

Middlesex University Research Repository

An open access repository of

Middlesex University research

<http://eprints.mdx.ac.uk>

Novak, Serguei ORCID: <https://orcid.org/0000-0001-7929-7641> (1992) Inference about the Pareto-type distribution. Transactions of the Eleventh Prague Conference on Information Theory, Statistical Decision Functions, Random Processes. In: Transactions of the Eleventh Prague Conference on Information Theory, Statistical Decision Functions, Random Processes, 26-31 Aug 1990, Prague. . [Conference or Workshop Item]

Final accepted version (with author's formatting)

This version is available at: <https://eprints.mdx.ac.uk/20934/>

Copyright:

Middlesex University Research Repository makes the University's research available electronically.

Copyright and moral rights to this work are retained by the author and/or other copyright owners unless otherwise stated. The work is supplied on the understanding that any use for commercial gain is strictly forbidden. A copy may be downloaded for personal, non-commercial, research or study without prior permission and without charge.

Works, including theses and research projects, may not be reproduced in any format or medium, or extensive quotations taken from them, or their content changed in any way, without first obtaining permission in writing from the copyright holder(s). They may not be sold or exploited commercially in any format or medium without the prior written permission of the copyright holder(s).

Full bibliographic details must be given when referring to, or quoting from full items including the author's name, the title of the work, publication details where relevant (place, publisher, date), pagination, and for theses or dissertations the awarding institution, the degree type awarded, and the date of the award.

If you believe that any material held in the repository infringes copyright law, please contact the Repository Team at Middlesex University via the following email address:

eprints@mdx.ac.uk

The item will be removed from the repository while any claim is being investigated.

See also repository copyright: re-use policy: <http://eprints.mdx.ac.uk/policies.html#copy>

Transactions 11th Prague Conf. Information Theory,
Statistical Decision functions, Random Processes.-
Prague, 1992, v. B, pp. 254-258.

INFERENCE ABOUT

THE PARETO-TYPE DISTRIBUTION

Sergey Novak

Novosibirsk

Key words: tail index , nonparametric inference.

ABSTRACT

Given a sample from the Pareto-type distribution $P(X \geq x) = L(x)x^{-b}$, where $L(x)$ slowly varies as $x \rightarrow \infty$, we estimate the tail index b . For the special case $L(x) \rightarrow L \equiv \text{const}$ as $x \rightarrow \infty$ we construct asymptotically normal estimator of L .

INTRODUCTION AND RESULTS

Let X, X_1, X_2, \dots be a sequence of independent identically distributed random variables with distribution

$$(A) \quad P_x \equiv \mathbb{P}(X \geq x) = L(x)x^{-b} \quad (b > 0),$$

where (unknown) function $L(x)$ slowly varies as $x \rightarrow \infty$; number b is to be estimated.

Well-known estimators of index b have been introduced by Hill (1975), Teugels (1985), S. Scorgo, Deheuvels, Mason (1985), M. Scorgo, Horvath, Révész (1987). Novak and Utev (1990) investigated the estimator

$$t_n = \frac{\sum_{i=1}^n \ln(X_i/N) 1_{\{X_i \geq N\}}}{\sum_{i=1}^n 1_{\{X_i \geq N\}}},$$

where $N = N(n)$ is chosen so that $np_N \rightarrow \infty$ as $n \rightarrow \infty$. They proved that

$$(1) \quad (np_N)^{1/2} (t_n - 1/b) \rightarrow \mathcal{N}(0; b^{-2}) \quad (n \rightarrow \infty)$$

if and only if $np_N g^2(N) \rightarrow 0 \quad (n \rightarrow \infty)$, where

$$g(N) = \mathbb{M}\{\ln(X/N) | X \geq N\} - b^{-1}$$

They considered also the situation

$$L(x) = L \cdot (1 + v_x)$$

(B)

$$L = \text{const}, \quad v_x \rightarrow 0 \text{ as } x \rightarrow \infty.$$

In the assumption

$$(2) \quad (\ln N)^2 / np_N \rightarrow 0, \quad g(n)(\ln N) \rightarrow a \in \mathbb{R} \quad (n \rightarrow \infty)$$

they proved that for $n \rightarrow \infty$ there holds

$$(3) \quad L_n^* \equiv n^{-1} N^{1/t_n} S_n \exp(ab^2) \xrightarrow{p} L,$$

where $S_n = \sum_{i=1}^n 1_{\{X_i \geq N\}}$.

The purpose of this article is to generalize results of (1) and (3).

Theorem 1. If $g(N) \rightarrow 0$ as $n \rightarrow \infty$ then

$$(4) \quad t_n \xrightarrow{p} 1/b \quad (n \rightarrow \infty)$$

If $\sqrt{np_N} g(N) \rightarrow A \in \mathbb{R}$ as $n \rightarrow \infty$ then for $n \rightarrow \infty$ we have

$$(5) \quad \left\{ (t_n - 1/b) S_n^{1/2} - A \right\} \cdot t_n^{-1} \Rightarrow \mathcal{N}(0; 1)$$

Corollary. If $\sqrt{np_N} g(N) \rightarrow A \in \mathbb{R}$ as $n \rightarrow \infty$ then for $n \rightarrow \infty$ we have

$$(5') \quad \left\{ t_n^2 (t_n^{-1} - b) S_n^{1/2} + A \right\} \cdot t_n^{-1} \Rightarrow \mathcal{N}(0; 1)$$

Consider now the situation (B).

Theorem 2. Let $\{r_n, n \geq 1\}$ be a sequence of positive numbers such that the following conditions hold for $n \rightarrow \infty$:

- (i) $r_n \rightarrow \infty$
- (ii) $r_n \sqrt{np_N} \rightarrow B \in \mathbb{R}$
- (iii) $b^2 r_n (a - g(N) \cdot \ln N) \rightarrow C \in \mathbb{R} \quad (a \in \mathbb{R})$
- (iv) $(\ln N) (np_N)^{-1/2} r_n \rightarrow c < \infty$
- (v) $r_n g(N) \rightarrow d \in \mathbb{R}$

Then for $n \rightarrow \infty$ we have

$$(6) \quad r_n (L_n/L - 1) \Rightarrow \mathcal{N}(B+C+D; b^2 c^2),$$

where $D = -2ab^3 d$, $L_n = n^{-1} N^{1/t_n} S_n \exp(a/t_n^2)$.

Note that assumption (2) suffices (i)-(v) .

Let us consider special cases

$$(a) \quad v_x = O(e^{-\beta x}) \quad (\beta > 0)$$

$$(b) \quad v_x = O(x^{-\beta}) \quad (\beta > 0)$$

$$(c) \quad v_x = O((\ln x)^{-\beta}) \quad (\beta > 1)$$

In the light of (8) it is easy to see that rate of convergence of $g(N)$ and v_N to zero as $N \rightarrow \infty$ is one and the same. Conditions of theorem 1 will be fulfilled for $r_n \times \min\{1/g(N); \sqrt{np_N}\}/(\ln N)$. Hence, convergence (6) holds with the following rate of r_n :

$$(a) \quad \{n(\ln n)^{-b}\}^{1/2}/(\ln \ln n) \quad N(n) = \frac{(\ln n - b \ln \ln n)}{2\beta} + O(1)$$

$$(b) \quad n^{\beta/(b+2\beta)}/(\ln n) \quad N(n) \times n^{1/(b+2\beta)}$$

$$(c) \quad (\ln n)^{\beta-1} \quad N(n) \times (n(\ln n)^{-2\beta})^{1/b}$$

PROOFS

Proof of theorem 1. Let

$$Y_i = 1_{\{x_i \geq N\}} \quad , \quad Z_i = Y_i \ln(X_i/N)$$

There follows from the definition (A) that

$$M \exp(itZ_1) = 1 + \int_0^\infty (e^{itz} - 1) P(Z_1 \in dz)$$

$$(7) \quad \begin{aligned} &= 1 + it \int_0^{\infty} e^{ity} \mathbb{P}(Z_1 \geq y) dy \\ &= 1 + it p_N (b-it)^{-1} + it p_N g(N, t) \end{aligned}$$

where

$$g(N, t) = \int_0^{\infty} e^{ity} \{L(Ne^y)/L(N) - 1\} e^{-by} dy$$

A remarkable fact is that

$$(8) \quad g(N) = \int_0^{\infty} \{L(Ne^y)/L(N) - 1\} e^{-by} dy$$

Taking into account (8) and properties of slowly varying functions (Seneta, 1976, p.2-6) we derive that for $N \rightarrow \infty$

$$g(N) \rightarrow 0$$

$$(9) \quad |g(N) - g(N, t)| = o(|t|)$$

It is easy to check that

$$(10) \quad \begin{aligned} &\mathbb{M} \exp\{it(S_n - np_N)\} = \\ &= \exp\{np_N(e^{it} - 1 - it) + O(np_N^2 t^2)\} \end{aligned}$$

First assertion of theorem 1 follows from (7), (9), (10).

Let us prove that for $n \rightarrow \infty$

$$(11) \quad \sum_{j=1}^n [Z_j - Y_j b^{-1} - p_N g(N)] (np_N)^{-1/2} \Rightarrow \mathcal{N}(0; b^{-2})$$

It is easy to see that (11) entails (5).

Similarly to (7) we derive

$$(12) \quad \begin{aligned} &\mathbb{M} \exp(it(Z - Yb^{-1})) \\ &= 1 + p_N \left[-t^2/2b^2 + ite^{-it/b} g(N, t) + O(t^3) \right] \end{aligned}$$

The convergence (9) follows from (7) and (12).

Proof of theorem 1 completed.

Remark. It is easy to see that assertions of theorem 1 and corollary remain true if we substitute $g(N)S_n^{1/2}$ for A in (5) and (5').

Proof of theorem 2. There follows from the definition (3) that

$$(13) \quad L_n/L = (np_N)^{-1} S_n \exp(\alpha t_n^{-2} + (t_n^{-1} - b) \cdot \ln N) \cdot (1 + v_N)$$

Hence, it is sufficient to prove that for $n \rightarrow \infty$ there holds

$$\tau_n r_n \ln(L_n/L) \Rightarrow \mathcal{N}(B+C+D; b^2 c^2),$$

where $\tau_n = 1\{S_n \geq np_N/2\}$. Note that

$$(14) \quad \left| \tau_n \ln(S_n / np_N) - \tau_n \sum_{i=1}^n (Y_i - p_N) / np_N \right| \leq 2 \left\{ \sum_{i=1}^n (Y_i - p_N) / np_N \right\}^2$$

There follows from (10), (13), (14), (i), (ii), (iv) that we need only to argue the convergence

$$r_n (\alpha t_n^{-2} + (t_n^{-1} - b) \cdot \ln N) \Rightarrow \mathcal{N}(C+D; b^2 c^2) \quad (n \rightarrow \infty)$$

One derives from (7), (12), (iv)-(v) that for $n \rightarrow \infty$

$$r_n(t_n^{-1} - b) \equiv$$

$$\equiv - \frac{b^2 r_n \sum_{i=1}^n (Z_i - Y_i b^{-1} - g(N) p_N)}{np_N} \cdot \frac{np_N}{\sum_{i=1}^n b Z_i} \xrightarrow{p} -b^2 d$$

Hence, $a r_n(t_n^{-2} - b^2) \xrightarrow{p} D$ as $n \rightarrow \infty$.

Now our purpose is to show that for $n \rightarrow \infty$ there holds

$$b^2 r_n \left\{ a - \frac{\sum_{i=1}^n (Z_i - Y_i b^{-1}) (\ln N)}{\sum_{i=1}^n b Z_i} \right\} \rightarrow$$

$$(15) \quad \rightarrow \mathcal{N}(C; b^2 c^2)$$

Taking into account (7), (12), (iv), (v), one can see that it is sufficient to prove the convergence

$$b^2 r_n \left\{ a - (np_N)^{-1} \sum_{i=1}^n (Z_i - Y_i b^{-1}) (\ln N) \right\} \rightarrow$$

$$(16) \quad \rightarrow \mathcal{N}(C; b^2 c^2) \quad (n \rightarrow \infty)$$

There follows from (12) and (iv) that

$$b^2 r_n (\ln N) \frac{\sum_{i=1}^n (Z_i - Y_i b^{-1} - g(N) p_N)}{np_N} \rightarrow$$

$$(17) \quad \rightarrow \mathcal{N}(0; b^2 c^2) \quad (n \rightarrow \infty)$$

The assertion (16) follows from (17) and (iii).

Theorem 2 is proved.

REFERENCES

- Csorgo M., Horvath L., Révész P. (1987) : On the optimality of estimating the tail index and a naive estimator . Austral. J. Math. 29 , 166 - 178 .
- Csorgo S., Deheuvels P., Mason D.M. (1985) : Kernel estimates of the tail index of the distribution . Ann. Statist. 13 , 1050 - 1077 .
- Hill B.M. (1975) : A simple general approach to inference about the tail index of a distribution. Ann. Statist. 3 , 1163 - 1174 .
- Novak S.Yu., Utev S.A. (1990) : On the asymptotic distribution of the ratio of sums of random variables. Siberian Math. J. 31 , No 5 , 781 - 788 .
- Seneta E. (1976) : Regularly varying functions . Lecture Notes Math. 508 , Springer Verlag, Heidelberg .
- Teugels J.L. (1985) : On the Pareto-type distribution. In: 4th Intern. Vilnius Conf. Probab. Th. Math. Statist., Abstracts Commun. iv , 302 - 304 .

Institute of Geodesy
ul. Plazkhotnogo 10
Novosibirsk 630072
Russia
Novak@niigaik.nsk.su