THE REGION OF ASYMPTOTIC LOCATION INVARIANCE FOR A GROUP ESTIMATOR

Agnė Jurevičiūtė, Marijus Vaičiulis
Vilnius University, Faculty of Mathematics and Informatics,
Šiauliai University, Faculty of Mathematics and Informatics

1. Introduction

A probability distribution function F is said to be heavy-tailed with tail index $\alpha > 0$ if for every $x > 0$,

$$\lim_{t\to\infty}\frac{1-F(tx)}{1-F(t)} = x^{-\alpha}. \quad (1)$$

The applications of the heavy-tailed distributions may be found in many fields, such as insurance, finance, climatology and environmental science. The problem of estimating the heavy-tailed index has been studied extensively by many authors (see e.g. [2] for an extensive review).

Let $X_1, X_2, \ldots, X_N$ be non-negative independent random variables with the common distribution function F, satisfying (1). Davydov, Paulauskas and Račkauskas [1] introduced the so-called group estimator for the parameter $p = \alpha/(\alpha+1)$. This estimator is the subject of research of this paper. To describe it, let us divide the sample $X_1, X_2, \ldots, X_N$ into $[N/m]$ groups $G_1, G_2, \ldots, G_{[N/m]}$, each group containing m random variables, where [*]stands for the integer part. Let $L_{m,k}$ and $\ell_{m,k}$ denote the largest and the second largest element in the group G_k. Then the group estimator has the functional expression

$$\hat{p}_{N,m} = \frac{1}{[N/m]}\sum_{k=1}^{[N/m]}\kappa_{m,k}, \quad \kappa_{m,k} = \frac{\ell_{m,k}}{L_{m,k}}. \quad (2)$$

It is easy to see that the estimator (2) is not location invariant, i.e., it changes when sample $X_1, X_2, \ldots, X_N$ is replaced by $X_1+\gamma, X_2+\gamma, \ldots, X_N+\gamma$, where $\gamma > 0$ is an arbitrary constant. Let us note that if some tail index is location invariant, then it automatically removes a constant-type trend from the data (see [4] for more on location invariant estimators). When a number of observations N is big enough, an important property of location invariance can be changed by an asymptotical location invariance of the tail index estimator. Let us recall that the tail index estimator asymptotically is a location invariant if its weak limit does not change when the sample $X_1, X_2, \ldots, X_N$ is replaced by $X_1+\gamma, X_2+\gamma, \ldots, X_N+\gamma$.

The aim of this paper is to answer the question whether the group estimator location is asymptotically invariant. This question is discussed in Section 2. An additional discussion on the class of group estimators is included in Section 3.

2. Main results

To derive asymptotic results of the group estimator (2), the assumption stronger than (1) is needed. Specifically, we will assume that an upper tail of the distribution function F satisfies the so-called second-order asymptotic relation, there exist constants $C_1 > 0,\ C_2 \in R,\ C_2 \neq 0$ and parameters $0 < \alpha < \beta$ such that

$$1 - F(x) = C_1 x^{-\alpha} + C_2 x^{-\beta} + o(x^{-\beta}), \quad x \to \infty. \quad (3)$$

Let us summarize results in [5] and [6].

Theorem 2.1 *Let $X_1, X_2, \ldots, X_N$ be non-negative independent random variables with the common distribution F satisfying (3). Let $m = m(N)$ be some integers satisfying $2 \le m(N) \le N,\ m(N) \to \infty$ and $N/m(N) \to \infty$ as $N \to \infty$.* (4)

(i) If $\beta < \infty$, assume in addition that there exists constant $0 \le \lambda < \infty$ such that

$$\frac{N}{m_N^{1+2\zeta}} \to \lambda^2,\ (N \to \infty). \quad (5)$$

Then

$$\sqrt{N/m}(\hat{p}_{N,m} - p) \Rightarrow \mathrm{N}(a\lambda, \sigma^2), \quad (N \to \infty). \quad (6)$$

where $\Rightarrow$ denotes convergence in distribution, N is normal random variable and

$$\zeta = \frac{\beta-\alpha}{\alpha},\quad a = \frac{C_2\beta\zeta\,\Gamma(\zeta+1)}{C_1^{\zeta+1}(\alpha+1)(\beta+1)},$$

$$\sigma^2 = \frac{\alpha}{(\alpha+1)^2(\alpha+2)}.$$

(ii) *If $\beta = \infty$, then*

$$\sqrt{N/m}(\hat{p}_{N,m} - p) \Rightarrow \mathrm{N}(0, \sigma^2), \quad (N \to \infty). \quad (7)$$

where σ^2 is the same as in (6).

Let us briefly discuss assumptions on

$m = m(N)$. The assumption (4) ensures convergence $\sqrt{N/m}(\hat{p}_{N,m} - \mathrm{E}\hat{p}_{N,m}) \Rightarrow \mathrm{N}(0,\sigma^2)$, while the assumption (5) allows us to control behaviour of the bias $\mathrm{E}\hat{p}_{N,m} - p$. If (5) is satisfied, we have $\sqrt{N/m}(\mathrm{E}\hat{p}_{N,m} - p) \to a\lambda$ as $N \to \infty$. The additional assumption $\beta = \infty$ in (3) gives Paretian case, i.e., in this case $X_1, X_2, \ldots, X_N$ are independent random variables with a distribution function

$$F(x) = 1 - C_1 x^{-\alpha}, \quad x \ge C_1^{1/\alpha}. \tag{8}$$

We note that the group estimator (2) is unbiased in Paretian case and thus the assumption (5) is not needed.

Let $p_{N,m}^{(\gamma)}$ denote the estimator $p_{N,m}$, for which observations $X_1, X_2, \ldots, X_N$ are replaced by $X_1 + \gamma, X_2 + \gamma, \ldots, X_N + \gamma$, where $\gamma > 0$. Our main result is a direct consequence of the Theorem 2.1.

Theorem 2.2 *Let $X_1, X_2, \ldots, X_N$ be non-negative independent random variables with the common distribution function F satisfying (3). Assume that $\gamma > 0$ satisfies relation $C_1\alpha\gamma \ne -C_2$ in the case when $\beta < \infty$, $\alpha + 1 = \beta$. Let $m = m(N)$ satisfies (4) and let the assumption (5) is satisfied with*

$$\zeta = \begin{cases} (\beta - \alpha)/\alpha, & \beta < \infty \text{ and } \alpha + 1 > \beta, \\ 1/\alpha, & \text{in other cases.} \end{cases} \tag{9}$$

Then

$$\sqrt{N/m}(\hat{p}_{N,m}^{(\gamma)} - p) \Rightarrow \mathrm{N}(a\lambda, \sigma^2) \quad (N \to \infty), \tag{10}$$

where σ^2 is the same as in Theorem 2.1. In the case $\beta < \infty$, $\alpha + 1 > \beta$, a constant a is the same as in Theorem 2.1. In the rest of cases it has the following form:

$$a = \begin{cases} \dfrac{(C_1\alpha\gamma + C_2)\Gamma(1/\alpha + 1)}{C_1^{1/\alpha+1}\alpha(\alpha+2)}, & \beta < \infty \text{ and } \alpha + 1 = \beta, \\ \dfrac{\gamma\Gamma(1/\alpha + 1)}{C_1^{1/\alpha}(\alpha+2)}, & \beta < \infty \text{ and } \alpha + 1 < \beta \text{ or } \beta = \infty. \end{cases} \tag{11}$$

Proof of Theorem 2.2 For application of the Theorem 2.1 we need to find asymptote of the upper tail $\mathrm{P}(X_k + \gamma > x)$. If $\beta < \infty$, use an asymptotic binomial expansion

$$(1 - \gamma/x)^{-\delta} = 1 + \gamma\delta x^{-1} + o(x^{-1}),$$

$$x \to \infty, \quad \delta > 0 \tag{12}$$

to get

$$\mathrm{P}(X_k + \gamma > x) \sim C_1 x^{-\alpha} + \{C_1\alpha\gamma x^{-\alpha-1} + C_2 x^{-\beta}\} + \{o(x^{-\alpha-1}) + o(x^{-\beta})\},$$

as $x \to \infty$. Whence follows

$$\mathrm{P}(X_k + \gamma > x) \sim C_1 x^{-\alpha} + \begin{cases} C_2 x^{-\beta} + o(x^{-\beta}), & \alpha + 1 > \beta, \\ (C_1\alpha\gamma + C_2)x^{-\alpha-1} + o(x^{-\alpha-1}), & \alpha + 1 = \beta,\ C_1\alpha\gamma \ne -C_2, \\ C_1\alpha\gamma x^{-\alpha-1} + o(x^{-\alpha-1}), & \alpha + 1 < \beta. \end{cases}$$

If X_k is the Paretian random variable with the distribution function (8), then we have

$$\mathrm{P}(X_k + \gamma > x) \sim C_1 x^{-\alpha} + C_1\alpha\gamma x^{-\alpha-1} + o(x^{-\alpha-1}),$$

$x \to \infty$.

Now assertions of the Theorem 2.2 follow easily. End of the proof.

Corollary 2.3 *An asymptotic location invariance of group estimator (2) depends on the distributional assumption (3) and the sequence m(N). In particular, if m(N) is such that the relation (5) is satisfied with $\lambda = 0$ and ζ given in (9), then the group estimator (2) is asymptotically location invariant. If $\lambda > 0$ in (5), then the group estimator (2) is asymptotically location invariant in the case $\beta < \infty$, $\alpha + 1 > \beta$ only.*

By the Corollary 2.3 a statistician should apply a group estimator with care –existence of a constant-type trend (or more complicated trend) can significantly aggravate estimation of a tail index of the heavy-tailed distribution. It is not easy to combine detrending and tail index estimation, but we intend to do this in the near future.

3. Comparison of the estimators $p_{N,m}$ and $p_{N,m}^{(\gamma)}$

The estimator $p_{N,m}^{(\gamma)}$, $\gamma > 0$, having a form

$$\hat{p}_{N,m}^{(\gamma)} = \frac{1}{[N/m]} \sum_{k=1}^{[N/m]} \kappa_{m,k}, \quad \kappa_{m,k} = \frac{\gamma + \ell_{m,k}}{\gamma + L_{m,k}}, \tag{13}$$

can be interpreted as an estimator of the parameter $p = \alpha / (\alpha + 1)$with tuning parameter γ, which must be chosen by a statistician.

Now we proceed to an asymptotic comparison of the estimators $p_{N,m}$ and $p_{N,m}^{(\gamma)}$ following the way proposed in [3]. In particular, from the sequences $m(N)$, $N \in \mathrm{N}$, satisfying (4) (and (5) in case $\beta < \infty$) we will choose such $m(N)$ (it is called optimal) that the asymptotic second moment of $\hat{p}_{N,m} - p$ is minimal. We will also choose another $m(N)$ satisfying assumptions (4) and (5) with ζ in (9), such that the asymptotic second moment of $\hat{p}^{(\gamma)}_{N,m} - p$ is minimal. Following de Haan and Peng [3], we will say that the estimator $p_{N,m}$ dominates the estimator $p^{(\gamma)}_{N,m}$ if

$$\lim_{N\to\infty} \frac{\mathrm{E}(\hat{p}_{N,m} - p)^2}{\mathrm{E}(\hat{p}^{(\gamma)}_{N,m} - p)^2} < 1. \tag{14}$$

It is known that taking

$$m(N) = \left(\frac{2\zeta\, a^2}{\sigma^2}\right)^{1/(1+2\zeta)} N^{1/(1+2\zeta)} \tag{15}$$

we get that a mean square error is minimal

$$\mathrm{E}(\hat{p}_{N,m} - p)^2 \sim \{(2\zeta)^{-2\zeta/(1+2\zeta)} +$$

$$+ (2\zeta)^{1/(1+2\zeta)}\} \left(\frac{a^2\sigma^{4\zeta}}{N^{2\zeta}}\right)^{1/(1+2\zeta)}, \quad N \to \infty, \tag{16}$$

see [5].

Since in the case $\beta < \infty$, $\alpha + 1 > \beta$ the estimators $p_{N,m}$ and $p^{(\gamma)}_{N,m}$ have the same limit law (see Theorems 2.1-2.2), the relation (16) holds true if we replace $\mathrm{E}(\hat{p}_{N,m} - p)^2$ by $\mathrm{E}(\hat{p}^{(\gamma)}_{N,m} - p)^2$. Thus, the limit in l.h.s. of (14) is equal to 1 and none of the estimators dominates the other.

Let now $\beta < \infty$, $\alpha + 1 = \beta$. Combining Theorem 2.2 with (16) we get that the minimal second moment of $\hat{p}^{(\gamma)}_{N,m} - p$ is asymptotically equal to

$$\{(2/\alpha)^{-1/(\alpha+2)} + (2/\alpha)^{\alpha/(\alpha+2)}\} \left(\frac{a^2\sigma^{4/\alpha}}{N^{2/\alpha}}\right)^{\alpha/(\alpha+2)},$$

$N \to \infty$, where a is given in (11). Then we have

$$\lim_{N\to\infty} \frac{\mathrm{E}(\hat{p}_{N,m} - p)^2}{\mathrm{E}(\hat{p}^{(\gamma)}_{N,m} - p)^2} = \left|\frac{C_2}{C_1\alpha\gamma + C_2}\right|^{2\alpha/(\alpha+2)}.$$

Whence follows that in this case none of the estimators dominates the other: it depends on the parameters C_1, C_2, α and γ.

Next, let $\beta < \infty$, $\alpha + 1 < \beta$. One can find that the minimal second moment of $\hat{p}^{(\gamma)}_{N,m} - p$ is asymptotically equal to $\tilde{C}N^{-2/(\alpha+2)}$, where $\tilde{C}$ is some positive constant. From

$$\lim_{N\to\infty} \frac{N^{-2\zeta/(1+2\zeta)}}{N^{-2/(\alpha+2)}} = \lim_{N\to\infty} N^{-2\alpha(\beta-\alpha-1)(2\beta-\alpha)^{-1}(\alpha+2)^{-1}} = 0$$

we deduce that the estimator $p_{N,m}$ dominates the estimator $p^{(\gamma)}_{N,m}$.

Similarly, one can draw a conclusion that in the Paretian case the estimator $p_{N,m}$ dominates the estimator $p^{(\gamma)}_{N,m}$ too. Moreover, Paulauskas in [5] obtained that claim (7) holds true without the assumption $m = m(N)$ and $\mathrm{E}(\hat{p}_{N,m} - p)^2$ is minimal for $m = 2$. Thus, a minimal second moment $\mathrm{E}(\hat{p}_{N,m} - p)^2$ tends to zero meaningfully faster than a minimal second order moment $\mathrm{E}(\hat{p}^{(\gamma)}_{N,m} - p)^2$.

Our final Corollary gives conclusion on the applicability of the estimator $p^{(\gamma)}_{N,m}$.

Corollary 3.1 *The estimator* $p^{(\gamma)}_{N,m}$ *does not outperform the estimator* $p_{N,m}$ *in the general case. Moreover, the estimator* $p^{(\gamma)}_{N,m}$ *contains a tuning parameter* $\gamma > 0$, *the choice of which is probably complicated. These two reasons allow us to assert that the estimator* $p^{(\gamma)}_{N,m}$ *is non-advisable in the applications.*

References

1. Davydov Yu., Paulauskas V., Račkauskas A., 2000, More on P-stable convex sets in Banach spaces. *Journal of Theoretical Prob.* No. 13 (1). P. 39–64.
2. Gomes M. I., Canto E Castro L., Fraga Alves M. I., Pestana D., 2008, Statistics of Extremes for IID Data: Laurens de Haan leading contributions. *Extremes*. No. 11 (1). P. 3–34.
3. L. de Haan, Peng L., 1998, Comparison of tail index estimators. *Statistica Neerlandica*. No. 52 (1). P. 60–70.
4. L. J., Peng Z., Nadarajah S., 2008, A class of unbiased location invariant Hill-type estimators for heavy tailed distributions. *Electronic Journal of Statistics*. No. 2. P. 829–847.
5. Paulauskas V., 2003, A new estimator for tail index. *Acta Appl. Math.* No. 79. P. 55–67.
6. Paulauskas V., Vaičiulis M., 2009, Once more on comparison of tail estimators. *In print*.
7. Peng L., Qi Y., 2006, A new calibration method of constructing empirical likelihood-based confidence intervals for the tail index. *Aust. N. Z. J. Stat.* No. 48 (1). P. 59–66.

THE REGION OF ASYMPTOTIC LOCATION INVARIANCE FOR A GROUP ESTIMATOR

Agnė Jurevičiūtė, Marijus Vaičiulis

Summary

This paper is devoted to estimation of the tail index of heavy-tailed distributions. We consider the situation when initial i.i.d. data is perturbed by a constant type positive trend. Our main message is that such kind of trend should be removed before estimation of the tail by using group estimator because this estimator asymptotically is not location invariant in general.
Keywords: group estimator, tail index, heavy-tailed distributions.

GRUPINIO ĮVERČIO ASIMPTOTINIO VIETOS INVARIANTIŠKUMO SRITIS

Agnė Jurevičiūtė, Marijus Vaičiulis

Santrauka

Šiuo straipsniu siekiama įvertinti skirstinių su dideliais nuokrypiais nuokrypio indeksą. Nagrinėjame situaciją, kai pirminiai i.i.d. duomenys trikdomi pastoviojo tipo teigiamojo trendo. Pagrindinis mūsų teiginys: tokios rūšies trendą reikėtų pašalinti prieš vertinant nuokrypį naudojant grupinį įvertį, kadangi šis įvertis asimptotiškai apskritai nėra nekintamas vietos atžvilgiu.
Prasminiai žodžiai: grupinis įvertis, nuokrypio indeksas, skirstiniai su dideliais nuokrypiais.

Įteikta 2009-09-03