=} (\tilde{\xi}_j/j)_{j=1}^n$, where $\xi_{0:n} := 0$, which is usually referred to as Rényi's representation. Thus $k \to \infty$ implies that $\xi_{n-2k+1:n} - \xi_{n-2k:n} \xrightarrow{p} 0$

and $\xi_{n-4k+1:n} - \xi_{n-4k-1:n} \xrightarrow{p} 0$ as $n \to \infty$. Also note that $k/n \to 0$ implies that $e^{-\xi_{n-2k+1:n}} \to 0$ a.s. as $n \to \infty$. Therefore,

$$\begin{aligned}
\frac{V_n(2k+1)}{V_n(2k)} &\stackrel{d}{=} \frac{F^{-1}(1-e^{-\xi_{n-2k:n}}) - F^{-1}(1-e^{-\xi_{n-4k-1:n}})}{F^{-1}(1-e^{-\xi_{n-2k+1:n}}) - F^{-1}(1-e^{-\xi_{n-4k+1:n}})} \\
&= \frac{F^{-1}(1-e^{-\xi_{n-2k+1:n}} \cdot e^{\xi_{n-2k+1:n}-\xi_{n-2k:n}}) - F^{-1}(1-e^{-\xi_{n-2k+1:n}})}{F^{-1}(1-e^{-\xi_{n-2k+1:n}}) - F^{-1}(1-e^{-\xi_{n-2k+1:n}} \cdot e^{\xi_{n-2k+1:n}-\xi_{n-4k+1:n}})} \\
&\quad + \frac{F^{-1}(1-e^{-\xi_{n-2k+1:n}}) - F^{-1}(1-e^{-\xi_{n-2k+1:n}} \cdot e^{\xi_{n-2k+1:n}-\xi_{n-4k-1:n}})}{F^{-1}(1-e^{-\xi_{n-2k+1:n}}) - F^{-1}(1-e^{-\xi_{n-2k+1:n}} \cdot e^{\xi_{n-2k+1:n}-\xi_{n-4k+1:n}})} \\
&\xrightarrow{p} \frac{1-1}{1-2^{-\beta}} + \frac{1-2^{-\beta}}{1-2^{-\beta}} = 1 \text{ as } n \to \infty
\end{aligned}$$

by Lemma A.1 since $\xi_{n-2k+1:n} - \xi_{n-4k+1:n} \xrightarrow{p} \log 2$ as $n \to \infty$ (see Corollary 2.1 of Dekkers and de Haan (1989)). This completes the proof. □

Proof of Theorem 2.2

Let $\xi_1, \xi_2, ...$ be i.i.d. standard exponential random variables and let $\xi_{1:n} \leq \cdots \leq \xi_{n:n}$ be the order statistics of $\xi_1, ..., \xi_n$. Then the conditions on the sequence $m = m(n)$ imply that $\xi_{n-m+1:n} + \log(m/n) \to 0$ a.s. as $n \to \infty$ by Corollary 4 of Wellner (1978). Thus, for $m = 2k+1$ with $k = k(n) \to \infty$, $k/n \to 0$ and $k/\log\log n \to \infty$ as $n \to \infty$, we have $\xi_{n-2k+1:n} - \xi_{n-2k:n} \to 0$ a.s. and $\xi_{n-2k+1:n} - \xi_{n-4k+1:n} \to \log 2$ a.s. as $n \to \infty$. The rest of the proof is similar to that of Theorem 2.1. □

The following lemma, which is a reformulation of Theorem 2.2.4 of Falk, Hüsler and Reiss (1994) (see also Corollary 5.5.5 of Reiss (1989)), is crucial for providing a rate of convergence in the asymptotic normality of $\hat{\beta}_n(m, a)$.

Lemma A.2. *Suppose that $F \in Q(\delta; H_\beta)$ for some $\delta > 0$ and $\beta \in \Re$. Then there exist constants $a_n > 0$ and $b_n \in \Re$ such that, for any $k \in \{1, ..., n\}$ and $n \in \{1, 2, ...\}$,*

$$\begin{aligned}
&\sup_{B \in \mathcal{B}^k} \left| P\left\{((X_{n-j+1:n} - b_n)/a_n)_{j=1}^k \in B\right\} - P\left\{\left(\left(\left(\sum_{i=1}^{j} \xi_i\right)^{-\beta} - 1\right)/\beta\right)_{j=1}^k \in B\right\}\right| \\
&\qquad = O((k/n)^\delta \sqrt{k} + k/n),
\end{aligned}$$

where $\xi_1, \xi_2, ...$ are i.i.d. standard exponential random variables and $\mathcal{B}^k$ denotes the Borel σ-field in $\Re^k$. Here the constants a_n and b_n coincide with those of (1.1).

Proof of Lemma 2.1

By Lemma A.2, there exist $a_n > 0$ and $b_n \in \Re$ such that, for $m \in \{2, ..., n/4\}$ and $n \in \{8, 9, ...\}$,

$$\sup_{B\in\mathcal{B}^{4m}} \left| P\left\{((X_{n-j+1:n} - b_n)/a_n)_{j=1}^{4m} \in B\right\} - P\left\{\left(\left(\left(\sum_{i=1}^{j} \xi_i\right)^{-\beta} - 1\right)/\beta\right)_{j=1}^{4m} \in B\right\}\right|$$
$$= O((m/n)^\delta \sqrt{m} + m/n), \tag{A.1}$$

where $\xi_1, \xi_2, ...$ are i.i.d. standard exponential random variables. Thus, if we put

$$A_n^{(1)} := \frac{a(X_{n-[m/2]+1:n} - X_{n-2[m/2]+1:n}) + (X_{n-m+1:n} - X_{n-2m+1:n})}{a(X_{n-2[m/2]+1:n} - X_{n-4[m/2]+1:n}) + (X_{n-2m+1:n} - X_{n-4m+1:n})} - 2^\beta,$$

then within the error bound $O((m/n)^\delta \sqrt{m} + m/n)$ in variational distance $A_n^{(1)}$ behaves like

$$A_n^{(2)} := \frac{a\{(\sum_{i=1}^{[m/2]} \xi_i)^{-\beta} - (\sum_{i=1}^{2[m/2]} \xi_i)^{-\beta}\} + \{(\sum_{i=1}^{m} \xi_i)^{-\beta} - (\sum_{i=1}^{2m} \xi_i)^{-\beta}\}}{a\{(\sum_{i=1}^{2[m/2]} \xi_i)^{-\beta} - (\sum_{i=1}^{4[m/2]} \xi_i)^{-\beta}\} + \{(\sum_{i=1}^{2m} \xi_i)^{-\beta} - (\sum_{i=1}^{4m} \xi_i)^{-\beta}\}} - 2^\beta$$
$$= \frac{\begin{pmatrix} a\{(1 + [m/2]^{-1} \sum_{i=1}^{[m/2]} \eta_i)^{-\beta} - (2 + [m/2]^{-1} \sum_{i=1}^{2[m/2]} \eta_i)^{-\beta}\} \\ +\{(m/[m/2] + [m/2]^{-1} \sum_{i=1}^{m} \eta_i)^{-\beta} \\ -(2m/[m/2] + [m/2]^{-1} \sum_{i=1}^{2m} \eta_i)^{-\beta}\} \end{pmatrix}}{\begin{pmatrix} a\{(2 + [m/2]^{-1} \sum_{i=1}^{2[m/2]} \eta_i)^{-\beta} - (4 + [m/2]^{-1} \sum_{i=1}^{4[m/2]} \eta_i)^{-\beta}\} \\ +\{(2m/[m/2] + [m/2]^{-1} \sum_{i=1}^{2m} \eta_i)^{-\beta} \\ -(4m/[m/2] + [m/2]^{-1} \sum_{i=1}^{4m} \eta_i)^{-\beta}\} \end{pmatrix}} - 2^\beta,$$

where $\eta_i = \xi_i - 1$, $i = 1, 2, ...$ Note here that

$$\sup_{B\in\mathcal{B}} \left| P\left\{\frac{1}{\sqrt{k}} \sum_{i=1}^{k} \eta_i \in B\right\} - N(0,1)(B)\right| = O(1/\sqrt{k}), \ k \in \{1, 2, ...\}, \tag{A.2}$$

where $N(0, 1)$ stands for the standard normal distribution. Thus, using the simple fact that $[m/2]^{-1/2} = (m/2)^{-1/2} + O(m^{-3/2})$, $[m/2]^{-1} = (m/2)^{-1} + O(m^{-2})$, and $m/[m/2] = 2 + O(m^{-1})$, it can be seen that within the error bound $O(1/\sqrt{m})$ in variational distance $A_n^{(2)}$ again behaves like

$$A_n^{(3)} := \frac{\begin{pmatrix} a\Big\{\Big(1+\frac{Z_1}{\sqrt{m/2}}+O_P(m^{-3/2})\Big)^{-\beta} \\ \quad -\Big(2+\frac{Z_1+Z_2}{\sqrt{m/2}}+O_P(m^{-3/2})\Big)^{-\beta}\Big\} \\ +\Big\{\Big(2+\frac{Z_1+Z_2}{\sqrt{m/2}}+O_P(m^{-1})\Big)^{-\beta} \\ \quad -\Big(4+\frac{Z_1+Z_2+\sqrt{2}Z_3}{\sqrt{m/2}}+O_P(m^{-1})\Big)^{-\beta}\Big\} \end{pmatrix}}{\begin{pmatrix} a\Big\{\Big(2+\frac{Z_1+Z_2}{\sqrt{m/2}}+O_P(m^{-3/2})\Big)^{-\beta} \\ \quad -\Big(4+\frac{Z_1+Z_2+\sqrt{2}Z_3}{\sqrt{m/2}}+O_P(m^{-3/2})\Big)^{-\beta}\Big\} \\ +\Big\{\Big(4+\frac{Z_1+Z_2+\sqrt{2}Z_3}{\sqrt{m/2}}+O_P(m^{-1})\Big)^{-\beta} \\ \quad -\Big(8+\frac{Z_1+Z_2+\sqrt{2}Z_3+2Z_4}{\sqrt{m/2}}+O_P(m^{-1})\Big)^{-\beta}\Big\} \end{pmatrix}} - 2^{\beta}$$

$$= \frac{\begin{pmatrix} a\Big\{\Big(1-\beta\frac{Z_1}{\sqrt{m/2}}\Big)-2^{-\beta}\Big(1-\beta\frac{Z_1+Z_2}{2\sqrt{m/2}}\Big)\Big\} \\ +\Big\{2^{-\beta}\Big(1-\beta\frac{Z_1+Z_2}{2\sqrt{m/2}}\Big)-4^{-\beta}\Big(1-\beta\frac{Z_1+Z_2+\sqrt{2}Z_3}{4\sqrt{m/2}}\Big)\Big\}+O_P(m^{-1}) \end{pmatrix}}{\begin{pmatrix} a\Big\{2^{-\beta}\Big(1-\beta\frac{Z_1+Z_2}{2\sqrt{m/2}}\Big)-4^{-\beta}\Big(1-\beta\frac{Z_1+Z_2+\sqrt{2}Z_3}{4\sqrt{m/2}}\Big)\Big\} \\ +\Big\{4^{-\beta}\Big(1-\beta\frac{Z_1+Z_2+\sqrt{2}Z_3}{4\sqrt{m/2}}\Big)-8^{-\beta}\Big(1-\beta\frac{Z_1+Z_2+\sqrt{2}Z_3+2Z_4}{8\sqrt{m/2}}\Big)\Big\} \\ +O_P(m^{-1}) \end{pmatrix}} - 2^{\beta}$$

$$= \frac{\beta\begin{pmatrix} 4a(-Z_1+Z_2)+2^{1-\beta}(a-1)(Z_1+Z_2-\sqrt{2}Z_3) \\ +4^{-\beta}(Z_1+Z_2+\sqrt{2}Z_3-2Z_4) \end{pmatrix}}{2^{2-\beta}(1-2^{-\beta})(a+2^{-\beta})\sqrt{2m}} + O_P(1/m), \qquad \text{(A.3)}$$

where Z_1, Z_2, Z_3, Z_4 are independent standard normal random variables and the second equality follows from the Taylor expansion $(1+x)^{-\beta} = 1-\beta x + O(x^2)$ as $x \to 0$. Using the Taylor expansion $\log(1+x) = x + O(x^2)$ as $x \to 0$, we therefore in all obtain that within the error bound $O((m/n)^{\delta}\sqrt{m}+m/n+1/\sqrt{m})$ in variational distance

$$\sqrt{m}(\hat{\beta}_n(m,a)-\beta) = \frac{\sqrt{m}}{\log 2}\log\Big(1+\frac{A_n^{(1)}}{2^{\beta}}\Big)$$

behaves like

$$
\begin{aligned}
&\frac{\sqrt{m}}{\log 2}\log\left(1+\frac{A_n^{(3)}}{2^\beta}\right)\\
&= \frac{\sqrt{m}}{\log 2}\Big(\frac{A_n^{(3)}}{2^\beta}+O_P((A_n^{(3)})^2)\Big)\\
&= \frac{\sqrt{m}}{2^\beta\log 2}A_n^{(3)}+O_P(1/\sqrt{m})\\
&= \frac{\beta\begin{pmatrix}(2^{-\beta-1}-1)(a+2^{-\beta-1})Z_1\\ +\{(2^{-\beta-1}+1)a+2^{-\beta-1}(2^{-\beta-1}-1)\}Z_2\\ -2^{-\beta-1/2}(a-1-2^{-\beta-1})Z_3-2^{-2\beta-1}Z_4\end{pmatrix}}{(1-2^{-\beta})(a+2^{-\beta})\sqrt{2}\log 2}+O_P(1/\sqrt{m})
\end{aligned}
$$

since $A_n^{(3)}$ is of order $O_P(1/\sqrt{m})$ from (A.3). The assertion now follows from elementary computations. □

Proof of Lemma 2.2

Put $g(x) := \log x/\log 2$ and

$$A_n(\beta) := \frac{a^*(\beta)(X_{n-[m/2]+1:n}-X_{n-2[m/2]+1:n})+(X_{n-m+1:n}-X_{n-2m+1:n})}{a^*(\beta)(X_{n-2[m/2]+1:n}-X_{n-4[m/2]+1:n})+(X_{n-2m+1:n}-X_{n-4m+1:n})}.$$

Then, since by the Taylor expansion of g

$$
\begin{aligned}
&\hat\beta_n(m,a^*(\tilde\beta_n))-\hat\beta_n(m,a^*(\beta))\\
&= g(A_n(\tilde\beta_n))-g(A_n(\beta))\\
&= g'(A_n(\beta))(A_n(\tilde\beta_n)-A_n(\beta))+\frac{g''(A_n(\beta))}{2!}(A_n(\tilde\beta_n)-A_n(\beta))^2+\cdots\\
&= \frac{A_n(\tilde\beta_n)-A_n(\beta)}{\log 2}\left\{\frac{1}{A_n(\beta)}-\frac{A_n(\tilde\beta_n)-A_n(\beta)}{2(A_n(\beta))^2}+\cdots\right\}
\end{aligned}
$$

and since by (2.1) $A_n(\beta)\xrightarrow{p}2^\beta$ as $n\to\infty$, it sufficies to show that

$$\sqrt{m}(A_n(\tilde\beta_n)-A_n(\beta))=o_P(1) \text{ as } n\to\infty. \tag{A.4}$$

Now

$$\sqrt{m}(A_n(\tilde\beta_n)-A_n(\beta))=(a^*(\tilde\beta_n)-a^*(\beta))\sqrt{m}R_n, \tag{A.5}$$

where

$$R_n := \frac{\left(\begin{array}{l} (X_{n-[m/2]+1:n} - X_{n-2[m/2]+1:n})(X_{n-2m+1:n} - X_{n-4m+1:n}) \\ -(X_{n-2[m/2]+1:n} - X_{n-4[m/2]+1:n})(X_{n-m+1:n} - X_{n-2m+1:n}) \end{array}\right)}{\left(\begin{array}{l} \{a^*(\tilde{\beta}_n)(X_{n-2[m/2]+1:n} - X_{n-4[m/2]+1:n}) + (X_{n-2m+1:n} - X_{n-4m+1:n})\} \\ \times\{a^*(\beta)(X_{n-2[m/2]+1:n} - X_{n-4[m/2]+1:n}) + (X_{n-2m+1:n} - X_{n-4m+1:n})\} \end{array}\right)}.$$

Then, by (A.1), within the error bound $O((m/n)^\delta\sqrt{m} + m/n)$ in variational distance R_n behaves like $R_n^{(1)} :=$

$$\frac{\left(\begin{array}{l} \{(1+[m/2]^{-1}\sum_{i=1}^{[m/2]}\eta_i)^{-\beta} - (2+[m/2]^{-1}\sum_{i=1}^{2[m/2]}\eta_i)^{-\beta}\} \\ \times\{(2m/[m/2]+[m/2]^{-1}\sum_{i=1}^{2m}\eta_i)^{-\beta} - (4m/[m/2]+[m/2]^{-1}\sum_{i=1}^{4m}\eta_i)^{-\beta}\} \\ -\{(2+[m/2]^{-1}\sum_{i=1}^{2[m/2]}\eta_i)^{-\beta} - (4+[m/2]^{-1}\sum_{i=1}^{4[m/2]}\eta_i)^{-\beta}\} \\ \times\{(m/[m/2]+[m/2]^{-1}\sum_{i=1}^{m}\eta_i)^{-\beta} - (2m/[m/2]+[m/2]^{-1}\sum_{i=1}^{2m}\eta_i)^{-\beta}\} \end{array}\right)}{\left(\begin{array}{l} \{a^*(\tilde{\beta}_n)((2+[m/2]^{-1}\sum_{i=1}^{2[m/2]}\eta_i)^{-\beta} - (4+[m/2]^{-1}\sum_{i=1}^{4[m/2]}\eta_i)^{-\beta}) \\ +((2m/[m/2]+[m/2]^{-1}\sum_{i=1}^{2m}\eta_i)^{-\beta} - (4m/[m/2]+[m/2]^{-1}\sum_{i=1}^{4m}\eta_i)^{-\beta})\} \\ \times\{a^*(\beta)((2+[m/2]^{-1}\sum_{i=1}^{2[m/2]}\eta_i)^{-\beta} - (4+[m/2]^{-1}\sum_{i=1}^{4[m/2]}\eta_i)^{-\beta}) \\ +((2m/[m/2]+[m/2]^{-1}\sum_{i=1}^{2m}\eta_i)^{-\beta} - (4m/[m/2]+[m/2]^{-1}\sum_{i=1}^{4m}\eta_i)^{-\beta})\} \end{array}\right)},$$

where $\eta_1 + 1, \eta_2 + 1, \ldots$ are i.i.d. standard exponential random variables. Again, similarly as in the proof of Lemma 2.1, by applying (A.2) and then by using the Taylor expansion $(1+x)^{-\beta} = 1 - \beta x + O(x^2)$ as $x \to 0$, it can be seen that within the error bound $O(1/\sqrt{m})$ in variational distance $R_n^{(1)}$ behaves like

$$\frac{\beta\{(2^{-\beta}-2)Z_1 + (2^{-\beta}+6)Z_2 - (3\cdot 2^{-\beta}+2)\sqrt{2}Z_3 + 2^{-\beta+1}Z_4\}}{4(1-2^{-\beta})(a^*(\tilde{\beta}_n)+2^{-\beta})(a^*(\beta)+2^{-\beta})\sqrt{2m}} + O_P(1/m),$$

where Z_1, Z_2, Z_3, Z_4 are independent standard normal random variables. Thus in all, within the error bound $O((m/n)^\delta\sqrt{m} + m/n + 1/\sqrt{m})$ in variational distance $\sqrt{m}R_n$ behaves like

$$\frac{\beta Y}{2(1-2^{-\beta})(a^*(\tilde{\beta}_n)+2^{-\beta})(a^*(\beta)+2^{-\beta})} + O_P(1/\sqrt{m}),$$

where Y is a normal random variable with mean 0 and variance $3\cdot 2^{-2\beta} + 4\cdot 2^{-\beta} + 6$. Since $a^*(\tilde{\beta}_n) \xrightarrow{p} a^*(\beta)$ as $n \to \infty$, (A.5) therefore implies (A.4). This completes the proof. □

REFERENCES

de Haan, L. (1984). "Slow variation and characterization of domains of attraction," in *Statistical Extremes and Applications*, ed. J. Tiago de Oliveira, Reidel, Dordrecht, pp. 31-48.

Dekkers, A. L. M. and de Haan, L. (1989). "On the estimation of the extreme-value index and large quantile estimation," *Ann. Statist.*, **17**, 1795-1832.

Drees, H. (1995). "Refined Pickands estimators of the extreme value index," *Ann. Statist.*, **23**, 2059-2080.

Falk, M. (1986). "Rates of uniform convergence of extreme order statistics," *Ann. Inst. Statist. Math.*, **38**, 245-262.

Falk, M. (1994). "Efficiency of convex combinations of Pickands estimator of the extreme value index," *J. Nonparametric Statist.*, **4**, 133-147.

Falk, M., Hüsler, J. and Reiss, R.-D. (1994). *Laws of Small Numbers: Extremes and Rare Events*, Birkhäuser, Basel.

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

Pickands, J. (1975). "Statistical inference using extreme order statistics," *Ann. Statist.*, **3**, 119-131.

Reiss, R.-D. (1989). *Approximate Distributions of Order Statistics*, Springer, New York.

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

Wellner, J. A. (1978). "Limit theorems for the ratio of the empirical distribution function to the true distribution function," *Z. Wahrsch. verw. Gebiete*, **45**, 73-88.