Centrum voor Wiskunde en Informatica
Centre for Mathematics and Computer Science

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

Large quantile estimation under extreme-value conditions

Department of Mathematical Statistics Report MS-R8711 September

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 Pure Research (Z.W.O.).

Large Quantile Estimation under Extreme-Value Conditions

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

L. de Haan
Erasmus University, Rotterdam, the Netherlands

A large quantile is estimated by a combination of extreme or intermediate order statistics. This leads to an asymptotic confidence interval.

1980 Mathematics Subject Classification: Primary 62F25, Secondary 62G30.
Key Words & Phrases: quantile estimation, extreme-value theory.

1. Introduction

After the 1953 flood the Dutch government set the following standard for the sea-dikes in the Netherlands: the probability that at any time in a given year the sea-level exceeds the level of the dikes is 1:10,000. The question how to give an estimate for such a level from past observations involves estimation of large quantiles of an unknown distribution function.

We consider the following idealized model: n i.i.d. observations $X_1, X_2, \ldots, X_n$ are available from an unknown distribution function F. In a future year k i.i.d. observations $Y_1, Y_2, \ldots, Y_k$ will be taken from F. We want to find a level x_p (where p is a given number <<1) such that $P\{\max(Y_1, \ldots, Y_k) \leq x_p\} = 1-p$, i.e.

Report MS-R8711
Centre for Mathematics and Computer Science
P.O. Box 4079, 1009 AB Amsterdam, The Netherlands

(1.1) $$F^k(x_p) = 1-p.$$

Define

$$U(x) := \left(\frac{1}{1-F}\right)^{\leftarrow}(x)$$

(the arrow means: inverse function) and $a_k(p) := k^{-1}p/\{1 - (1-p)^{1/k}\}$. Note that (loosely speaking)

(1.2) $$x_p = U(a_k(p) \cdot \frac{k}{p})$$

and

(1.3) $$a_k(p) \to 1 \quad \text{for } p \to 0,\ k \text{ fixed.}$$

We want to estimate x_p on the basis of the observations $X_1, X_2, \ldots, X_n$. Let $X_{(1)}^{(n)} \leq X_{(2)}^{(n)} \leq \ldots \leq X_{(n)}^{(n)}$ be their order statistics, F_n the empirical distribution function and

(1.4) $$U_n := \left(\frac{1}{1 - F_n}\right)^{\leftarrow}.$$

Note that $F_n(X_{(r)}^{(n)}) = r/n$ $(r = 1, 2, \ldots, n)$ so that

(1.5) $$X_{(n-r)}^{(n)} = U_n(\frac{n}{r}) \quad \text{for } r = 1, 2, \ldots, n.$$

In case $p < 1/n$ nothing can be done without imposing extra conditions on F. We choose for an asymptotic theory and for imposing the extra condition that F is in the domain of attraction of some extreme-value distribution. We use the von Mises-parametrization of the extreme value distributions ($\gamma \in \mathbb{R}$)

(1.6) $$G_\gamma(x) := \exp -(1 + \gamma x)^{-1/\gamma}$$

for values of x such that $1 + \gamma x > 0$ (for $\gamma = 0$ the right-hand side must be interpreted as $\exp - \exp - x$). Now $F \in D(G_\gamma)$ if and only if for $x > 0$, $y > 0$ $(y \neq 1)$

(1.7) $$\lim_{t\to\infty} \frac{U(tx) - U(t)}{U(ty) - U(t)} = \frac{x^\gamma - 1}{y^\gamma - 1}$$

(for $\gamma = 0$ the right-hand side must be interpreted as logx/logy).
We shall give asymptotic results for $n\to\infty$ and $p = p_n \to 0$ (the latter assumption is reasonable at least for the specific problem mentioned above, since there $p \ll 1/n$). For the unknown function U we write the identity ($r = 1, 2, \ldots$)

$$(1.8) \qquad x_p = U(a_k(p)\cdot\frac{k}{p}) = \frac{U(a_k(p)\cdot\frac{k}{p}) - U(\frac{n}{r})}{U(\frac{n}{r}) - U(\frac{n}{2r})}\{U(\frac{n}{r}) - U(\frac{n}{2r})\} + U(\frac{n}{r}).$$

Upon replacing U by its empirical counterpart U_n in appropriate places and using (1.7) and (1.3) we arrive at the following proposed estimate for x_p:

$$(1.9) \qquad \hat{x}_{p,n} := \frac{(\frac{kr}{pn})^{\hat{\gamma}_n} - 1}{1 - 2^{-\hat{\gamma}_n}}\{X^{(n)}_{(n-r)} - X^{(n)}_{(n-2r)}\} + X^{(n)}_{(n-r)}$$

where $\hat{\gamma}_n$ is a consistent estimate of γ (see J. Pickands III (1975) and Dekkers and de Haan (1987)). Intuitively this means that in the absence of more observations (that would have allowed us to simply use the inverse empirical distribution function) one uses observed spacing to make up (modulo a multiplicative constant) for the missing spacings; like a surgeon who uses a piece of skin from elsewhere to cover a wound.
Note that we do not use the largest observation explicitly. One can argue that this makes sense because the largest observation may add too much uncertainty (larger variance if applicable).
In order to deal with the asymptotics we need to require that $kr/(pn)$ has a positive limit ($n\to\infty$). We shall consider two cases: r fixed, hence $p_n \sim c/n$ ($c < 1$) - section 2 - and $r\to\infty$, $r(n)/n \to 0$ hence $p_n \sim c.r(n)/n \to 0$ and $np_n \sim c.r(n) \to \infty$ (section 3). In either case we give an asymptotic confidence interval for x_p. In order to emphasize the difference between the situations in the two sections, in section 3 we shall use the letter m instead of the letter r for the number of order statistics involved.
Somewhat related papers are Weissman (1978) and Boos (1984).

2. Results for a bounded number of order statistics

We shall prove the following result providing an asymptotic confidence interval for x_{p_n}.

Theorem 2.1

Suppose the extreme-value condition (1.7) holds. If k and r are fixed, $n\to\infty$, $p = p_n$ and $\lim_{n\to\infty} np_n = c \in (0, 1)$, then

$$\frac{\hat{x}_{p_n,n} - x_{p_n}}{X^{(n)}_{(n-r)} - X^{(n)}_{(n-2r)}} \tag{2.1}$$

converges in distribution to the distribution of the random variable

$$\frac{(\frac{kr}{c})^{\gamma} - 2^{-\gamma}}{1 - 2^{-\gamma}} + \{1 - (\frac{kQ_r}{c})^{\gamma}\}/\{e^{\gamma H_r} - 1\}, \tag{2.2}$$

where H_r and Q_r are independent, Q_r has a gamma (2r+1)-distribution and H_r the distribution of $\sum_{j=r+1}^{2r} Z_j/j$ with $Z_1, Z_2, \ldots$ i.i.d. standard exponential.

For the proof we need some Lemma's.

Lemma 2.2

Let $E^{(n)}_{(1)} \leq E^{(n)}_{(2)} \leq \ldots \leq E^{(n)}_{(n)}$ be standard exponential order statistics. Then for $n\to\infty$ and r fixed the random vector

$$(E^{(n)}_{(n-r)} - E^{(n)}_{(n-2r)}, E^{(n)}_{(n-2r)} - \log n)$$

converges in distribution to the distribution of $(H_r, -\log Q_r)$ where H_r and Q_r are independent and have the probability distributions mentioned in theorem 2.1.

Proof

We use the so-called Rényi-representation: $\{E_{(m)} - E_{(m+1)}\}_{m=1}^{n-1} \stackrel{d}{=} \{Z_m/m\}_{m=1}^{n-1}$ with $Z_1, Z_2, \ldots$ (depending on n) i.i.d. standard exponential. The independence is immediate. Note that the distribution of $E^{(n)}_{(n-r)} - E^{(n)}_{(n-2r)}$ does not depend on n. For the asymptotic distribution of $E^{(n)}_{(n-2r)} - \log n$ see Smirnov (1949).

Lemma 2.3

Under the conditions of theorem 2.1 the random vector

$$\left(\frac{X^{(n)}_{(n-r)} - X^{(n)}_{(n-2r)}}{a(n)}, \frac{X^{(n)}_{(n-2r)} - U(n)}{a(n)}\right) \tag{2.3}$$

where $a(t) := \{U(2t) - U(t)\}\, \gamma/(2^{\gamma} - 1)$, converges in distribution to the distribution of

$$(2.4) \qquad \left(Q_r^{\gamma}\, \frac{e^{\gamma H_r} - 1}{\gamma},\ \frac{Q_r^{-\gamma} - 1}{\gamma}\right)$$

where H_r and Q_r are independent and have the probability distributions mentioned in theorem 2.1.

Proof

We may write the order statistics as functions of the exponential order statistics.

$$\left(\frac{X^{(n)}_{(n-r)} - X^{(n)}_{(n-2r)}}{a(n)},\ \frac{X^{(n)}_{(n-2r)} - U(n)}{a(n)}\right) \stackrel{d}{=}$$

$$\left(\frac{a\left(e^{E^{(n)}_{(n-2r)} - \log n} \cdot n\right)}{a(n)}\ \frac{U\left(e^{E^{(n)}_{(n-r)} - E^{(n)}_{(n-2r)}} \cdot e^{E^{(n)}_{(n-2r)}}\right) - U\left(e^{E^{(n)}_{(n-2r)}}\right)}{a\left(e^{E^{(n)}_{(n-2r)}}\right)},\right.$$

$$\left.\frac{U\left(e^{E^{(n)}_{(n-2r)} - \log n} \cdot n\right) - U(n)}{a(n)}\right) \to \left(Q_r^{-\gamma} \cdot \frac{e^{\gamma H_r} - 1}{\gamma},\ \frac{Q_r^{-\gamma} - 1}{\gamma}\right) \quad (n \to \infty)$$

in distribution by Lemma 2.2 and (1.7), since the latter relation holds locally uniformly and implies $a(tx) \sim x^{\gamma} a(t)$ $(t \to \infty)$ for $x > 0$, locally uniformly.

Proof of theorem 2.1

$$(\hat{x}_{p_n,n} - x_{p_n})/(X^{(n)}_{(n-r)} - X^{(n)}_{(n-2r)}) =$$

$$= \frac{\left(\frac{kr}{pn}\right)^{\hat{\gamma}_n} - 1}{1 - 2^{-\hat{\gamma}_n}} + 1 + \frac{a(n)}{X^{(n)}_{(n-r)} - X^{(n)}_{(n-2r)}} \left\{\frac{X^{(n)}_{(n-2r)} - U(n)}{a(n)} + \right.$$

$$+\frac{U(n)-U(a_k(p)\frac{k}{p.n}\cdot n)}{a(n)}\}$$

$$\to \frac{(\frac{kr}{c})^{\gamma}-2^{-\gamma}}{1-2^{-\gamma}}+(Q_r^{-\gamma}\cdot\frac{e^{\gamma H_r}-1}{\gamma})^{-1}\cdot\{\frac{Q_r^{-\gamma}-1}{\gamma}+\frac{1-(\frac{k}{c})^{\gamma}}{\gamma}\}$$

$(n\to\infty)$ in distribution by (1.7) and Lemma 2.2.

Remark 1

It should be possible, and maybe even preferable to use different linear combinations of spacings instead of $X_{(n-r)}^{(n)} - X_{(n-2r)}^{(n)}$ in the definition of $\hat{x}_{p,n}$. This is a subject of later research.

Remark 2

We do not enter here into the queston of how to choose r in an optimal way.

Remark 3

Theorem 2.1 only implies consistency of the estimate $\hat{x}_{p_n,n}$ for x_{p_n} in the trivial case $\gamma < 0$. It is doubtful if in general a consistent estimate is possible within the present set-up. By exploiting the fact that $\sqrt{2r}\,(H_r - \log 2)$ and $\{Q_r - (2r+1)\}/\sqrt{2r+1}$ are asymptotically standard normal for $r\to\infty$, one can show however that for $\gamma < \frac{1}{2}$ the expression (2.2) converges to zero, but for $\gamma = \frac{1}{2}$ it converges to a normal distribution and for $\gamma > \frac{1}{2}$ it diverges $(n\to\infty)$.

3. Results for an unbounded number of order statistics

We shall require the following strengthening of the extreme-value condition (1.7).

Condition 3.1

Suppose that the distribution function F has a positive derivative $F'(x)$ for all $x < U(\infty)$. Suppose with either choice of sign one of the following conditions holds:

i) $\pm\, t^{1+1/\gamma}F'(t) \in \Pi$ for some $\gamma > 0$,

ij) $U(\infty) < \infty$ and $\mp\, t^{-1-1/\gamma}F'(U(\infty) - t^{-1}) \in \Pi$ for some $\gamma < 0$,

iij) Let $f_0 := (1-F)/F'$. There exists a function $\pm\,\alpha: \mathbb{R} \to \mathbb{R}^+$ with $\alpha(t) \to 0$ $(t\to\infty)$ such that for $x > 0$

$$\lim_{t\to\infty}\left(\frac{1-F(t+x\,f_0(t))}{1-F(t)} - e^{-x}\right)/\alpha(t) = \frac{x^2}{2}\,e^{-x},$$

(this is the case $\gamma = 0$).

Lemma 3.2

Condition 3.1 implies $\pm\, t^{1-\gamma}U'(t) \in \Pi$ and hence in particular (1.7) for the corresponding value of γ (Dekkers and de Haan 1987).
The reader is also referred to that paper for examples and additional information.

Theorem 3.3

Suppose condition 3.1 holds. Let $m = m(n) \in \mathbb{N}$, $m(n) \to \infty$, $m(n)/n \to 0$ $(n\to\infty)$ and $p = p_n \sim c.m(n)/n$ $(n\to\infty)$ with $c > 0$. Define

$$(3.1)\qquad \hat{x}_{p,n} := \frac{\left(a_k(p)\,\frac{km}{p_n n}\right)^{\hat{\gamma}_n} - 1}{1 - 2^{-\hat{\gamma}_n}}\,\{X^{(n)}_{(n-m)} - X^{(n)}_{(n-2m)}\} + X^{(n)}_{(n-m)}$$

where $\hat{\gamma}_n := (\log 2)^{-1}\log\dfrac{X^{(n)}_{(n-m)} - X^{(n)}_{(n-2m)}}{X^{(n)}_{(n-2m)} - X^{(n)}_{(n-4m)}}$ (Note that this definition differs slightly from (1.9)). Then

$$(3.2)\qquad \sqrt{2m}\,.\,\frac{\hat{x}_{p,n} - x_p}{X^{(n)}_{(n-m)} - X^{(n)}_{(n-2m)}}$$

is asymptotically normal $(n\to\infty)$ with mean zero and variance

$$(D_1 + D_2)^2 2^{2\gamma} + \frac{\left(\frac{2k}{c}\right)^{2\gamma}\gamma^2}{(2^\gamma - 1)^2\,.\,2} + \tfrac{1}{2}\left(\frac{\left(\frac{2k}{c}\right)^\gamma\gamma}{(2^\gamma - 1)} - D_1\right)^2$$

where D_1 and D_2 are defined by

$$D_1 := \{\log(\frac{2k}{c}) \cdot (\frac{2k}{c})^\gamma(2^\gamma - 1) - ((\frac{2k}{c})^\gamma - 1)\log(2)2^\gamma\}/(2^\gamma - 1)^2$$

$$D_2 := \{(\frac{2k}{c})^\gamma - 1\} \cdot \gamma/(2^\gamma - 1)^2.$$

Remark

Let us compare (3.1) with an obvious way to estimate x_p. Using $F^n(a_n x + b_n) \approx G_\gamma(x)$ one gets from (1.1)

$$(1 - p)^{1/k} = F(x_p) \approx G_\gamma^{1/n}(\frac{x_p - b_n}{a_n}),$$

i.e.

$$x_p \approx b_n + a_n . G_\gamma^{\leftarrow}(1 - p)^{n/k}) \approx b_n + a_n \cdot \frac{(\frac{k}{np})^\gamma - 1}{\gamma}.$$

We can identify approximately b_n with $X^{(n)}_{(n-m)}$, a_n with $\{X^{(n)}_{(n-m)} - X^{(n)}_{(n-2m)}\}$. $\hat{\gamma}_n/\{1 - 2^{-\hat{\gamma}_n}\}$ and $\{(\frac{k}{np})^\gamma - 1\}/\gamma$ with $[\{a_k(p)k.m/(np_n)\}^{\hat{\gamma}_n} - 1]/\hat{\gamma}_n$ figuring in (3.1). Note that $\{X^{(n)}_{(n-m)} - X^{(n)}_{(n-2m)}\}\hat{\gamma}_n/\{1 - 2^{-\hat{\gamma}_n}\}$ is exactly the estimate of the scale parameter recommended by J. Pickands III (1975). Note also that our m is missing in the formula for $G_\gamma^{\leftarrow}((1-p)^{n/k})$; the factor m comes in because we use $X^{(n)}_{(n-m)}$ rather than $X^{(n)}_{(n)}$ to fix the location.

For the proof we need a series of lemma's. For each of them we shall suppose that condition 3.1 is in force and the condition $m(n) \to \infty$, $m(n)/n \to 0$, although sometimes a weaker condition suffices.

Lemma 3.4

Let $E^{(n)}_{(1)} \leq E^{(n)}_{(2)} \leq \ldots \leq E^{(n)}_{(n)}$ be standard exponential order statistics. Then

(3.3) $\sqrt{2m}\,\{E^{(n)}_{(n-2m)} - \log\frac{n}{2m}\}$ and $\sqrt{2m}\,\{E^{(n)}_{(n-m)} - E^{(n)}_{(n-2m)} - \log 2\}$

are independent and have asymptotically a standard normal distribution.

Proof

Smirnov (1949) and Dekkers and de Haan (1987) respectively.

Corollary 3.5

$E^{(n)}_{(n-m)} - \log \frac{n}{m} \to 0$ in probability.

Lemma 3.6

Condition 3.1 implies

$$(3.4) \qquad \lim_{n\to\infty} \sqrt{m} \left\{ \frac{V(\log \frac{n}{m} + x) - V(\log \frac{n}{m})}{V'(\log \frac{n}{m})} - \frac{e^{\gamma x} - 1}{\gamma} \right\} = 0$$

locally uniformly for sequences $m(n) \to \infty$ satisfying $m(n) = o(n/g^{+}(n))$ $(n\to\infty)$ where

$$(3.5) \qquad g(t) := \tfrac{1}{2}\, t^{3-2\gamma} \{U'(t)\ (2^{\gamma} - 1)/\lambda(U(2t) - U(t))\}^2$$

and

$$(3.6) \qquad V(x) := U(e^{x}).$$

Proof

The expression at the left-hand side of (3.4) equals

$$\lim_{n\to\infty} \int_0^x \sqrt{m}\, \frac{V'(\log \frac{n}{m} + s) - e^{\gamma s} V'(\log \frac{n}{m})}{V'(\log \frac{n}{m})}\, ds.$$

Now

$$(3.7) \qquad \lim_{n\to\infty} \sqrt{m}\, \frac{V'(\log \frac{n}{m} + s) - e^{\gamma s} V'(\log \frac{n}{m})}{V'(\log \frac{n}{m})} = 0 \text{ for all } s$$

by the reasoning in the proof of theorem 2.3 of Dekkers and de Haan (1987).

Lemma 3.7

The random vector

$$(3.8) \qquad \sqrt{2m}\left(\frac{V(E^{(n)}_{(n-2m)}) - V(\log \frac{n}{2m})}{V'(\log \frac{n}{2m})},\ \frac{V(E^{(n)}_{(n-m)}) - V(E^{(n)}_{(n-2m)})}{2^{\gamma} V'(E^{(n)}_{(n-2m)})} - \frac{1 - 2^{-\gamma}}{\gamma} \right)$$

is asymptotically standard normal.

Proof

$$(3.9)\qquad \sqrt{2m}\,\frac{V(E^{(n)}_{(n-2m)}) - V(\log \frac{n}{2m})}{V'(\log \frac{n}{2m})} = \sqrt{2m}\int_0^{N_n/\sqrt{2m}} \frac{V'(\log \frac{n}{2m} + s)}{V'(\log \frac{n}{2m})}ds$$

with

$$(3.10)\qquad N_n := \sqrt{2m}\,\{E^{(n)}_{(n-2m)} - \log(\tfrac{n}{2m})\}$$

and

$$(3.11)\qquad \sqrt{2m}\{\frac{V(E^{(n)}_{(n-m)}) - V(E^{(n)}_{(n-2m)})}{2^{\gamma}V'(\,E^{(n)}_{(n-2m)})} - \frac{1 - 2^{-\gamma}}{\gamma}\} =$$

$$2^{-\gamma}.\sqrt{2m}\int_0^{\log 2}\{\frac{V'(E^{(n)}_{(n-2m)} + s)}{V'(E^{(n)}_{(n-2m)})} - e^{\gamma s}\}ds +$$

$$2^{-\gamma}\sqrt{2m}\int_0^{M_n/\sqrt{2m}}\frac{V'(E^{(n)}_{(n-2m)}) + \log 2 + s)}{V'(E^{(n)}_{(n-2m)})}ds$$

with

$$(3.12)\qquad M_n := \sqrt{2m}\,\{E^{(n)}_{(n-m)} - E^{(n)}_{(n-2m)} - \log 2\}.$$

By (3.7) and Corollary 3.5 the first term at the right-hand side of (3.11) tends to zero in probability. The joint distribution of the right-hand side of (3.9) and the second term at the right-hand side of (3.11) converges to a standard two-dimensional normal distribution by Lemma 3.4 and (from condition 3.1)

$$(3.13)\qquad \lim_{t\to\infty}\frac{V'(t+s)}{V'(t)} = e^{\gamma s} \quad \text{for all } s.$$

Corollary 3.8

$$\lim_{n\to\infty}\frac{V(E^{(n)}_{(n-m)}) - V(E^{(n)}_{(n-2m)})}{V'(\log \frac{n}{2m})} = \frac{2^{\gamma} - 1}{\gamma} \text{ in probability.}$$

Lemma 3.9

The random variable

$$\sqrt{2m}\ \{\frac{V'(E^{(n)}_{(n-2m)})}{V'(\log \frac{n}{2m})} - 1\}$$

is asymptotically normal with mean zero and variance γ^2 $(n\to\infty)$. The asymptotic correlation with the first component of the random vector from Lemma 3.7 is one.

Proof

$$\sqrt{2m}\ \{\frac{V'(E^{(n)}_{(n-2m)})}{V'(\log \frac{n}{2m})} - 1\} =$$

$$= \sqrt{2m}\ \frac{V'(\log \frac{n}{2m} + (E^{(n)}_{(n-2m)} - \log \frac{n}{2m})) - e^{\gamma(E^{(n)}_{(n-2m)} - \log \frac{n}{2m})}\, V'(\log \frac{n}{2m})}{V'(\log \frac{n}{2m})} +$$

$$+ \sqrt{2m}\ \{e^{\gamma(E^{(n)}_{(n-2m)} - \log \frac{n}{2m})} - 1\}.$$

The first term tends to zero by (3.7) (holding locally uniformly) and the second term is asymptotically $\gamma\sqrt{2m}\ \{E^{(n)}_{(n-2m)} - \log \frac{n}{2m}\}$. Use Lemma 3.4. ◊

Corollary 3.10

$\lim_{n\to\infty} V'(E^{(n)}_{(n-2m)})/V'(\log n/2m) = 1$ in probability.

Proof of theorem 3.1

$$(3.14)\qquad \sqrt{2m}\ \frac{\hat{x}_{p,n} - x_p}{V(E^{(n)}_{(n-m)}) - V(E^{(n)}_{(n-2m)})} =$$

$$\sqrt{2m}\ \{\frac{(a_k(p)\ \frac{km}{pn})^{\hat{\gamma}_n} - 1}{1 - 2^{-\hat{\gamma}_n}} + 1 + \frac{V(E^{(n)}_{(n-2m)}) - V(\log \frac{n}{2m})}{V(E^{(n)}_{(n-m)}) - V(E^{(n)}_{(n-2m)})} -$$

$$- \frac{V(\log \frac{n}{2m} + \log a_k(p) \frac{2km}{pn}) - V(\log \frac{n}{2m})}{V(E^{(n)}_{(n-m)}) - V(E^{(n)}_{(n-2m)})}\}$$

$$= \sqrt{2m}\ \{\frac{(a_k(p) \frac{km}{pn})^{\hat{\gamma}_n} - 2^{-\hat{\gamma}_n}}{1 - 2^{-\hat{\gamma}_n}} - \frac{(a_k(p) \frac{2km}{pn})^{\gamma} - 1}{2^{\gamma} - 1}\}$$

$$+ \sqrt{2m}\ \frac{V(E^{(n)}_{(n-2m)}) - V(\log \frac{n}{2m})}{V'(\log \frac{n}{2m})} \cdot \frac{V'(\log \frac{n}{2m})}{V(E^{(n)}_{(n-m)}) - V(E^{(n)}_{(n-2m)})} +$$

$$- \sqrt{2m}\ \{\frac{V(\log \frac{n}{2m} + \log a_k(p) \frac{2km}{pn}) - V(\log \frac{n}{2m})}{V'(\log \frac{n}{2m})} - \frac{(a_k(p) \frac{2km}{pn})^{\gamma} - 1}{\gamma}\}$$

$$\frac{V'(\log \frac{n}{2m})}{V(E^{(n)}_{(n-m)}) - V(E^{(n)}_{(n-2m)})} + \frac{(a_k(p) \frac{2km}{pn})^{\gamma} - 1}{2^{\gamma} - 1} \cdot \frac{V'(\log \frac{n}{2m})}{V(E^{(n)}_{(n-m)}) - V(E^{(n)}_{(n-2m)})} \cdot$$

$$2^{\gamma} \cdot \frac{V'(E^{(n)}_{(n-2m)})}{V'(\log \frac{n}{2m})} \cdot \sqrt{2m}\ \{\frac{V(E^{(n)}_{(n-m)}) - V(E^{(n)}_{(n-2m)})}{2^{\gamma} V'(E^{(n)}_{(n-2m)})} - \frac{1 - 2^{-\gamma}}{\gamma}\}$$

$$+ \frac{(a_k(p) \frac{2km}{pn})^{\gamma} - 1}{\gamma} \cdot \frac{V'(\log \frac{n}{2m})}{V(E^{(n)}_{(n-m)}) - V(E^{(n)}_{(n-2m)})} \sqrt{2m}\ \{\frac{V'(E^{(n)}_{(n-2m)})}{V'(\log \frac{n}{2m})} - 1\}$$

Note that the asymptotics of all terms except the first one are covered by Lemma's 3.6, 3,7 and 3.9 and Corollaries 3.8 and 3.10. For the first term we proceed as follows (write a_n for $a_k(p) \frac{2km}{pn}$).

$$\sqrt{2m}\ \{\frac{a_n^{\gamma + (\hat{\gamma}_n - \gamma)} - 1}{2^{\gamma + (\hat{\gamma}_n - \gamma)} - 1} - \frac{a_n^{\gamma} - 1}{2^{\gamma} - 1}\} \sim [\frac{d}{dx} \frac{a_n^x - 1}{2^x - 1}]_{x=\gamma} \cdot \sqrt{2m}(\hat{\gamma}_n - \gamma)$$

$$= \frac{(\log a_n)\ a_n^{\gamma}(2^{\gamma} - 1) - (a_n^{\gamma} - 1)(\log 2) 2^{\gamma}}{(2^{\gamma} - 1)^2} \sqrt{2m}(\hat{\gamma}_n - \gamma).$$

Examining the proof of theorem 2.3 of Dekkers and de Haan (1987) one sees that $(n\to\infty)$

$$\sqrt{2m}(\hat{\gamma}_n - \gamma) \sim 2^{\gamma}.\sqrt{2m}\{E_{(n-m)}^{(n)} - E_{(n-2m)}^{(n)} - \log 2\} -$$

$$- \sqrt{2m}\{E_{(n-2m)}^{(n)} - E_{(n-4m)}^{(n)} - \log 2\}.$$

Denoting the limit random variables of the independent r.v.'s

$$\sqrt{2m}\{E_{(n-m)}^{(n)} - E_{(n-2m)}^{(n)} - \log 2\}$$

$$\sqrt{2m}\{E_{(n-2m)}^{(n)} - E_{(n-4m)}^{(n)} - \log 2\}$$

$$\sqrt{2m}\{E_{(n-4m)}^{(n)} - \log \frac{n}{4m}\}$$

by M, $2^{-\frac{1}{2}}P$ and $2^{-\frac{1}{2}}Q$ respectively we note that
$\sqrt{2m}\{E_{(n-2m)}^{(n)} - \log \frac{n}{2m}\} \to (P + Q)/\sqrt{2}$ in distribution $(n\to\infty)$.

Assembling the now known asymptotics of all terms of (3.14) we get

$$\sqrt{2m}\,\frac{\hat{x}_{p,n} - x_p}{V(E_{(n-m)}^{(n)}) - V(E_{(n-2m)}^{(n)})} \to$$

$$\to \frac{(\log \frac{2k}{c})(\frac{2k}{c})^{\gamma}(2^{\gamma} - 1) - ((\frac{2k}{c})^{\gamma} - 1)(\log 2)2^{\gamma}}{(2^{\gamma} - 1)^2}\{2^{\gamma}M - 2^{-\frac{1}{2}}P\} + \frac{\gamma}{2^{\gamma} - 1}\,\frac{P+Q}{\sqrt{2}}$$

$$+ \frac{(\frac{2k}{c})^{\gamma} - 1}{2^{\gamma} - 1} \cdot \frac{\gamma}{2^{\gamma} - 1} \cdot 2^{\gamma}M + \frac{(\frac{2k}{c})^{\gamma} - 1}{2^{\gamma} - 1} \cdot \gamma \cdot \frac{P+Q}{\sqrt{2}}$$

in distribution $(n\to\infty)$.

References

D.D. Boos (1984). Using extreme-value theory to estimate large percentiles. Technometrics 26, 33-39.

A.L.M. Dekkers and L. de Haan (1987). On a consistent estimate of the index of an extreme-value distribution. Submitted.

W. Feller (1966). An introduction to probability theory and its applications (Vol II). Wiley New York.

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

N.V. Smirnov (1949). Limit distributions for the terms of a variational series. In: Amer. Math. Soc. Transl. Ser. 1, no. 67

I. Weissman (1978). Estimation of parameters and large quantiles based on the k largest observations. J.A.S.A. 73, 812-815.