SEQUENTIAL METHODS IN STATISTICS
BANACH CENTER PUBLICATIONS, VOLUME 16
PWN-POLISH SCIENTIFIC PUBLISHERS
WARSAW 1985

LIMIT THEOREMS FOR PARETO-TYPE DISTRIBUTIONS

A. DE MEYER and J. L. TEUGELS

Department of Mathematics, Catholic University of Leuven, Leuven, Belgium

We derive some limit theorems for large order statistics from a Pareto-type distribution. In a few cases these theorems yield asymptotic confidence intervals for the index of the Pareto-type distribution.

1. Introduction

Let $X_1, X_2, \ldots, X_n$ be a sample of size n from X with d.f. F on $[0, \infty)$. The order statistics of the sample will be denoted by

$$X_1^* \leqslant X_2^* \leqslant \ldots \leqslant X_n^*.$$

We denote by $\mathscr{P}(\alpha)$ the set of all distributions on $[0, \infty)$ for which

$$1-F(x) \sim x^{-\alpha} L^{-\alpha}(x), \qquad \alpha > 0$$

where L is slowly varying (s.v.). F is then of Pareto-type.

One of the main statistical problems connected with $\mathscr{P}(\alpha)$ is the estimation of α; see [3], [4], [6]. Recently the construction of confidence intervals received some attention as well [3], [4]. Our main goal in this paper is to derive limit theorems that are useful in the construction of asymptotic confidence intervals for α^{-1}.

2. Limit theorems for one order statistic

Recall from general theory that for any $1 \leqslant k \leqslant n$

$$G_n(x) \equiv P\{X_{n-k+1}^* \leqslant x\} = \frac{n!}{(n-k)!\,(k-1)!} \int_0^x F^{n-k}(u)[1-F(u)]^{k-1}\, dF(u). \quad (1)$$

Choosing $\{a_n\}$ as a normalizing sequence and substituting $u = F^i\left(1-\frac{s}{n}\right)$ (where F^i denotes the inverse of F) one obtains easily

$$P\{a_n^{-1} X_{n-k+1}^* \leqslant x\} = \frac{n!\, n^{-k}}{(n-k)!} \frac{1}{(k-1)!} \int_{\Delta_n(x)}^{n} \left(1-\frac{s}{n}\right)^{n-k} s^{k-1}\, ds$$

where $\Delta_n(x) = n\{1-F(a_n x)\}$. If $F \in \mathscr{P}(\alpha)$ then it is natural to take $n\{1-F(a_n)\} = \Delta_n(1) \to 1$; since then $\Delta_n(x) \to x^{-\alpha}$. We obtain the well-known result ([10]):

THEOREM 1. *If* $F \in \mathscr{P}(\alpha)$ *and* $n\{1-F(a_n)\} \to 1$ *then for fixed* k

$$P\{a_n^{-1} X_{n-k+1}^* \leqslant x\} \to \frac{1}{(k-1)!} \int_{x^{-\alpha}}^{\infty} e^{-s} s^{k-1}\, ds.$$

As is well known from the theory of regular variation ([9]) one can obtain an asymptotic expression for $\{a_n\}$, i.e.,

$$a_n \sim n^{1/\alpha} L^*(n^{1/\alpha})$$

where L^* is the conjugate of L.

COROLLARY 1. *If* $F \in \mathscr{P}(\alpha)$ *and* $n\{1-F(a_n)\} \to 1$ *then for fixed* k

$$(\log n)^{-1} \cdot \log X_{n-k+1}^* \xrightarrow{P} 1/\alpha.$$

Proof.

$$P\{(\log n)^{-1} \cdot \log X_{n-k+1}^* \leqslant w\} = P\{a_n^{-1} X_{n-k+1}^* \leqslant n^w a_n^{-1}\}.$$

However

$$n^w a_n^{-1} \sim n^{w-1/\alpha}\left(L^*(n^{1/\alpha})\right)^{-1} \to \begin{cases} \infty & \text{if } w > 1/\alpha, \\ 0 & \text{if } w < 1/\alpha. \end{cases}$$

Since the limit in Theorem 1 is continuous the convergence in Theorem 1 is uniform [5, p. 139]. Hence the result. ∎

From the statistical point of view Theorem 1 is useless since $\{a_n\}$ is unknown. The corollary shows that we can get a consistent estimator for α^{-1} but k plays no role at all.

3. Limit theorems for one order statistic with floating k

We would like to make k depending on n. One could assume $k/n \to \lambda \in (0, 1)$ as is done in [10]; for $\lambda = 1$ see [2]. For a general treatment see the two papers by Balkema and de Haan ([1]). We shall assume $\lambda = 0$ since the largest order statistics will contain most information on α.

Change $1-F(u) = s$ in (1) and put then $s = q_n + p_n x$ where $\{p_n\}_2^\infty$ and $\{q_n\}_2^\infty$ will be determined shortly. Then

$$G_n(u) = I_1 \int_{\Delta_n(u)}^{\delta_n} I_2(x)\,dx$$

where

$$I_1 = \frac{n!}{(n-k)!\,(k-1)!}\, p_n(1-q_n)^{n-k} q_n^{k-1},$$

$$I_2(x) = \left(1 - \frac{p_n}{1-q_n} x\right)^{n-k} \left(1 + \frac{p_n}{q_n} x\right)^{k-1},$$

$$\delta_n = (1-q_n)/p_n,$$

$$\Delta_n(u) = p_n^{-1}\{1 - F(u) - q_n\}.$$

As shown in [10], [11] a proper choice for $\{p_n\}$ and $\{q_n\}$ to let $I_2(x)$ converge for $n \to \infty$, and $k_n/n \to 0$ is

$$q_n = \frac{k-1}{n-1} \quad \text{or} \quad 1 - q_n = \frac{n-k}{n-1},$$

$$p_n^2 = (n-k)(k-1)/(n-1)^3.$$

It then easily follows that for $T > 0$, fixed, $\log I_2(x) \to -\frac{1}{2}x^2$ uniformly on $[-T, +T]$. Stirling's formula yields then $I_1 \to 1/\sqrt{2\pi}$ while $\delta_n \to +\infty$. We only need to find a limit for $\Delta_n(u)$.

Since $q_n \to 0$, $p_n \to 0$ also $u \to \infty$. Replace u by $c_n u$ where $1 - F(c_n) \sim k/n$. Given $\varepsilon > 0$ pick $n_0(\varepsilon)$ such that for $n > n_0$ and $u > 0$

$$(1-\varepsilon)k/n \leqslant 1 - F(c_n) \leqslant (1+\varepsilon)k/n,$$

$$(u^{-\alpha} - \varepsilon)\{1 - F(c_n)\} \leqslant 1 - F(c_n u) \leqslant (u^{-\alpha} + \varepsilon)\{1 - F(c_n)\}.$$

Combining these inequalities we obtain

$$\Delta_n(c_n u) \to \begin{cases} -\infty & \text{if } u > 1, \\ +\infty & \text{if } u < 1. \end{cases}$$

Now take $u > 1$. Then for a fixed T

$$|G_n(c_n u) - 1| \leqslant \left| I_1 \int_{-T}^{T} I_2(x)\,dx - \frac{1}{\sqrt{2\pi}} \int_{-T}^{T} e^{-x^2/2}\,dx \right| +$$

$$+ \frac{2}{\sqrt{2\pi}} \int_{T}^{\infty} e^{-x^2/2}\,dx + I_1 \int_{T}^{\delta_n} I_2(x)\,dx + I_1 \int_{\Delta_n(c_n u)}^{-T} I_2(x)\,dx.$$

The last two terms are easily handled ([10], [11]). A similar argument works for $u < 1$; we obtain

THEOREM 2. *If* $F \in \mathscr{P}(\alpha)$, $n \to \infty$, $k \to \infty$, $k/n \to 0$ *and* $1-F(c_n) \sim k/n$ *then*

$$c_n^{-1} X^*_{n-k+1} \xrightarrow{P} 1.$$

One can do better by looking at $d_n^{-1}\{X^*_{n-k+1} - c_n\}$. The only change in the argument is the limiting value of $\Delta_n(d_n u + c_n)$ for $|u| \leqslant T$. We write

$$\begin{aligned}\Delta_n(d_n u + c_n) &= p_n^{-1}\{1 - F(d_n u + c_n) - q_n\}\\ &= p_n^{-1}[1-F(c_n)]\left(1+\frac{d_n}{c_n}u\right)^{-\alpha}\left\{\frac{L(d_n u + c_n)}{L(c_n)} - 1\right\} +\\ &\quad + p_n^{-1}[1-F(c_n)]\left\{\left(1+\frac{d_n}{c_n}u\right)^{-\alpha} - 1\right\} + p_n^{-1}\{1-F(c_n) - q_n\}\\ &\equiv I_{n1} + I_{n2} + I_{n3}.\end{aligned}$$

It seems natural to assume L to be so-called normalized, i.e.,

$$L(x) = c\exp\int_1^x \frac{\varepsilon(u)}{u}\,du \tag{2}$$

where $c > 0$ and $\varepsilon(u) \to 0$ as $u \to \infty$. In particular one can take $1-F(c_n) = k/n \sim q_n$ so that $I_{n3} \to 0$. Further

$$I_{n2} \sim p_n^{-1} q_n \left\{\left(1+\frac{d_n}{c_n}u\right)^{-\alpha} - 1\right\} \sim -\alpha\sqrt{k}\,\frac{d_n}{c_n}u$$

if $d_n/c_n \to 0$. So pick $d_n = c_n k^{-1/2}$. Finally

$$\begin{aligned} I_{n1} &\sim \sqrt{k}\left\{\exp\int_{c_n}^{d_n u + c_n} \varepsilon(u)\cdot u^{-1}\,du - 1\right\}\\ &\sim \sqrt{k}\left\{\exp\int_1^{1+uk^{-1/2}} \varepsilon(c_n v)\cdot v^{-1}\,dv - 1\right\} \to 0\end{aligned}$$

since $k \to \infty$ and $\varepsilon(c_n v) \to 0$. We have proved

THEOREM 3. *Let* $1-F(x) = x^{-\alpha}L^{-\alpha}(x)$ *where* L *satisfies* (2); *take* $1-F(c_n) = k/n \to 0$ *and* $d_n = k^{-1/2}c_n$. *Then for* $n \to \infty$, $k \to \infty$

$$d_n^{-1}\{X^*_{n-k+1} - c_n\} \xrightarrow{\mathscr{D}} Y \sim \mathscr{N}\left(0, \frac{1}{\alpha}\right).$$

EXAMPLE 1. Assume $F(x) = \exp -x^{-\alpha}$. Then $F \in \mathscr{P}(\alpha)$ and L satisfies (2) with $c = 1$ and $\varepsilon(x) = 1 + x^{-\alpha}\{1 - \exp x^{-\alpha}\}^{-1}$. As in the proof of Corollary 1 one can write

$$P\left\{\sqrt{k}\left[\alpha \log X^*_{n-k+1} - \log \frac{n}{k}\right] \leqslant u\right\} = P\{d_n^{-1}(X^*_{n-k+1} - c_n) \leqslant \varphi_n(u)\}$$

where $\varphi_n(u) = \sqrt{k}\left\{c_n^{-1}\left(\frac{n}{k}\right)^{1/\alpha} \exp \frac{u}{\alpha\sqrt{k}} - 1\right\}$. Direct calculation shows that $c_n = \left(\frac{n}{k}\right)^{1/\alpha}\left(1 + O\left(\frac{k}{n}\right)\right)$. Henceforth

$$\varphi_n(u) = \sqrt{k}\left\{c_n^{-1}\left(\frac{n}{k}\right)^{1/\alpha} - 1\right\} + \sqrt{k}\, c_n^{-1}\left(\frac{n}{k}\right)^{1/\alpha}\left\{\exp \frac{u}{\alpha\sqrt{k}} - 1\right\}.$$

The second term tends to u/α; the first tends to zero if we take $k = o(n^{2/3})$. Hence if $k = o(n^{2/3})$ then by Theorem 3

$$\sqrt{k}\left[\alpha \log X^*_{n-k+1} - \log \frac{n}{k}\right] \xrightarrow{\mathscr{D}} \mathscr{N}(0, 1).$$

We obtain an asymptotic confidence interval for α. Needless to say that the same procedure works for a large class of members of $\mathscr{P}(\alpha)$. ■

4. Limit theorems for two order statistics

Put

$$G^{(n)}_{m,k}(u, v) = P\{X^*_{n-k-m+1} \leqslant u,\ X^*_{n-k+1} \leqslant v\}$$
$$= \frac{n!}{(n-m-k)!\,(k-1)!\,(m-1)!} \int_0^{u \wedge v} dF(x) \int_x^v dF(y)\, F^{n-k-m}(x)\,[F(y) - F(x)]^{m-1}\,[1 - F(y)]^{k-1}$$

where $u \wedge v = \min(u, v)$.

The analogue of Theorem 1 can be found in [8].

THEOREM 4. *If $F \in \mathscr{P}(\alpha)$ and $n\{1 - F(a_n)\} \to 1$ then for m and k fixed*

$$G^{(n)}_{m,k}(a_n u, a_n v) \to \frac{1}{(k-1)!\,(m-1)!} \cdot \int_{(u \wedge v)^{-\alpha}}^{\infty} dx \int_{v^{-\alpha}}^{x} dy e^{-x}(x-y)^{m-1} y^{k-1}.$$

Since $\{a_n\}$ norms both order statistics we get

COROLLARY 2. *If $F \in \mathscr{P}(\alpha)$ then for m and k fixed and $w \in (0, 1)$*

$$P\left\{\log(X^*_{n-m-k+1}/X^*_{n-k+1}) \leqslant \frac{1}{\alpha}\log w\right\} \to \frac{1}{B(m, k)} \int_0^w r^{k-1}(1-r)^{m-1}\, dr. \quad (3)$$

For every pair (k, m), (3) provides an asymptotic confidence interval for α.

We make m dependent on n, keeping k fixed. Write

$$c_{n,k,m} = n!\,\{(n-m-k)!\,(k-1)!\,(m-1)!\}^{-1}.$$

In a similar fashion as in Section 3 we write

$$G^{(n)}_{m,k}(u, v) = I_1 \int_{\Delta_n}^{\delta_n} \left(1 - \frac{p_n}{1-q_n} x\right)^{n-k-m} \left(1 + \frac{p_n}{q_n} x\right)^{m+k-1} dx \times$$

$$\times \int_{\Delta'_n}^{m} \left(1 - \frac{z}{m}\right)^{m-1} z^{k-1}\, dz$$

where

$$I_1 = c_{n,k,m}\, m^{-k} p_n (1-q_n)^{n-k-m} q_n^{m+k-1},$$

$$\delta_n = p_n^{-1}(1-q_n),$$

$$\Delta_n \equiv \Delta_n(u, v) = p_n^{-1}\{1 - F(u \wedge v) - q_n\},$$

$$\Delta'_n \equiv \Delta'_n(v, x) = \{q_n + p_n x\}^{-1}\, m\,\{1 - F(v)\}.$$

A good choice of $\{p_n\}$ and $\{q_n\}$ is now

$$q_n = (n-1)^{-1}(m+k-1), \qquad p_n^2 = (n-1)^{-3}(m+k-1)(n-k-m).$$

If we replace v by $a_n v$ then $\Delta'_n(a_n v, x) \to v^{-\alpha}$. Replace u by $d_n u$ where $1 - F(d_n) \sim m/n \to 0$ then

$$\Delta_n(d_n u, a_n v) \sim \Delta_n(d_n u, a_n u) \to \begin{cases} +\infty & \text{if} \quad u < 1, \\ -\infty & \text{if} \quad u > 1. \end{cases}$$

We obtain:

THEOREM 5. *Let $F \in \mathscr{P}(\alpha)$, $n[1 - F(a_n)] \to 1$, $1 - F(d_n) \sim m/n \to 0$ as $m \to \infty$, $n \to \infty$. Then*

$$P\{X^*_{n-m-k+1} \leqslant d_n u,\ X^*_{n-k+1} \leqslant a_n v\} \to \frac{1}{(k-1)!} U_1(u) \int_{v^{-\alpha}}^{\infty} e^{-z} z^{k-1}\, dz$$

$$\textit{where} \qquad U_c(u) = \begin{cases} 0 & \textit{if} \quad u < c, \\ 1 & \textit{if} \quad u \geqslant c. \end{cases}$$

Comparison with Theorem 2 for the first and with Theorem 1 for the second component yields asymptotic independence of the normalized order statistics.

COROLLARY 3. *Let* $1-F(x) \sim cx^{-\alpha}$ $(c > 0)$. *Then for* $m \to \infty$, $m/n \to 0$ *as* $n \to \infty$

$$P\{\alpha \log(X^*_{n-m-k+1}/X^*_{n-k+1}) + \log m \leqslant v\} \to \frac{1}{(k-1)!} \int_0^{e^v} e^{-t} t^{k-1}\, dt.$$

Proof. The left side equals

$$\int_0^\infty P\{X^*_{n-m+k+1} \leqslant a_n m^{-1/\alpha} e^{v/\alpha} y,\ X^*_{n-k+1} a_n^{-1} \in [y, y+dy)\}.$$

However $d_n \sim (nc/m)^{1/\alpha}$ while $a_n \sim (nc)^{1/\alpha}$ so that $d_n^{-1} a_n m^{-1/\alpha} \to 1$. By Theorem 4 and easy reduction, the result follows. ∎

Corollary 3 for $k = 1$ leads to the double exponential law, known in extreme value theory. In this form the corollary was derived by different methods by de Haan and Resnick ([3]) for the case where F is stable.

EXAMPLE 2. Let

$$1 - F(x) = \exp -\alpha\psi(x)$$

where $\psi(x) = \log x + (\log x)^\beta$ for $x \geqslant 1$ where $\beta \in (0, 1/2)$. Then it easily shown ([9]) that

$$L^*(x) \sim \exp(\log x)^\beta.$$

Hence in Theorem 4

$$\frac{a_n}{d_n} \sim \frac{n^{1/\alpha} L^*(n^{1/\alpha})}{(n/m)^{1/\alpha} L^*[(n/m)^{1/\alpha}]}.$$

Choose $m = \sqrt{n}$ then

$$\frac{a_n}{d_n} \sim m^{1/\alpha} \exp\{(\log n^{1/\alpha})^\beta [1 - 2^{-\beta}]\}.$$

Hence $m^{1/\alpha}$ is not a norming constant for the ratio of $X^*_{n-m-k+1}$ and X^*_{n-k+1}. This implies that some condition on F is necessary in Corollary 3. Nevertheless it would be sufficient to require that $m \to \infty$, $m/n \to 0$ in such a way that

$$L^*(n)/L^*(n/m) \to 1.$$ ∎

Finally we generalize Theorem 3:

THEOREM 6. *Assume* $F \in \mathscr{P}(\alpha)$ *where* L *satisfies* (2). *Define* $1-F(b_k) = k/n$ *and* $\alpha d_k = k^{-1/2} b_k$. *Assume* $k \to \infty$, $m \to \infty$, $k/n \to 0$, $m/n \to 0$ *and* $k/m \to \theta$. *Then*

$$\left(\frac{X^*_{n-m-k+1} - b_{m+k}}{d_{m+k}}, \frac{X^*_{n-k+1} - b_k}{d_k}\right) \xrightarrow{\mathscr{D}} (U, V)$$

where (U, V) *is bivariate normal with standard marginals and* $\varrho = \{\theta/(1+\theta)\}^{1/2}$.

We omit the proof which is somewhat lengthy but follows the same pattern as the proof of Theorem 3, [7].

EXAMPLE 3. For the situations discussed in Example 1 one obtains that if $k+m = o(n^{2/3})$ then with the conditions as in Theorem 6

$$\left(\sqrt{m+k}\left[\alpha \log X^*_{n-m-k+1} - \log\frac{n}{m+k}\right], \sqrt{k}\left[\alpha \log X^*_{n-k+1} - \log\frac{n}{k}\right]\right) \xrightarrow{\mathscr{D}} (U, V).$$

■

References

[1] A. A. Balkema and L. de Haan, *Limit distributions for order statistics*, Theory Probab. Appl., Part I, **23** (1978), 77–92, Part II, **23** (1978), 341–358.

[2] D. M. Chibisov, *On limit distributions for order statistics*, Theory Probab. Appl. **9** (1964), 142–147.

[3] L. de Haan and S. Resnick, *A simple asymptotic estimate for the index of a stable distribution*, J. Roy. Statist. Soc. Ser. **B 42** (1980), 83–88.

[4] P. Hall, *On the Haan and Resnick's estimates of the index of a stable law*, preliminary report, 1981.

[5] I. A. Ibragimov and Y. V. Linnik, *Independent and Stationary Sequences of Random Variables*, Wolters–Noordhoff, Groningen 1971.

[6] D. V. Iliescu and V. Voda, *Some notes on Pareto distributions*, Rev. Roumaine Math. Pures Appl. **24** (1979), 597–609.

[7] A. Maes, *Schattingen van parameters in verdelingsfunkties van het Pareto-type*, Licence Thesis, Leuven 1981.

[8] H. J. Rossberg, *Über die Verteilungsfunktionen der Differenzen und Quotienten von Ranggrössen*, Math. Nachr. **21** (1960), 37–79.

[9] E. Seneta, *Regularly varying functions*, Lecture Notes in Math. **508**, Springer-Verlag, Berlin 1976.

[10] N. V. Smirnov, *Limit distributions for the terms of a variational series*, Amer. Math. Soc. Transl. **11** (1962), 82–143.

[11] J. L. Teugels, *Limit theorems on order statistics*, Ann. Probab. **9** (1981), 868–880.

Presented to the semester
Sequential Methods in Statistics
September 7–December 11, 1981