PLISKA

STUDIA MATHEMATICA BULGARICA

ПЛИСКА

БЪЛГАРСКИ МАТЕМАТИЧЕСКИ СТУДИИ

Pliska Stud. Math. Bulgar. **22** (2013), 89–108

PLISKA
STUDIA MATHEMATICA
BULGARICA

ON ESTIMATION AND TESTING FOR PARETO TAILS

Pavlina Jordanova*, Milan Stehlík, Zdeněk Fabián†,
Luboš Střelec

The t-Hill estimator for independent data was introduced by Fabian and Stehlik (2009). It estimates the extreme value index of distribution function with regularly varying tail. This paper considers sampling of an infinite moving average model. We prove that in the discussed case the t-Hill estimator is weak consistent. However, in contrast to independent identically distributed case here it is shown that the t-Hill and the Hill estimator applied to the moving average model are not robust with respect to large observations.

1. Introduction. Here we suppose that $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n$ are possibly dependent copies of $\mathbf{X}$ with d.f. F, upper order statistics

$$\mathbf{X}_{(1,n)} \leq \mathbf{X}_{(2,n)} \leq \cdots \leq \mathbf{X}_{(n,n)}$$

and

$$1 - F \in RV_{-\alpha}, \quad \alpha > 0. \tag{1}$$

We consider the t-Hill estimator of α^{-1}

$$H^*_{k,n} = \frac{1}{\widehat{\alpha}_{k,n}} = \left\{ \frac{1}{k} \sum_{i=1}^{k} \frac{\mathbf{X}_{(n-k,n)}}{\mathbf{X}_{(n-i+1,n)}} \right\}^{-1} - 1, \quad k = 1, 2, \ldots, n-1. \tag{2}$$

*The first author is supported by Grant No. RD-05-278/15.03.2012 of Shumen University, Bulgaria.

†Zdenek Fabian acknowledges the institutional support RVO: 67985807.

2010 *Mathematics Subject Classification:* 62F10, 62F12.

Key words: Point estimation, asymptotic properties of estimators, testing against heavy tails.

Hill (1975) derived a procedure of Pareto tail estimation by the MLE. He obtain the following Hill estimator

$$H_{k,n} = \frac{1}{\widehat{\alpha}_{k,n}} = \frac{1}{k}\sum_{i=1}^{k} \log\left\{\frac{\mathbf{X}_{(n-i+1,n)}}{\mathbf{X}_{(n-k,n)}}\right\}, \quad k = 1, 2, \ldots, n-1. \tag{3}$$

Later on, many authors tried to robustify the Hill estimator, but they still rely on maximum likelihood. Fraga Alves (2001) has introduced a new lower bound for sample fraction and studied its properties. Gomes and Oliveira (2003) and Li, Peng and Nadarajah (2010) introduce powers of original statistics. However, the influence function of Hill estimator is slowly increasing, but unbounded, therefore the Hill procedure is no robust. Many authors tried to make the original Hill estimator robust (see e.g. Beran and Schell (2010) and Vandewalle, Beirlant, Christmann and Hubert (2007)). In Fabián (2001) a new score method of score moment estimators has been proposed. It appeared that these score moment estimators are robust for very heavy tailed distributions (see Stehlík, Potocký, Waldl and Fabián (2010)). Jordanova and Pancheva (2012) consider i.i.d. sample and find the limit distribution of the t-Hill estimator for fixed number k of the threshold order statistic. They prove that for Pareto distributed observations we do not need large sample to have the corresponding limit distribution for fixed k. In that case under suitable normalizations and large sample the t-Hill estimator is asymptotically normal for $k(n) \to \infty$. Under the more general conditions, the t-Hill estimator is asymptotically normal for $k(n) = o(n)$. The Hill estimator procedure with the score moment estimator has been investigated in Stehlík, Fabián and Střelec (2012) for optimal testing for normality against Pareto tail.

A specific problem is finding of an optimal threshold k, say, yielding a trade off in between of variance and bias of the Hill estimator. Simulation results by Embrechts et al. (1997) showed that the Hill estimator and its alternatives work well over large ranges of values for k in the case of Pareto distribution. However, Hill estimator is often giving wrong results for distributions different from the Pareto one. Their "Hill horror plots" actually show deviations of the Hill estimates trending farther away from the true value of the tail index as k is increased.

When the mean square error (MSE) is employed to study the quality of estimation, then we are getting a bath-tube shape of MSE against threshold k, since higher order statistics did not see the different underlying distribution, however, first order statistics are very different (e.g. Frechet distribution as a Pareto tail distribution, see Gomes and Oliveira (2003)).

Resnick and Starica (1993) generalize the Hill estimator for a more general settings with possibly dependent data. In this paper we continue these investigations. We obtain weak consistency of the t-Hill estimator for a special class of dependent data, the infinite moving average model. In the end of the paper we give some examples showing that in contrast to independent identically distributed case the t-Hill and the Hill estimator applied to the moving average model are not robust with respect to large observations. Since the score moment estimator is simple, it is easy to implement it to the Hill procedure. Under the concept of the Hill estimator we understand the successive averaging of ordered values up to given k. In this paper we understand *"The Hill estimator"* as a specific procedure for studying the tail of Pareto like distribution. Instead of implementing "The Hill estimator" procedure, we implement the score moment procedure. We illustrate t-Hill and we also quantify the robustness and compare efficiency with other competitors. The paper is organized as follows. In the next section we recall the theory of scalar score. In section 3 we discuss the t-Hill estimator, introduced firstly by Fabián and Stehlík (2009) and Stehlík, Fabián and Střelec (2012). The next section concerns the moving average model. Here we show that under certain conditions the t-Hill estimator is weakly consistent. In section 5 we introduce the t-Hill plot and compare t-Hill and Hill estimators. Therein contamination of underling data is controlled by means of score variance of Pareto distribution. Comparisons show that t-Hill estimator outperforms Hill estimator. We end with powers of selected tests for normality against Pareto distribution.

2. Scalar score. Fabian and Stehlik (2009) and Fabian (2011) have introduced a simple scalar inference function, called a scalar score, which reflects main features of a continuous probability distribution. Its simplicity has made it possible to introduce new relevant numerical characteristics of continuous distributions.

The scalar score has been introduced in three steps.

i) Let $\mathcal{X}$ be support of distribution F with density f, continuously differentiable according to $x \in \mathcal{X}$ and let $\eta : \mathcal{X} \to \mathbb{R}$. The *transformation-based score* or shortly the *t-score* is defined by

$$T(x) = -\frac{1}{f(x)}\frac{d}{dx}\left(\frac{1}{\eta'(x)}f(x)\right). \tag{4}$$

where η is given for most of distributions by

$$\eta(x) = \begin{cases} x & \text{if} \quad \mathcal{X} = \mathbb{R} \\ \log(x-a) & \text{if} \quad \mathcal{X} = (a,\infty) \\ \log\dfrac{x}{1-x} & \text{if} \quad \mathcal{X} = (0,1). \end{cases} \tag{5}$$

Then (4) expresses a relative change of a 'basic component of the density', the density divided by the Jacobian of mapping (5).

ii) As a measure of central tendency of distribution $F(x)$ has been suggested the zero of the t-score,

$$x^* : T(x) = 0,$$

called the *transformation-based mean* or shortly the *t-mean*.

iii) The function

$$S(x;\theta) = \eta'(x^*)T(x;\theta), \tag{6}$$

called *scalar score*, has been suggested as a scalar inference function of distribution F.

For a particular class of distributions with support $\mathbb{R}$ and location parameter μ (6) is identical with the score function of distribution F for μ. For a particular class of distributions with "partial" support $\mathcal{X} \neq \mathbb{R}$ and "transformed location parameter" $\tau = \eta^{-1}(\mu)$ (distributions on $\mathcal{X} \neq \mathbb{R}$ are taken as transformed "prototypes" with support $\mathbb{R}$"), (6) was proved to be identical with the score function of distribution F for this parameter. For other distributions, (6) is a new function. The t-mean of distributions with support $\mathbb{R}$ is the mode (if the distribution has the location parameter, the t-mean is its value), the t-mean of distributions with partial support is the transformed mode of the prototype.

Instead of the ordinary moments, the score moments were introduced for any $k \in \mathcal{N}$ by relation

$$M_k(\theta) = ES^k = \int_{\mathcal{X}} S(x;\theta)^k f(x;\theta)\, dx, \tag{7}$$

existing if f satisfies the usual regularity requirements. It appeares that the score moments are often expressed by elementary functions of parameters. $M_1 = 0$. The value $M_2 = ES^2$ of location and transformed location distributions is, respectively, Fisher information for the location and transformed location parameter. Accordingly, ES^2 is the Fisher information for the t-mean. The reciprocal value

$$\omega^2 = \frac{1}{ES^2}, \tag{8}$$

the score variance, appeared to be a natural measure of the variability (dispersion) of the distribution even in cases in which the usual variance does not exist.

For parametric distributions with vector parameter θ, $x^* = x^*(\theta)$. Given data $x_1, \ldots, x_n$ and a model family $\{F_\theta, \theta \in \Theta\}$, the sample characteristics of central tendency ("center") and dispersion (square of the "radius") can be obtained as functions of the estimated parameters: the sample t-mean $\hat{x}^* = x^*(\hat{\theta})$ and the sample score variance $\hat{\omega}^2 = \omega^2(\hat{\theta})$. Estimates $\hat{\theta}$ of θ are usually the maximum likelihood estimates or some robust M-estimates in cases of heavy-tailed distributions or if considering gross errors models. We introduced the score moment estimate as the solution of equations

$$\text{(9)} \qquad \hat{\theta}_{SM}: \qquad \frac{1}{n}\sum_{i=1}^{n} S^k(x_i;\theta) = \mathrm{E}_\theta S^k, \qquad k = 1,\dots,m,$$

derived from (7) using the substitution principle.

In Fabián Z. (2007) and (2008) it is shown that $\hat{x}^*$ is consistent and asymptotically normal. The t-score moment estimators take into account the assumed form of the distribution, similarly as the maximum likelihood (ML) ones. However, since x_i enters into estimation equations by means of $S(x_i;\theta)$ only and scalar scores of heavy-tailed distributions are bounded, the score moment estimates are in cases of heavy-tailed distributions robust, or, in other words, the t-score estimates of all parameters of heavy-tailed distributions are protected against outliers.

3. t-Hill estimator in case of Pareto distribution. In some cases, the first equation of (9) has a form

$$\text{(10)} \qquad \hat{x}^*_{SM}: \qquad \sum_{i=1}^{n} S(x_i;x^*) = 0.$$

This is the case of the Pareto distribution $P(\alpha)$ with support $\mathcal{X} = [1,\infty)$ and density

$$f(x) = \frac{\alpha}{x^{\alpha+1}}.$$

Using the mapping $\eta = \log(x-1)$, $\eta'(x) = 1/(x-1)$, the t-score (4) is

$$T(x) = -1 - (x-1)f'(x)/f(x) = \alpha(1 - x^*/x)$$

where the t-mean $x^* = (\alpha+1)/\alpha$. From (10), $\hat{x}^* = \bar{x}_H$, where $\bar{x}_H = n\Big/\sum_1^n 1/x_i$ is the harmonic mean, and

$$\hat{\alpha} = 1/(\hat{x}^* - 1).$$

It suggests to introduce a variant of the Hill estimator as

$$\hat{\gamma}_k = \frac{1}{\hat{\alpha}_k} = H^*_{k,n} = \frac{1}{\frac{1}{k}\sum_{i=1}^{k}\frac{X_{(n-k,n)}}{X_{(n-i+1,n)}}} - 1, \tag{11}$$

where harmonic mean is taken from the last k observed values with threshold $X_{n-k,n}$.

Let us call the estimator (11) t-Hill estimator. Since it is based on harmonic mean, it is expected to be to a certain extent resistant to large observations so that it could yield more realistic values than the ordinary Hill estimator.

4. The Moving Average Model. Now we apply the following result about weak consistency of the t-Hill estimator, obtained in Jordanova, Dusek and Stehlik (2012).

Theorem 1. Suppose that $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n$ are possibly dependent copies of $\mathbf{X}$ with d.f. F, upper order statistics

$$\mathbf{X}_{(1,n)} \leq \mathbf{X}_{(2,n)} \leq \cdots \leq \mathbf{X}_{(n,n)}$$

and F satisfy condition (1). Let

$$\mu_{X,k(n),n}(\cdot) := \frac{1}{k(n)}\sum_{i=1}^{n} \varepsilon\left\{\frac{\mathbf{X}_i}{b\left(\frac{n}{k(n)}\right)} \in (\cdot)\right\},$$

be a random element in the space $\mathbb{E}^+$ of positive Radon measures on $(0, \infty]$ endowed with the vague topology,

$$b(t) := F^{\leftarrow}(1 - \frac{1}{t}) = \left(\frac{1}{\overline{F}}\right)^{\leftarrow}(t)$$

and $\mu_{X,k(n),n} \Longrightarrow \mu, \quad n \to \infty$, where $\mu : \sigma((0,\infty]) \to [0, \infty)$ and $\mu(x; \infty] = x^{-\alpha}$, $x > 0$. Assume the following Mason's condition

$$\frac{k(n)}{n} \to 0, \quad k(n) \to \infty, \quad n \to \infty$$

holds. Then $H^*_{k,n}$ is an weakly consistent estimator for $\dfrac{1}{\alpha}$.

In a particular case of infinite moving average sequence this theorem looks in the following way.

Theorem 2. Suppose at least one of the real numbers $c_j, j = 0, 1, \ldots$ is positive and there exists $\delta \in (0, 1)$, $\delta < \alpha$ such that

$$\sum_{j=0}^{\infty} |c_j|^{\delta} < \infty. \tag{12}$$

Consider the moving average sequence

$$\mathbf{X}_n = \sum_{j=0}^{\infty} c_j \mathbf{Z}_{n-j}, \quad -\infty < n < \infty, \tag{13}$$

where $\mathbf{Z}_i, -\infty < i < \infty$, are non-negative independent identically distributed (i.i.d.) innovations with d.f. G, such that $\overline{G} \in RV_{-\alpha}$, $\alpha > 0$.

If

$$k(n) \to \infty, \quad \frac{k(n)}{n} \to 0, \tag{14}$$

then $H^*_{k,n}$ is an weakly consistent estimator for α^{-1}.

P r o o f. Resnick and Starica (1993) show that

$$\mu_{X,k(n),n} \Rightarrow \mu, \quad n \to \infty, \tag{15}$$

in $\mathbb{E}^+$, where $\mu : \sigma((0, \infty]) \to [0, \infty)$ and $\mu(x; \infty] = x^{-\alpha}$, $x > 0$.

The random, variables $\mathbf{X}_n$, $-\infty < n < \infty$, are identically distributed. Cline (1983) proves that under these settings

$$\overline{F}(x) = P\left(\sum_{j=0}^{\infty} c_j \mathbf{Z}_j > x\right) \sim \sum_{j:c_j>0}^{\infty} c_j^{\alpha} \overline{G}(x) \in RV_{-\alpha}.$$

We apply Theorem 1 and complete the proof. $\square$

5. Empirical investigation of the robustness of t-Hill estimator. Comparisons. It seems difficult to compare the mean values and variances of Hill and t-Hill estimators for all distribution functions with regularly varying tails. Their robustness is not defined clearly and investigated theoretically yet. In this section we compare empirically their properties.

Let $n \in \{1, 2, \ldots\}$ be fixed. Analogously to the Hill plot we consider the set of points with coordinates

$$\left(k(n), \frac{1}{\widehat{\alpha}_{k(n),n}}\right), \quad k(n) \in \{1, 2, \ldots, n\}.$$

Further on we call their plot “t-Hill plot”. In each of the following graphs, the dotted line represents the true value of $1/\alpha$.

Note: The score variance of the Pareto distribution is $\omega^2 = (\alpha+2)/\alpha^3$. If $\omega = 1$, we have $1/\alpha = 0.657$. Pareto distribution with score variance ω^2 will be denoted by $P(\omega)$.

We consider two cases.

Case 1. i.i.d. observations. The next investigations refine the conclusions made in Fabián and Stehlík (2009).

a.) Pareto distribution. We simulate $n = 2500$ observations with Pareto d.f. and $\alpha = 0.3, 1$. Then we plot the Hill and t-Hill estimators for $k = 50, 51, \ldots, 2500$ and determine the corresponding 0.05-confidence intervals. To obtain good estimators in each of these cases it is important k to be large and $k < n$. The rate of convergence of the Hill and t-Hill estimators could be seen in Fig. 1.

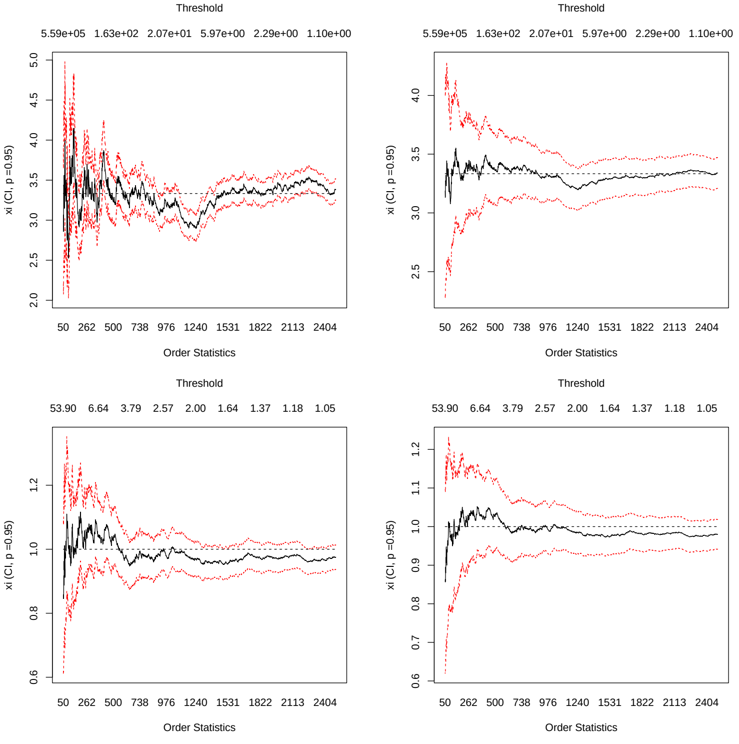

Fig. 1. The t-Hill plot – left and Hill plot – rigth, for $F(x) = 1 - x^{-\alpha}, x > 1$, $\alpha = 0.3$ – above and $\alpha = 1$ below

It is apparent that the t-Hill in his first part oscillates more than the ordinary Hill. The reason is that the t-Hill is sensitive to an abrupt change of the threshold value. It could be, however, suppressed by a suitable smoothing. According to our empirical investigations both estimators behave similarly. The Hill estimator has smaller dispersion than t-Hill estimator and therefore it is slightly better. The rate of convergence of both estimators increases with α.

b.) D.f. with regularly varying tail at very slow rate. The Hill and t-Hill estimators may perform very poorly if the slowly varying function in the tail is faraway from a constant. In (4.16) Embrechts et al. (1997) consider

$$F^{\leftarrow}(p) = (1-p)^{-1/\alpha}(-ln(1-p)), \quad p \in (0,1), \tag{16}$$

with respect to the Hill estimator. We simulated a sample of $n = 10000$ observations of random variables with quantile functions (16) for $\alpha = 0.3, 0.5$ and plotted the Hill and the t-Hill plots for $k(n) = 21, 22, \ldots, 500$. The rate of convergence of these estimators could be observed on Fig. 2.

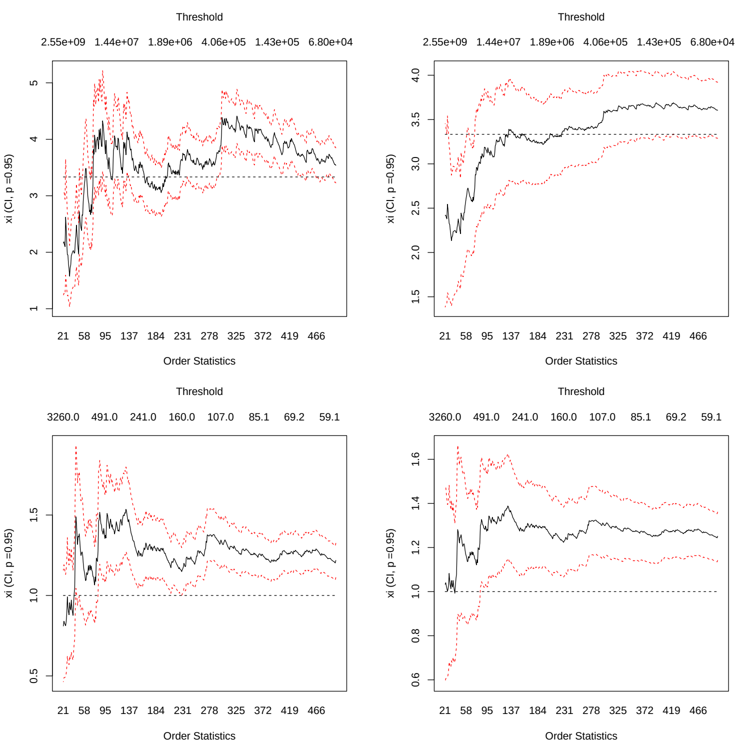

Fig. 2 The t-Hill (left) and Hill plot (right), $\alpha = 0.3$ above and $\alpha = 1$, below

The Hill estimator seems slightly closer to the estimated value, but both estimators are not good for such d.f's.

c.) Log-gamma distribution. The score moment estimator is usually simple so that it makes possible for many distributions to apply a score moment Hill-like estimator. Consider for instance data generated from the log-gamma distribution with support $\mathcal{X} = (1, \infty)$ and density

$$f(z) = \frac{c^\alpha}{\Gamma(\alpha)} (\log z)^{\alpha-1} z^{-(c+1)}. \tag{17}$$

In this case, a simple scalar score is obtained by use of the mapping $\eta : (1, \infty) \to \mathbb{R}$ in the form

$$\eta(x) = \log(\log x).$$

Since $\eta'(x) = 1/(x \log x)$, by (4)

$$T(x) = \frac{1}{f(x)} \frac{d}{dx} \left(-(\log x)^c \alpha x^{-c} \right) = c \log x - \alpha$$

so that the 'loglog' t-mean is $x^* = e^{\alpha/c}$. As the 'second log-log moment' $ET^2 = E[c^2 \log^2(x/x^*)] = \alpha$, the estimation equations (9) are

$$\sum_{i=1}^{n} c \log x_i - \alpha \quad = \quad 0 \tag{18}$$

$$\sum_{i=1}^{n} (c \log x_i - \alpha)^2 \quad = \quad \alpha \tag{19}$$

By setting $\hat{s}_1 = \frac{1}{k} \sum_{i=1}^{k} \log x_i$ and $\hat{s}_2 = \frac{1}{k} \sum_{i=1}^{k} \log^2 x_i$, it follows from (18) that the estimates $\hat{\alpha}$ and $\hat{c}$ correspondingly of α and c are $\hat{\alpha} = \hat{s}_1 \hat{c}$ and from (19) $\hat{c}(\hat{s}_2 - \hat{s}_1^2) = \hat{s}_1$. So that the Hill-like estimate of the tail index (cf. Beirlant, Goegebeur, Segers and Teugels, (2005)) is given by closed-form expression

$$\hat{\gamma}_k = \frac{1}{\hat{\alpha}_k} = \frac{\hat{s}_2}{\hat{s}_1^2} - 1.$$

The Hill-like estimates based on log-gamma distribution is given on Fig. 3 and Fig. 4. It is apparent that the log-gamma hill-like estimator estimates the tail index properly whereas both t-Hill and Hill, expecting heavier Pareto tail, show systematic decrease.

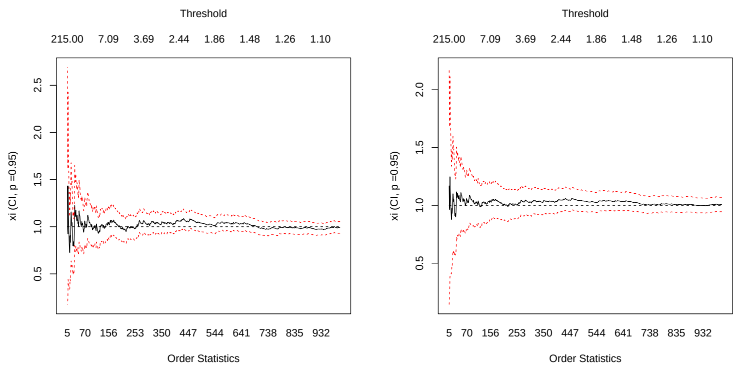

Fig. 3. T-Hill (left), Hill (right) based on log-gamma distribution

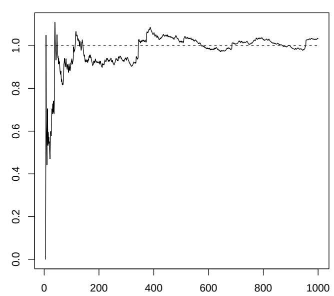

Fig. 4. Hill-like estimates based on log-gamma distribution

d.) Contaminated data from Pareto sample. Usually the data are contaminated in practice. See e.g. Fabian and Stehlik (2009). Frequently there are outliers in the right tail of the distribution. It is known that the Hill estimator is not robust. This is due to the fact that in its formula, the data are entered by their logarithm. What about the t-Hill estimator? In (2) the data are involved by their reciprocal value. We could compare the charts of $y(x) = 1/x$ and $y(x) = ln(x)$ and to deduce that the t-Hill estimators are more robust with respect to large values than the Hill estimators. They are sensitive to the center of the distribution.

We simulated $n = 2500$ data of Pareto(α) distribution with probability $1 - \epsilon = 0.9$, contaminated with Pareto(δ) distribution with probability $\epsilon = 0.1$ for different $\alpha = 1, 1.7$ and $\delta = 0.5, 1$, i.e.

$$F(x) = (1-\epsilon)F_{Patero(\alpha)}(x) + \epsilon F_{Patero(\delta)}(x), \quad x \in \mathbb{R}.$$

Then we plotted the Hill and t-Hill estimators for $k(n) = 1, 2, \ldots, n$, together with the corresponding 0.05-confidence intervals. The Hill estimator has more narrow confidence intervals than t-Hill estimator. To reach a good estimation for $1/\alpha$ we need $k(n)$ to be large and $k(n) < n$. When $\alpha < 1$ and $\delta \geq 1$ the Hill estimators are better than the t-Hill estimators. When $\alpha < 1$ and $\delta < 1$, or $\alpha \geq 1$ and $\delta > 1$ the t-Hill estimators are comparable to the corresponding Hill estimators. In view of Figs 5–7 we can conclude that, for $\alpha \geq 1$ and $\delta \leq 1$, the t-Hill estimators are better than the Hill estimators.

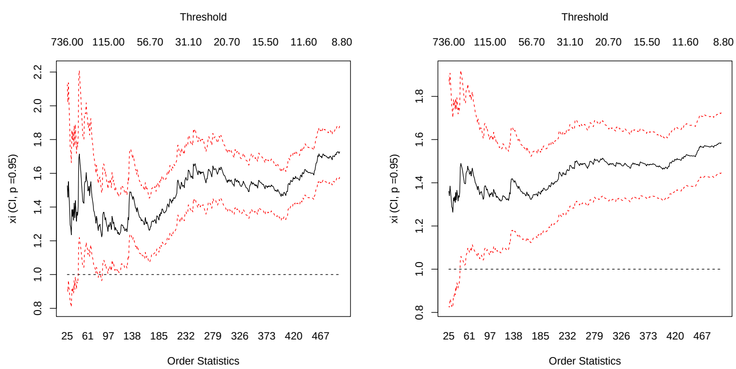

Fig. 5. Rate of convergence of the t-Hill (left) and Hill (right) estimators, $\alpha = 1, \delta = 0.5$

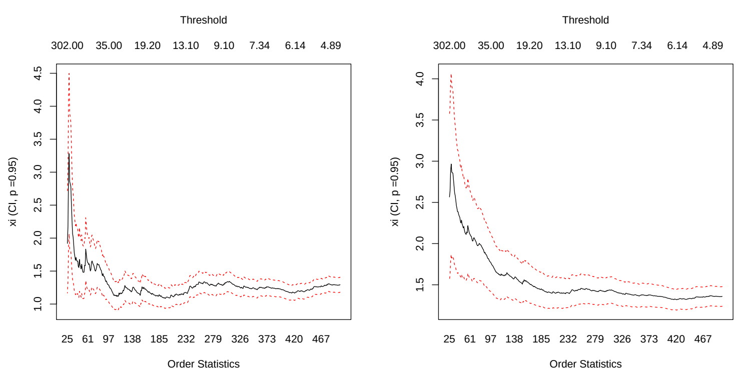

Fig. 6. Rate of convergence of the t-Hill (left) and Hill (right) estimators, $\alpha = 1.7$ and $\delta = 0.5$

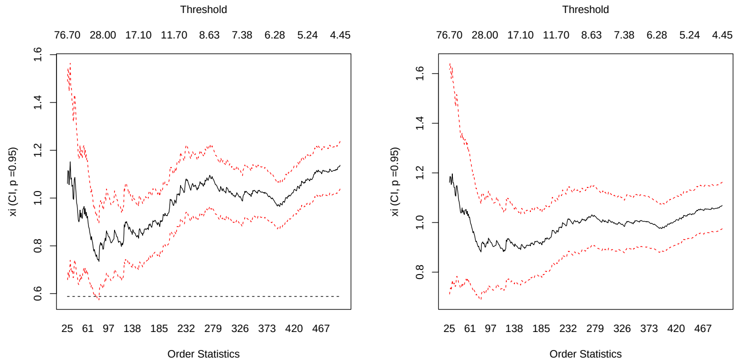

Fig. 7. Rate of convergence of the t-Hill (left) and Hill (right) estimators, $\alpha = 1.7$ and $\delta = 1$

Ratios of $\hat{\gamma}(k)/\gamma(1)$ for $k = 100, 250$ and 900 and $n = 1000$ are given for $H_{k,n}$ and $H^*_{k,n}$ in Table 1 and Table 2.

Table 1. Values $r(k) = H_{k,n}/\gamma(\omega^*) - 1$ and $r^*(k) = H^*_{k,n}/\gamma(\omega^*) - 1$ for P_{cont} with three different ω^* and $\epsilon = 0.05$

ϵ	$\omega*$	$\gamma(\omega*)$	$r(100)$	$r^*(100)$	$r(250)$	$r^*(250)$	$r(900)$	$r^*(900)$
	1.5	0.879	.0363	.0376	.0235	.0187	.0184	.0170
.05	3	1.500	.2027	.1517	.1333	.0982	.0679	.0473
	5	2.165	.4982	.3049	.2792	.1573	.1203	.0656

Table 2. Values $r(k) = H_{k,n}/\gamma(\omega^*) - 1$ and $r^*(k) = H^*_{k,n}/\gamma(\omega^*) - 1$ for P_{cont} with three different ω^* and 0.1

ϵ	ω^*	$\gamma(\omega^*)$	$r(100)$	$r^*(100)$	$r(250)$	$r^*(250)$	$r(900)$	$r^*(900)$
	1.5	0.879	.0683	.0707	.0502	1.0458	.0339	.0296
.10	3	1.500	.3842	.2828	.2512	.1799	.1340	.0970
	5	2.165	.9298	.6250	.5536	.3352	.2383	.1310

Case 2. Infinite moving average process Consider the infinite moving average process with the following autoregressive form

$$\mathbf{X}_i = 1.3\,\mathbf{X}_{i-1} - 0.7\,\mathbf{X}_{i-2} + \mathbf{Z}_i, \quad i = 1, 2, \ldots, n.$$

This process is considered in Resnick and Starica (1993) with respect to the Hill estimator. In each of the following examples we simulated a sample of $n = 2500$

such data and plotted the Hill and the t-Hill plots of the corresponding estimators together with their 0.05-confidence intervals.

a.) Pareto noise. Here $\mathbf{Z}_i$, $i = 1, 2, \dots$ are independent and Pareto(α) distributed. For each sample we calculated $\mathbf{X}_i, i = 1, 2, \dots, n$, Hill and the t-Hill estimators for $k(n) = 10, 11, \dots, 500$. The charts of corresponding estimators together with their 0.05-confidence intervals are given on Fig. 8 and Fig. 9, for $\alpha = 0.3, 1$.

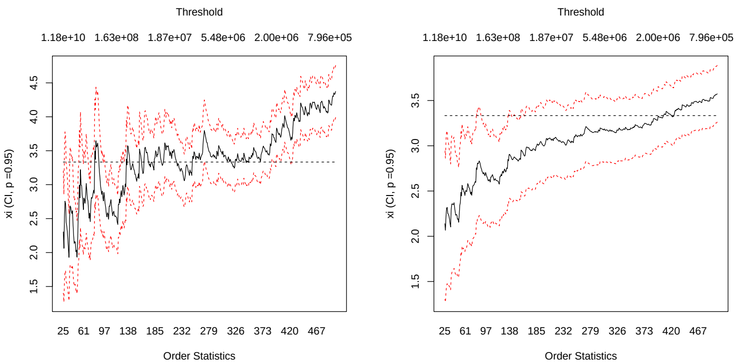

Fig. 8. Rate of convergence of the t-Hill (left) and Hill (right) estimators, $\alpha = 0.3$

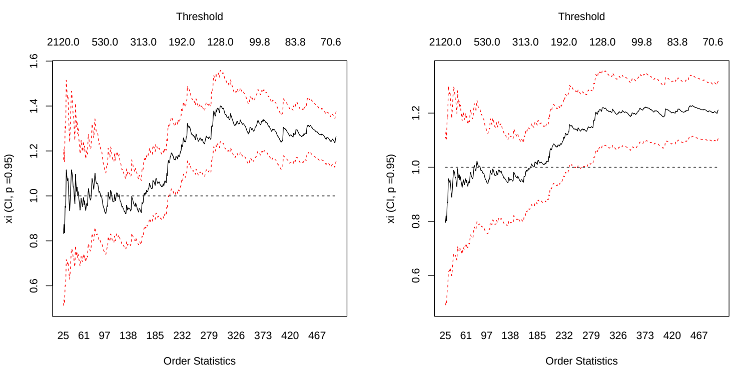

Fig. 9. Rate of convergence of the t-Hill (left) and Hill (right) estimators, $\alpha = 1$

Both estimators have similar behaviour for fixed number of upper order statistics. In order to obtain consistent estimators we need $k(n) = o(n)$, $k(n) \to \infty$ and $n \to \infty$. Therefore both estimators are appropriate only for very large samples.

b.) Infinite moving average data with contaminated Pareto noise. In this case $\mathbf{Z}_i$, $i = 1, 2, \dots, n$ are independent Pareto(α) distributed on $(1, \infty)$

with probability 0.9, contaminated with Pareto(δ) distribution with probability 0.1 for different $\alpha > 0$ and $\delta > 0$. We take a sample of $n = 10000$ observations and calculated the Hill and the t-Hill estimators. The corresponding estimators together with their 0.05-confidence intervals are given on Figs 10–12. The results are very different from **Case 1), d.)**.

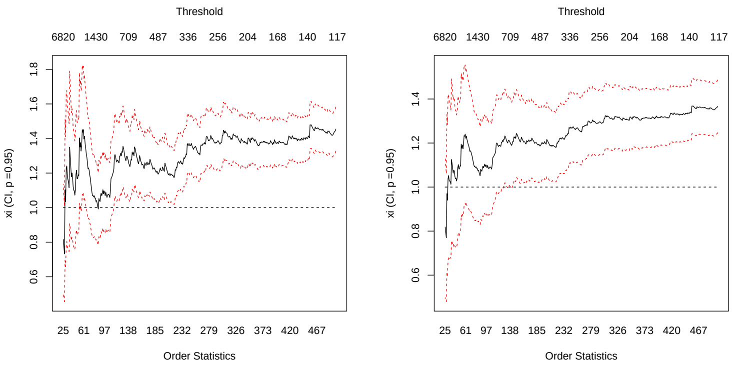

Fig. 10. Rate of convergence of the t-Hill (left) and Hill (right) estimators, $\alpha = 1, \delta = 0.5$

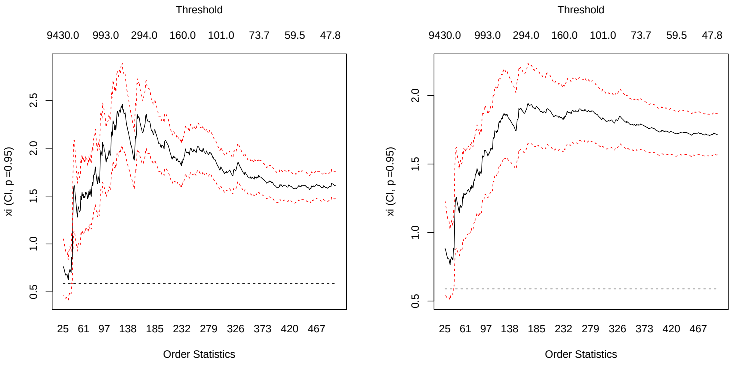

Fig. 11. Rate of convergence of the t-Hill (left) and Hill (right) estimators, $\alpha = 1.7, \delta = 0.5$

Here $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n$ are dependent. Both estimators depend on upper order statistics, but situation is very different from the i.i.d. case. In view of regular variation of the tail of the distribution of the noise components, $1/\min(\alpha, \delta)$ determines the largest values among $\mathbf{Z}_1, \mathbf{Z}_2, \ldots, \mathbf{Z}_n$. If for some $i = 1, 2, \ldots, n$ $\mathbf{Z}_i$ is huge then it influences all of $\mathbf{X}_j$, $j = i+1, i+2, \ldots, n$. Therefore in this case

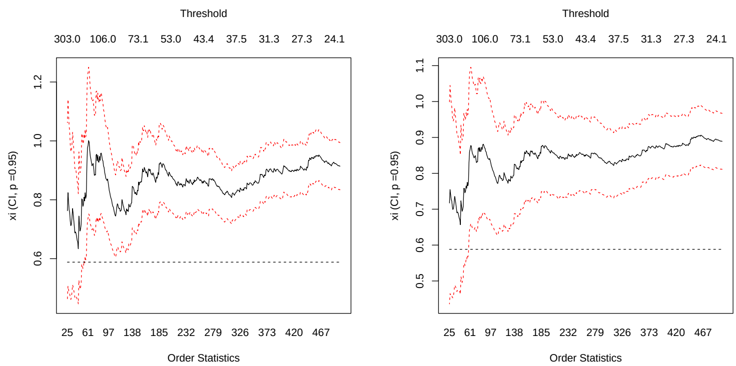

Fig. 12. Rate of convergence of the t-Hill (left) and Hill (right) estimators, $\alpha = 1.7, \delta = 1$

both estimators are not robust. They are good not for $1/\alpha$, but for $1/\min(\alpha, \delta)$.

6. Power of selected tests for normality against Pareto distribution. In this section we present power of selected classical and robust tests for normality against Pareto alternative distributions. For this purpose we assume Pareto (α, c) distribution for $\alpha \in \{0.5, 1, 2, 5, 10\}$, $c = 1$. Simulation study has been performed with sample sizes $n \in \{10, 15, 20, 25\}$, 100000 repetitions and the following tests of normality: the classical Jarque-Bera test (JB), the robust Jarque-Bera test (RJB), the Shapiro-Wilk test (SW), the Anderson-Darling test (AD), the Lilliefors test (LT), directed SJ test (SJ_{dir}), three medcouple tests ($MC1, MC2, MC3$) and selected RT tests which were introduced in Stehlík, Fabián and Střelec (2012), i.e. $RT_{JB}9$, $RT_{JB}39$ and $RT_{JB}42$ tests. Therein we substantially used t-Hill estimator for Pareto tail to classify optimal test again given alternative. Therefore, Tables 3–6 present the results of Monte Carlo simulations of power of analyzed tests against Pareto (α, $c = 1$) alternative distributions. From them we may conclude that:

- The Shapiro-Wilk test outperforms the other tests for normality.
- For small sample sizes $n = 10, 15, 20$ and 25 the $RT_{JB}9$ outperforms JB test for all analyzed Pareto alternatives. For example, power of the classical Jarque-Bera test against Pareto ($\alpha = 1$, $c = 1$) alternative and very small sample size $n = 10$ is 0.757. In comparison, power of $RT_{JB}9$ test (test based on "mean-median" robustification of the classical Jarque-Bera test - see Stehlík, Fabián and Střelec (2012)) against the same alternative is 0.857 and is comparable with the Shapiro-Wilk test, which has power of 0.870
- The power of the tests is decreasing with the increase of the parameter

Table 3. Power of analyzed tests against Pareto (α, $c = 1$) alternative distributions for $n = 10$

	$n = 10$				
test	$\alpha = 0.5$	$\alpha = 1$	$\alpha = 2$	$\alpha = 5$	$\alpha = 10$
JB	0.886	0.757	0.610	0.444	0.403
AD	0.958	0.855	0.712	0.545	0.483
LT	0.916	0.765	0.585	0.416	0.367
RJB	0.885	0.745	0.579	0.414	0.366
SJ_{dir}	0.875	0.712	0.530	0.359	0.306
SW	**0.966**	**0.870**	**0.741**	**0.578**	**0.518**
$MC1$	0.635	0.445	0.330	0.251	0.218
$MC2$	0.217	0.150	0.121	0.107	0.105
$MC3$	0.732	0.507	0.361	0.261	0.229
$RT_{JB}9$	0.956	0.857	0.707	0.556	0.484
$RT_{JB}39$	0.825	0.669	0.501	0.366	0.308
$RT_{JB}42$	0.441	0.284	0.193	0.137	0.130

Table 4. Power of analyzed tests against Pareto (α, $c = 1$) alternative distributions for $n = 15$

	$n = 15$				
test	$\alpha = 0.5$	$\alpha = 1$	$\alpha = 2$	$\alpha = 5$	$\alpha = 10$
JB	0.979	0.912	0.799	0.650	0.585
AD	0.997	0.968	0.896	0.766	0.710
LT	0.990	0.920	0.787	0.611	0.540
RJB	0.976	0.897	0.767	0.602	0.533
SJ_{dir}	0.970	0.868	0.702	0.507	0.435
SW	**0.998**	**0.976**	**0.921**	**0.809**	**0.760**
$MC1$	0.770	0.550	0.398	0.285	0.259
$MC2$	0.425	0.253	0.174	0.141	0.121
$MC3$	0.914	0.711	0.530	0.375	0.327
$RT_{JB}9$	0.994	0.953	0.863	0.721	0.656
$RT_{JB}39$	0.973	0.894	0.758	0.588	0.516
$RT_{JB}42$	0.748	0.543	0.394	0.286	0.244

Table 5. Power of analyzed tests against Pareto (α, $c = 1$) alternative distributions for $n = 20$

	$n = 20$				
test	$\alpha = 0.5$	$\alpha = 1$	$\alpha = 2$	$\alpha = 5$	$\alpha = 10$
JB	0.997	0.974	0.908	0.789	0.716
AD	**1.000**	0.995	0.968	0.894	0.845
LT	0.999	0.978	0.902	0.753	0.676
RJB	0.996	0.963	0.879	0.740	0.656
SJ_{dir}	0.993	0.941	0.810	0.624	0.528
SW	**1.000**	**0.997**	**0.983**	**0.926**	**0.891**
$MC1$	0.891	0.706	0.529	0.399	0.358
$MC2$	0.518	0.322	0.212	0.170	0.167
$MC3$	0.972	0.845	0.664	0.504	0.453
$RT_{JB}9$	0.999	0.991	0.948	0.854	0.801
$RT_{JB}39$	0.995	0.965	0.879	0.737	0.661
$RT_{JB}42$	0.892	0.738	0.579	0.432	0.384

Table 6. Power of analyzed tests against Pareto (α, $c = 1$) alternative distributions for $n = 25$

	$n = 25$				
test	$\alpha = 0.5$	$\alpha = 1$	$\alpha = 2$	$\alpha = 5$	$\alpha = 10$
JB	**1.000**	0.993	0.961	0.875	0.830
AD	**1.000**	0.999	0.989	0.956	0.928
LT	**1.000**	0.994	0.958	0.862	0.791
RJB	0.999	0.987	0.939	0.825	0.765
SJ_{dir}	0.998	0.973	0.880	0.705	0.609
SW	**1.000**	**1.000**	**0.995**	**0.976**	**0.957**
$MC1$	0.918	0.741	0.570	0.429	0.388
$MC2$	0.557	0.320	0.222	0.169	0.160
$MC3$	0.981	0.868	0.695	0.519	0.466
$RT_{JB}9$	**1.000**	0.997	0.978	0.919	0.876
$RT_{JB}39$	0.999	0.989	0.944	0.838	0.770
$RT_{JB}42$	0.959	0.857	0.716	0.574	0.518

α. For example, power of the Shapiro-Wilk test against Pareto ($\alpha = 1$, $c = 1$) alternative and small sample size $n = 15$ is 0.976 and against Pareto ($\alpha = 10$, $c = 1$) alternative and the same sample size is 0.760.

- The smallest power show the medcouple tests and $RT_{JB}42$ (test based on "trimm-trimm" robustification of the classical Jarque-Bera test – see Stehlík, Fabián and Střelec (2012)).

Based upon our experience we can recommend the Shapiro-Wilk test and $RT_{JB}9$ test instead of the classical Jarque-Bera test (JB) and robust Jarque-Bera test (RJB), especially for small and very small sample sizes.

REFERENCES

[1] J. Beirlant, Y. Goegebeur, J.Segers, J.Teugels. Statistics of Extremes. Theory and applications. Wiley, 2005.

[2] J. Beran, D. Schell. On robust tail index estimation. *Computational Statistics and Data Analysis*, (2010). doi:10.1016/j.csda.2010.05.028

[3] D. Cline. Infinfite series of random variables with regularly varying tailes. Tech. Report No 83–24, Indtitute of Applied Mathematics and Statistics, University of British Columbia, 1983.

[4] P. Embrechts, Cl. Klueppelberg, Th. Mikosch. Modeling extremal events for insurance and finance. Springer Verlag, 1997.

[5] Zd. Fabián. Induced cores and their use in robust parametric estimation. *Communication in Statistics, Theory Methods* **30** (2001), 537–556.

[6] Zd. Fabián. Estimation of simple characteristics of samples from skewed and heavy-tailed distribution. Recent Advances in Stochastic Modeling and Data Analysis (ed. C. Skiadas), Singapore, World Scientific, 2007, 43–50.

[7] Zd.Fabián. New measures of central tendency and variability of continuous distributions. *Communication in Statistics, Theory Methods*, **37** (2008), 159–174.

[8] Zd. Fabián. A New Statistical Tool: Scalar Score Function, *Computer Technology and Application*, **2**, (2011), 29–35.

[9] Zd. Fabian, M. Stehlik. On robust and distribution sensitive Hill like method. *IFAS Reasearch Paper Series*, **43**, (2009), No 4.

[10] M. I. Fraga Alves. A location invariant Hill-type estimator. *Extremes* **4**, (2001), 199–217.

[11] M. Ivette Gomes and O. Oliveira. Censoring estimators of a positive tail index. *Statistics & Probability Letters* **65**, (2003), No 3, 147–159.

[12] B. M. Hill. A simple general approach to inference about the tail of the distribution. *Ann. Statist.* **3**, (1975), 1163–1174.

[13] P. Jordanova, J. Dusek, M. Stehlik. Modelling methane emission by the infinite moving average process. *Chemometrics and Intelligent Laboratory Systems* (submitted 2012).

[14] P. Jordanova, E. Pancheva Weak Asymptotic Results for t-Hill Estimator. *C. R. Acad. Bulg. Sci* **65** (2012), No 12, 1649–1656.

[15] J. Li, Z. Peng, S. Nadarajah. Asymptotic normality of location invariant heavy tail index estimator. *Extremes* **13**, (2010), No 3, 269–290.

[16] S. Resnick, C. Starica. Consistency of Hill's Estimator for Dependent Data. Technical report No. 1077, School of Operational Research and Industrial Engeneering Cornell University Itaca NY, **14853**, 1993, 43–50.

[17] Stehlík M., Potocký R., Waldl H. and Fabián Z. On the favourable estimation of fitting heavy tailed data. *Computational Statistics* **25** (2010), 485–503.

[18] M. Stehlík, Zd. Fabián, L. Střelec. Small Sample Robust Testing for Normality Against Pareto Tails. *Communications in Statistics – Similation and Computation* **41** (2012), No 7, 1167–1194.

[19] B. Vandewalle B., J. Beirlant, A. Christmann, M. Hubert. A robust estimator for the tail index of Pareto-type distributions. *Computational Statistics & Data Analysis* **51**, (2007), No 12, 6252–6268.

Pavlina K. Jordanova
Faculty of Mathematics and Informatics
Shumen University
115, Alen Mak Str.
9712 Shumen, Bulgaria
e-mail: `pavlina_kj@abv.bg`

Milan Stehlík
Department of Applied Statistics
Johannes Kepler University in Linz
Freistädter Straße 315, 2. Stock
A-4040 Linz a. D., Austria
e-mail: `Milan.Stehlik@jku.at`

Zdeněk Fabián
Institute of Computer Science
Academy of Sci. of the Czech Republic
Prague, Czech Republic
e-mail: `zdenek@cs.cas.cz`

Luboš Střelec
Dep. of Stat. and Oper. Analysis
Mendel University in Brno
Brno, Czech Republic
e-mail: `lubos.strelec@mendelu.cz`