Probab. Th. Rel. Fields 85, 499–503 (1990)

Probability Theory and Related Fields

A note on generalized Pareto distributions and the k upper extremes

Michael Falk

Mathematisch-Geographische Fakultät, Katholische Universität Eichstätt, Ostenstrasse 26–28 D-8078 Eichstätt, Federal Republic of Germany

Received May 18, 1989; in revised form November 29, 1989

Summary Denote by $D(n, k)$ the maximum distance between the distribution function of the kth-largest order statistic in an *iid* sample of size n, equally standardized in k, and of its corresponding limiting extreme value distribution.

Suppose that the underlying distribution function is ultimately continuous and strictly increasing. It is shown in the present paper that $D(n, k)$ converges to zero for any sequence $k = k(n)$ with $\lim_{n\to\infty} k/n = 0$ if and only if the underlying distribution is ultimately a generalized Pareto distribution. Thus, *gPds* do not only yield the best rate of joint convergence of extremes, but they are also the only distributions where there actually is convergence.

Introduction and main result

Let $X_1, \ldots, X_n$ be independent and identically distributed random variables ($\equiv rvs$) with common distribution function ($\equiv df$)F and denote by $X_{1:n} \leqq \cdots \leqq X_{n:n}$ the corresponding order statistics.

It is well-known that $(X_{n:n} - b_n)/a_n$ converges in distribution to some nondegenerate limiting distribution G for some choice of constants $a_n > 0$, $b_n \in \mathbb{R}$, if and only if $((X_{n-i+1:n} - b_n)/a_n)_{i=1}^k$ converges weakly to $G^{(k)}$ for any integer k, where $G^{(k)}$ has k-dimensional Lebesgue density $g^{(k)}(x_1, \ldots, x_k) = G(x_k) \prod_{i=1}^{k} G'(x_i)/G(x_i)$ if $x_1 > \cdots > x_k$ and zero elsewhere (Dwass (1966), Weissman (1975)).

Moreover, it is well-known since Gnedenko (1943) that G must be of one of the following types, where $\alpha > 0$: $G_{1,\alpha}(x) := \exp(-x^{-\alpha})$, $x > 0$, $G_{2,\alpha}(x) := \exp(-(-x)^\alpha)$, $x \leqq 0$, $G_3(x) := \exp(-e^{-x})$, $x \in \mathbb{R}$, being the Fréchet, (reversed) Weibull and Gumbel distribution.

Denote by W a generalized Pareto distribution ($\equiv gPd$), i.e. $W(x):=1+\log(G(x))$ if $1/e \leqq G(x) \leqq 1$, yielding

$$W_{1,\alpha}(x)=1-x^{-\alpha}, x \geqq 1$$

$$W_{2,\alpha}(x)=1-(-x)^{\alpha}, x \in [-1,0],$$

$$W_3(x)=1-e^{-x}, x \geqq 0.$$

One of the significant properties of *gPds* is the fact that they yield the best rate of joint convergence of the upper extremes, equally standardized, if the underlying *df* F is ultimately continuous and strictly increasing in its upper tail. This is captured in the following Theorem, where we denote by $\mathscr{B}^k$ the Borel-σ-algebra in $\mathbb{R}^k$.

Theorem 1. *Let F be continuous and strictly increasing in a left neighborhood of $\omega(F):=\sup\{x \in \mathbb{R}: F(x)<1\}$. There exist norming constants $a_n>0$, $b_n \in \mathbb{R}$ and a positive constant C such that for any $k \in \{1,\ldots,n\}$, $n \in \mathbb{N}$,*

$$\sup_{B \in \mathscr{B}^k} |P\{((X_{n-i+1:n}-b_n)/a_n))_{i=1}^k \in B\} - G^{(k)}(B)| \leqq Ck/n$$

if and only if there exist $c>0$, $d \in \mathbb{R}$ such that $F(x)=W(cx+d)$ for x near $\omega(F)$, where W is the gPd pertaining to G.

The *if*-part of this result is due to Reiss (1981, Theorems 2.6 and 3.2) while the *only if*-part has been proved in Falk (1989, Theorem 10).

It is clear from this result that *gPds* play a central role in extreme value theory if the joint distribution of the k largest observations is considered; this is usually done in statistical applications and was the starting point in the paper by Pickands (1975), who first observed the importance of *gPds*. A rigorous approach based on *gPds* to the problem of estimating tails of probability distributions was carried out by Smith (1987). For a detailed description of properties of *gPds* and of their particular role in extreme value theory we refer to Falk (1986, 1989) and to the recent monograph by Reiss (1989, Sects. 1 and 5). In particular it is shown there that the rate of joint convergence uniformly over all Borel sets of the k largest order statistics, i.e. the left hand side of the displayed formula in the preceding result, is determined by the distance of the underlying distribution F from the corresponding shifted *gPd*.

Now the bound in Theorem 1 tends to zero as n tends to infinity if $k=k(n)$ satisfies $k/n \to_{n\to\infty} 0$ and hence, the problem suggests itself to characterize those *dfs* F such that the distance between the joint distribution of the k largest order statistics in a sample of size n, equally standardized in k, and its limit $G^{(k)}$ converges to zero for *any* sequence $k=k(n)$ such that $k/n \to_{n\to\infty} 0$.

This problem led to the following main result of the present paper. It reveals that *gPds* do not only yield the best rate of joint convergence of the upper $k=k(n)$ extremes, but that they are also the only distributions where there actually is convergence for *any* choice of k satisfying the above conditions.

Throughout the rest of this paper we suppose that F is ultimately continuous and strictly increasing in its upper tail. By $G_{(k)}$ we denote the k-th one-dimensional marginal distribution of $G^{(k)}$, i.e. $G_{(k)}(x)=G(x)\sum_{j=0}^{k-1}(-\log(G(x)))^j/j!$ if $G(x) \in (0,1)$.

Theorem 2. *If there exist constants $a_n>0$, $b_n \in \mathbb{R}$ such that*

$$\sup_t |P\{(X_{n-k+1:n}-b_n)/a_n \leqq t\} - G_{(k)}(t)| \to_{n\to\infty} 0$$

for any sequence $k = k(n) \in \{1, \ldots, n\}$, $n \in \mathbb{N}$, *with* $k/n \to_{n\to\infty} 0$, *then there exist* $c > 0$, $d \in \mathbb{R}$, *such that* $F(x) = W(cx + d)$ *for* x *near* $\omega(F)$, *where* W *is the gPd pertaining to* G.

The following consequence is obvious.

Corollary. *If there exist constants* $a_n > 0$, $b_n \in \mathbb{R}$ *such that for any* $k \in \{1, \ldots, n\}$, $n \in \mathbb{N}$

$$\sup_t |P\{(X_{n-k+1:n} - b_n)/a_n \leqq t\} - G_{(k)}(t)| \leqq g(k/n),$$

where $g: \mathbb{R} \to \mathbb{R}$ *satisfies* $\lim_{x\to 0} g(x) = 0$, *then the conclusion of Theorem 2 holds.*

Remarks. (i) With the particular choice $g(x) = Cx$ the preceding result obviously yields the *only if*-part of Theorem 1.

(ii) We do not know whether the smoothness assumptions on the upper tail of F can be dropped in Theorem 2.

Auxiliary results and proofs

The following auxiliary result will be basic for the proof of Theorem 2. By $[x]$ we denote the integer part of $x \in \mathbb{R}$.

Lemma. *Denote by* $D(n, k)$ *the left hand side of the displayed formula in Theorem 2. Suppose that* $\lim_{n\to\infty} D(n, k) = 0$ *if* $k/n \to_{n\to\infty} 0$. *Then, for any* $\delta > 0$, *there exists* $\varepsilon_\delta > 0$ *such that* $D(n, [\varepsilon n]) < \delta$ *as well as* $D(n, 2[\varepsilon n/2]) < \delta$ *if* n *is large for any* $0 < \varepsilon \leqq \varepsilon_\delta$.

Proof. Negate the assertion, i.e. suppose that for any $\varepsilon_0 > 0$ there exists $\varepsilon \leqq \varepsilon_0$ such that for any n there exists $m \geqq n$ with $D(m, [\varepsilon m]) \geqq \delta$.

Choose $\varepsilon_k \downarrow 0$ and $m_k \uparrow \infty$ such that $D(m_k, [\varepsilon_k m_k]) \geqq \delta$. Put for $m \in \mathbb{N}$

$$k(m) := [\varepsilon_k m] \quad \text{if } m \in [m_k, m_{k+1} - 1).$$

Then, obviously $k(m)/m \to_{m\to\infty} 0$ but clearly

$$\limsup_{m\to\infty} D(m, k(m)) \geqq \delta$$

which is a contradiction to the assumptions of the lemma. The same arguments apply to $2[\varepsilon n/2]$ in place of $[\varepsilon n]$. This yields the assertion. □

Proof of Theorem 2. At first we prove the assertion for $G_{2,1,(k)}$. This is the easiest case, but it is also the easiest way to demonstrate the basic ideas of the proof.

First note that $G_{2,1,(k)}$ is the *df* of $-\sum_{j=1}^{k} Y_j$, where $Y_1, Y_2, \ldots$ are *iid rvs* with common standard exponential distribution. Moreover, let $X^{(i)}_{n-k+1:n}$, $i = 1, 2$, be independent replica of $X_{n-k+1:n}$.

With $D(n, k)$ as defined in the preceding lemma we obtain by Fubini's theorem for $k \leqq n$

$$\sup_t \left| P\left\{ \sum_{i=1}^{2} (X^{(i)}_{n-k+1:n} - b_n)/a_n \leqq t \right\} - G_{2,1,(2k)}(t) \right|$$
$$\leqq 2 \sup_t |P\{(X_{n-k+1:n} - b_n)/a_n \leqq t\} - G_{2,1,(k)}(t)|$$
$$= 2D(n, k).$$

On the other hand, if $2k \leqq n$,

$$\sup_t |P\{(X_{n-2k+1:n}-b_n)/a_n \leqq t\} - G_{2,1,(2k)}| = D(n, 2k)$$

and hence,

$$\sup_t \left| P\left\{ \sum_{i=1}^{2} (X^{(i)}_{n-k+1:n} - b_n) \leqq t \right\} - P\{(X_{n-2k+1:n} - b_n) \leqq t\} \right|$$
$$\leqq 2D(n,k) + D(n,2k). \tag{1}$$

By the preceding lemma there exists $\varepsilon_0 > 0$ such that for any $\varepsilon \leqq \varepsilon_0$

$$D(n, [\varepsilon n]) \leqq 1/4, \qquad D(n, 2[\varepsilon n/2]) \leqq 1/4$$

if n is large. Chooce $\varepsilon \in (0, \varepsilon_0)$. Then, with the particular choice $k = k(n) = [\varepsilon n/2]$ we obtain from (1) if n is large

$$\sup_t \left| P\left\{ \sum_{i=1}^{2} (X^{(i)}_{n-[\varepsilon n/2]+1:n} - b_n) \leqq t \right\} - P\{(X_{n-2[\varepsilon n/2]+1:n} - b_n) \leqq t\} \right|$$
$$\leqq 2D(n, [\varepsilon n/2]) + D(n, 2[\varepsilon n/2]) \leqq 3/4. \tag{2}$$

From classical extreme value theory (e.g. 2.1, 2.2 in the book by Galambos (1987) or Theorem 2.3.2 in the book by de Haan (1975)), we conclude that $b_n \to_{n\to\infty} \omega(F) =: b < \infty$. Moreover, if ε_0 is small enough, continuity and strict monotonicity of F yield

$$X_{n-[n\varepsilon/2]:n} \to_{n\to\infty} F^{-1}(1-\varepsilon/2), \qquad X_{n-2[n\varepsilon/2]:n} \to_{n\to\infty} F^{-1}(1-\varepsilon)$$

in probability, and hence, (2) implies

$$2(F^{-1}(1-\varepsilon/2) - b) = F^{-1}(1-\varepsilon) - b \tag{3}$$

which now holds for any $\varepsilon \in (0, \varepsilon_0)$.

From Theorem 2.3.2 and Corollary 1.2.1, 5 by de Haan (1975) we know that the function $u(\varepsilon) := F^{-1}(1-\varepsilon) - b$, $\varepsilon \in (0,1)$, satisfies $u(t\varepsilon)/u(\varepsilon) \to_{\varepsilon\to\infty} t$, $t > 0$. Iterating (3) we obtain that for any $\varepsilon \in (0, \varepsilon_0)$ and $m \in \mathbb{N}$

$$u(\varepsilon) = 2^m u(\varepsilon/2^m)$$

and thus, we deduce for any $\varepsilon_1, \varepsilon_2 \in (0, \varepsilon_0)$

$$\frac{u(\varepsilon_1)}{u(\varepsilon_2)} = \frac{u((\varepsilon_1/\varepsilon_2)\varepsilon_2/2^m)}{u(\varepsilon_2/2^m)} \to_{m\to\infty} \varepsilon_1/\varepsilon_2,$$

i.e. $u(\varepsilon) = c\varepsilon$, $0 < \varepsilon < \varepsilon_0$, for some $c > 0$. This completes the proof of the case $G = G_{2,1}$.

Observe now that $G_{2,\alpha,(k)}$ is the distribution of $-\left(\sum_{j=1}^{k} Y_j\right)^{1/\alpha}$, $G_{1,\alpha,(k)}$ the one of $\left(\sum_{j=1}^{k} Y_j\right)^{-1/\alpha}$, and $G_{3,(k)}$ the one of $-\log\left(\sum_{j=1}^{k} Y_j\right)$. The general case $G_{2,\alpha,(k)}$ as well as $G_{1,\alpha,(k)}$ can then easily be dealt with in complete analogy to the preceding proof.

Some extra remarks might be useful for the final case $G_{3,(k)}$. Repeating the above arguments we obtain that in this case there exist $a > 0$ and $\varepsilon_0 > 0$ such that

$$2\exp(-F^{-1}(1-\varepsilon/2)/a) = \exp(-F^{-1}(1-\varepsilon)) \tag{4}$$

for any $\varepsilon \in (0, \varepsilon_0)$. From Theorem 2.4.1 in the book by de Haan (1975) we know that the function $v(\varepsilon) := F^{-1}(1-\varepsilon)$, $\varepsilon \in (0,1)$, satisfies

$$\frac{v(\varepsilon x) - v(\varepsilon)}{v(\varepsilon y) - v(\varepsilon)} \to_{\varepsilon \to 0} \frac{\log(x)}{\log(y)},$$

$x, y > 0$, $y \neq 1$. Iterating (4) and taking logarithms we obtain that for $\varepsilon \in (0, \varepsilon_0)$ and $m \in \mathbb{N}$

$$v(\varepsilon) = -m \log(2) + v(\varepsilon/2^m)/a$$

which implies $\omega(F) = \infty$. We obtain further that for $\varepsilon, \varepsilon_1, \varepsilon_2 \in (0, \varepsilon_0)$, $\varepsilon_1 \neq \varepsilon_2$

$$\frac{v(\varepsilon) - v(\varepsilon_1)}{v(\varepsilon_2) - v(\varepsilon_1)} = \frac{v(\varepsilon/2^m) - v(\varepsilon_1/2^m)}{v(\varepsilon_2/2^m) - v(\varepsilon_1/2^m)} \to_{m \to \infty} \frac{\log(\varepsilon/\varepsilon_1)}{\log(\varepsilon_2/\varepsilon_1)},$$

i.e. $v(\varepsilon) = c_1 - c_2 \log(\varepsilon)$, $\varepsilon \in (0, \varepsilon_0)$, where c_1, c_2 are positive constants. This implies the assertion. □

Acknowledgement. I would like to thank the referee for useful comments which led to a more concise proof of Theorem 2.

References

Dwass, J.P.: Extremal processes II. Ill. J. Math. **10**, 381–391 (1966)

Falk, M.: Rates of uniform convergence of extreme order statistics. Ann. Inst. Stat. Math. A**38**, 245–262 (1986)

Falk, M.: Best attainable rate of joint convergence of extremes. In: Hüsler, J., Reiss, R.-D. (eds.) Extreme value theory. (Lect. Notes Stat., pp. 1–9) New York Berlin Heidelberg: Springer 1989

Galambos, J.: The asymptotic theory of extreme order statistics, 2nd. ed. Melbourne, Florida: Krieger 1987

Gnedenko, B.: Sur la distribution limite du terme maximum d'une série aléatoire. Ann. Math. **44**, 423–453 (1943)

Haan, L. de: On regular variation and its application to the weak convergence of sample extremes. 2nd. ed. Amsterdam: Math. Centre Tracts 1975

Pickands, J.III: Statistical inference using extreme order statistics. Ann. Stat. **3**, 119–131 (1975)

Reiss, R.-D.: Uniform approximation to distributions of extreme order statistics. Adv. Appl. Probab. **13**, 533–547 (1981)

Reiss, R.-D.: Approximate distributions of order statistics (with applications to nonparametric statistics). Springer Ser. Stat. New York Berlin Heidelberg: Springer 1989

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

Weissman, I.: Multivariate extremal processes generated by independent nonidentically distributed random variables. J. Appl. Probab. **12**, 477–487 (1975)