Transactions 11th Prague Conf. Information Theory, Statistical Decision functions, Random Processes.- Prague, 1992, v. B, pp. 251-258.

INFERENCE ABOUT THE PARETO-TYPE DISTRIBUTION

Sergey Novak

Novosibirsk

Key words: tail index , nonparametric inference.

ABSTRACT

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

INTRODUCTION AND RESULTS

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

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

where (unknown) function $L(x)$ slowly varies as $x\to\infty$; 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,Horváth,Révész (1987). Novak and Utev (1990) investigated the estimator

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

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

(1) $$(np_n)^{1/2}(t_n-1/b) \Rightarrow \mathcal{N}(0;b^{-2}) \qquad (n\to\infty)$$

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

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

They considered also the situation

(B) $$L(x)=L_\circ(1+v_x) , \qquad L\equiv\text{const}, \quad v_x\to 0 \text{ as } x\to\infty .$$

In the assumption

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

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

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

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

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

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

(4) $$t_n \xrightarrow[p]{} 1/b \qquad (n\to\infty)$$

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

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

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

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

Consider now the situation $(\mathbb{B})$.

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

(i) $r_n \to \infty$

(ii) $r_n v_N \to B\in\mathbb{R}$

(iii) $b^2 r_n(a-g(N)\cdot\ln N) \to C\in\mathbb{R}$ $\qquad (a\in\mathbb{R})$

(iv) $(\ln N)(np_N)^{-1/2} r_n \to c<\infty$

(v) $r_n g(N) \to d\in\mathbb{R}$

Then for $n\to\infty$ we have

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

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

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

Let us consider special cases

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

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

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

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

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

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

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

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

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

(7) $$= 1 + it\int_0^{\infty} e^{ity}\mathbb{P}(Z_1 \geq y)dy$$

$$= 1 + itp_N(b-it)^{-1} + itp_N g(N,t) \quad ,$$

where

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

A remarkable fact is that

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

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

$$g(N) \to 0 \ ,$$

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

It is easy to check that

$$\mathbb{M} \exp\{it(S_n - np_N)\} =$$

(10) $$= \exp\{np_N(e^{it} - 1 - it) + O(np_N^2 t^2)\}$$

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

Let us proof that for $n \to \infty$

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

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

Similarly to (7) we derive

(12) $$\mathbb{M} \exp(\ it(Z - Yb^{-1})\ )$$

$$= 1 + p_N\Big[-t^2/2b^2 + ite^{-it/b}g(N,t) + O(t^3)\Big]$$

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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) $$\Rightarrow \mathcal{N}(C;b^2c^2) \qquad (n\to\infty)$$

There follows from (12) and (iv) that

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

(17) $$\Rightarrow \mathcal{N}(0;b^2c^2) \qquad (n\to\infty)$$

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

Theorem 2 is proved.

REFERENCES

Csorgo M., Horvath L., Revesz 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