Johnson Point and Johnson Variance

Fabián, Zdeněk
2006

Institute of Computer Science
Academy of Sciences of the Czech Republic

Johnson point and Johnson variance

Zdeněk Fabián

Technical report No. 971

July 2006

Pod Vodárenskou věží 2, 182 07 Prague 8, phone: +420 266 053 210, fax: +420 286 585 789,
e-mail:zdenek@cs.cas.cz

Institute of Computer Science
Academy of Sciences of the Czech Republic

Johnson point and Johnson variance*

Zdeněk Fabián

Technical report No. 971

July 2006

Abstract:

New measures of central tendency and dispersion of continuous probability distributions were introduced in [1]. In this paper we show that they offer a simple and suitable description of heavy-tailed distributions, i.e., the distributions which may not have the mean and/or variance.

MSC: 62A01, 62F01

Keywords:
basic characteristics, Johnson transform, parametric estimates

*The author is indebted to I. Vajda for valuable critical remarks and suggestions. The work was supported by Grant Agency AS CR under grant number IAA1075403.

1 Introduction

The vector of parameters of parametric distributions has often no component characterizing the central tendency and/or dispersion of the distribution. On the other hand, the often used characteristics, the mean and the variance, may not exist if the distribution is heavy-tailed.

New measures of the central tendency and dispersion of continuous probability distributions were proposed in [1].

Distribution with distribution function F is said to be supported by interval $(a, b) \subseteq \mathbb{R}$ if its density $f(x) = dF(x)/dx$ satisfies relation

$$f(x) \begin{cases} > 0 & \text{for} \quad x \in (a, b) \\ = 0 & \text{for} \quad x \in \mathbb{R} \setminus (a, b). \end{cases}$$

As shown in [2], the distributions supported by $(a, b) \neq \mathbb{R}$ can be viewed as transformed 'prototypes' supported by $\mathbb{R}$. A suitable transformation $\eta^{-1} : \mathbb{R} \to (a, b)$ is the inverse of the Johnson transformation [3] adapted to arbitrary interval support, $\eta : (a, b) \to \mathbb{R}$, given by

$$\eta(x) = \begin{cases} x & \text{if} \quad (a, b) = \mathbb{R} \\ \log(x - a) & \text{if} \quad -\infty < a < b = \infty \\ \log \dfrac{(x - a)}{(b - x)} & \text{if} \quad -\infty < a < b < \infty \\ \log(b - x) & \text{if} \quad -\infty = a < b < \infty. \end{cases} \tag{1.1}$$

Let F supported by (a, b) be given by $F = G \circ \eta$, where G is a distribution supported by $\mathbb{R}$, called a *prototype*. Denote by g the density of G and by Q its score function

$$Q(y) = -\frac{g'(y)}{g(y)}. \tag{1.2}$$

The density of F is the transformed density of the prototype

$$f(x) = g(\eta(x))\eta'(x). \tag{1.3}$$

As an important characteristic of distribution F was shown to be a transformed score function of the prototype,

$$T(x) = Q(\eta(x)), \tag{1.4}$$

called in [2] a *core function*. It was shown in [2] that from (1.4) it follows that

$$T(x) = \frac{1}{f(x)} \frac{d}{dx} \left(-\frac{1}{\eta'(x)} f(x) \right). \tag{1.5}$$

Thus, the core function of any distribution with differentiable density can be determined without reference to its prototype by differentiating according to x.

An unusual feature of function T is that it is 'support-dependent'. It should be more accurately called a Johnson core since η is, in fact, to a certain extent arbitrary. On the other hand, for many distributions 1.1 represents a best choice: the logarithmic transformation is suitable for exponential densities and, moreover,

i/ the prototype of the lognormal distribution is the normal one
ii/ core function of the uniform distribution on $(0,1)$ is linear
iii/ for a given $t \in (a,b)$, $\eta(x) - \eta(t)$ is continuous when $b \to \infty$.

However, for some distributions, simpler core functions can be obtained by the use of other η. For instance, for distribution expressed by means of trigonometric functions on $(-\pi/2, \pi/2)$, a more suitable mapping is $\eta(x) = \tan(x)$. To avoid ambiguities, we fixed the mapping (1.1).

2 Johnson score

The Johnson score of a continuous distribution was introduced in [1].

Definition 2.1 Let $(a,b) \subseteq \mathbb{R}$ and η be given by (1.1). Let F be an absolutely continuous distribution supported by (a,b) with twice continuously differentiable density $f(x)$. Let function $T(x)$ be given by (1.5) and the solution x^* of equation

$$T(x) = 0$$

be unique. x^* will be called a *Johnson point* and function

$$S(x) = \eta'(x^*)T(x) \tag{2.1}$$

a *Johnson score* of distribution F.

If F has support $\mathbb{R}$ then $\eta'(x) = 1$, Johnson score is the score function and Johnson point is the mode. A large class of distributions supported by $(a,b) \neq \mathbb{R}$, the Johnson scores of which are well-known functions, is described in the next section.

3 Distributions with location and scale prototypes

Let G be a prototype with continuously differentiable and unimodal density g,

$$g'(0) = 0, \tag{3.1}$$

and let S be its Johnson score. Let $\mathcal{G}_{\mu,s} = \{G_{\mu,s} : \mu \in \mathbb{R}, s \in (0,\infty)\}$ be a parametrized family with parent G and densities and score functions in forms

$$g_{\mu,s}(y) = \frac{1}{s} g\left(\frac{y-\mu}{s}\right) \tag{3.2}$$

$$Q_{\mu,s}(y) = \frac{1}{s} Q\left(\frac{y-\mu}{s}\right). \tag{3.3}$$

Denote by

$$t = \eta^{-1}(\mu) \tag{3.4}$$

the transformed location of prototype $G_{\mu,s}$ and construct a family of transformed distributions with support (a,b),

$$\mathcal{F}_{t,s}^{(a,b)} = \{F_{t,s} : F_{t,s} = G_{\mu,s} \circ \eta\}. \tag{3.5}$$

(1.3) and (3.2) with (1.4) and (3.3) imply that the density and Johnson score of $F_{t,s} \in \mathcal{F}_{t,s}^{(a,b)}$ are

$$f_{t,s}(x) = \frac{1}{s} g\left(\frac{\eta(x)-\eta(t)}{s}\right)\eta'(x) \tag{3.6}$$

$$T_{t,s}(x) = \frac{1}{s} Q\left(\frac{\eta(x)-\eta(t)}{s}\right) \tag{3.7}$$

(see Proposition 6 in [4]). A set of families in form (3.5) will be called a *set of distributions on* (a,b) *with location and scale prototypes.*

The main proposition is as follows.

Proposition 3.1 Let distribution $F_{t,s} \in \mathcal{F}_{t,s}^{(a,b)}$. Then its Johnson point is t and its Johnson score

$$S_{t,s}(x) = \frac{\partial}{\partial t} \log f_{t,s}(x). \tag{3.8}$$

Proof. By (2.1), (3.7), (1.2) and (3.1), $T_{t,s}(t) = 0$. The second assertion is proven in [1].

The parameter t (the 'Johnson image' of the location of the prototype) will be called the *Johnson parameter*. By Proposition 3.1, the Johnson score of a distribution with location and scale prototype is the likelihood score for the Johnson parameter.

Example 3.2 Let $g(z) = e^{-z}e^{-e^{-z}}$ be density of the parent of the extreme value family $\mathcal{G}_{\mu,s}$. A location and scale prototype $G_{\mu,s} \in \mathcal{G}_{\mu,s}$ has density

$$g_{\mu,s}(y) = s^{-1}e^{-\frac{y-\mu}{s}}e^{-e^{-\frac{y-\mu}{s}}}$$

and score function

$$Q_{\mu,s}(y) = s^{-1}(1 - e^{-\frac{y-\mu}{s}}). \tag{3.9}$$

$G_{\mu,s}$ is the prototype of $F_{t,s} = G_{\mu,s} \circ \eta$, the density of which is, by (3.6),

$$f_{t,s}(x) = \frac{1}{s} g\left(\frac{\ln x - \ln t}{s}\right)\frac{1}{x} = \frac{\beta}{x}\left(\frac{x}{t}\right)^{-\beta} e^{-(x/t)^{-\beta}}, \tag{3.10}$$

where $\beta = 1/s$. The transformed family is the Fréchet family with Johnson parameter $t = e^{\mu}$. By (3.7), (3.9) and (2.1), Johnson score of $F_{t,s}$ is

$$S_{t,s}(x) = \frac{1}{t}\beta\left[1 - (x/t)^{-\beta}\right],$$

which is equal to the likelihood score for t.

4 Johnson score of arbitrary distributions

In preceding section we see that Johnson scores of a large group of distributions supported by $(a, b) \neq \mathbb{R}$ with location and scale prototypes are well-known important functions. The Johnson scores of distributions supported by $(a, b) \neq \mathbb{R}$ without Johnson parameter are unknown, but we expect that they can be as useful as the likelihood scores of distributions of the first group.

Example 4.1 Consider gamma distribution with support $(0, \infty)$ and density

$$f_{\alpha,\gamma}(x) = \frac{\gamma^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\gamma x},$$

where $\alpha > 0, \gamma > 0$. From (1.5), $T_{\alpha,\gamma}(x) = -1 - x f'_{\alpha,\gamma}(x)/f_{\alpha,\gamma}(x) = \gamma x - \alpha$. The zero of $T_{\alpha,\gamma}$ is $x^* = \alpha/\gamma$. By (2.1) one obtains Johnson score

$$S_{\alpha,\gamma}(x) = \frac{1}{x^*}(\gamma x - \alpha) = \gamma \left(\frac{x}{\alpha/\gamma} - 1 \right). \tag{4.1}$$

By setting $t = \alpha/\gamma$ we obtain a reparametrized form of the gamma distribution with Johnson parameter,

$$f_{t,\alpha}(x) = \frac{\alpha^\alpha}{\Gamma(\alpha) x} \left(\frac{x}{t} \right)^\alpha e^{-\alpha x/t}.$$

Let us now present a typical example of a distribution which cannot be reparametrized into a form with Johnson parameter.

Example 4.2 Consider Fisher-Snedecor distribution with support $(0, \infty)$ and density

$$f_{p,q}(x) = \frac{1}{xB(p,q)} \frac{(x/\rho)^p}{(x/\rho + 1)^{p+q}} \tag{4.2}$$

where $\rho = q/p$, $p > 0, q > 0$. The usual notation can be introduced by setting $p = \nu_1/2, q = \nu_2/2$ and $\rho = \nu_2/\nu_1$. The mean of (4.2) exists only if $q > 1$ and equals to $\tau = q/(q-1)$, the variance exists only if $q > 2$ and equals to

$$\sigma^2 = \frac{q^2(p+q-1)}{p(q-1)^2(q-2)} \tag{4.3}$$

(see [5]). By (1.5),

$$T_{p,q}(x) = -1 - x f'_{p,q}(x)/f_{p,q}(x) = \frac{qx/\rho - p}{x/\rho + 1}, \tag{4.4}$$

and from $T_{p,q}(x^*) = 0$ we obtain $x^* = \rho p/q = 1$. Johnson point of any member of the Fisher-Snedecor family is thus equal to one and Johnson scores are equal to core functions (4.4). Some densities (4.2) and Johnson scores (4.4) are plotted in Fig.1.

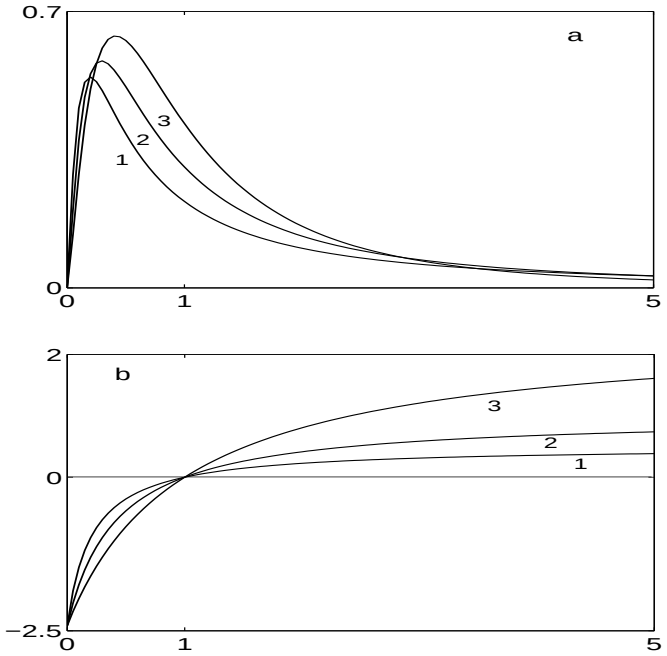

Figure 1. Densities (a) and Johnson scores (b) of the Fisher-Snedecor distributions, $p = 1 + \sqrt{2}$. 1: $q = 0.5$, 2 : $q = 1.0$, 3 : $q = 1 + \sqrt{2}$.

5 Johnson variance

The Johnson point of a distribution exists if its prototype is unimodal and Johnson score continuous. In [1], we proposed to take it as a measure of the central tendency of a distribution.

A number which can serve as a measure of dispersion of a distribution around its Johnson point was introduced in [1] as follows.

Definition 5.1 Let density f of a distribution F satisfy the assumptions of Definition 2.1, let S be its Johnson score and let integral

$$0 < ES^2 = \int_a^b S^2(x) f(x)\, dx$$

be finite. A value

$$\sigma_S^2 = (ES^2)^{-1} \tag{5.1}$$

will be called a *Johnson variance* of F and its square root a *standard Johnson deviation.*

Consider particular cases of (5.1).

i/ Taking the expectation of (2.1) squared, the Johnson variance of a distribution supported by $(0,\infty)$ is

$$\sigma_S^2 = (x^*)^2/ET^2. \tag{5.2}$$

ii/ By (3.7), for $F_{t,s} \in \mathcal{F}_{t,s}^{(a,b)}$ it holds that $ET^2 = s^{-2}EQ^2$ and, by Proposition 1, $x^* = t$. Taking the expectation of (2.1) squared, one obtains

$$\sigma_S^2 = \frac{s^2}{[\eta'(t)]^2 EQ^2}$$

where EQ^2 is the Fisher information of the parent of the prototype. For distributions with location and scale prototypes supported by $(0,\infty)$, $\sigma_S^2 = s^2t^2/EQ^2$.

Example 5.2 (Example 4.1 continued) For gamma distribution, $ET^2_{\alpha,\gamma} = \alpha$ and $x^* = \alpha/\gamma$, so that by (5.2)

$$\sigma_S^2 = (\alpha/\gamma)^2\alpha^{-1} = \alpha/\gamma^2.$$

The Johnson point of the gamma distribution is equal to the mean and the Johnson variance equals to the usual variance.

Example 5.3 (Example 4.2 continued) The second moment of the core function of Fisher-Snedecor distribution is $ET^2 = pq/(p+q+1)$ and $x^* = 1$. The Johnson variance is thus

$$\sigma_S^2 = \frac{p+q+1}{pq}. \tag{5.3}$$

It is interesting to compare (5.3) with (4.3). Consider the case $q = p$. For large p it holds that $\sigma^2 \sim \sigma_S^2 \sim 2/p$. For small p, the usual variance behaves by an incomprehensible manner, whereas the behavior of the Johnson variance is quite reasonable.

Fig.2 shows the densities of prototypes of distributions from Fig.1a,

$$g_{p,q}(y) = \frac{1}{B(p,q)}\frac{e^{p(y-y_0)}}{(e^{y-y_0}+1)^{p+q}},$$

where $y_0 = \ln\gamma = \ln(q/p)$. The distributions do not have a location parameter (y_0 is a constant). Let us find a symmetric prototype distribution with unique Johnson variance. The solution of equation (5.3) for $\sigma_S = 1$ and $p = q$ is $p = 1+\sqrt{2}$. In Fig. 2, a striking coincidence of $g_{1+\sqrt{2},1+\sqrt{2}}$ (the curve denoted by 3) with the density of the standard normal distribution (dotted curve) is apparent. By the dashed line, $g_{4.85,4.85}$ with $\sigma = 1$ is plotted.

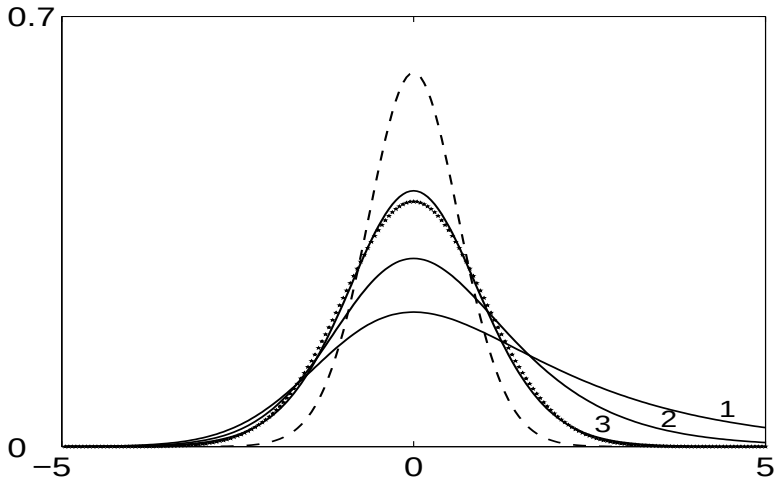

Figure 2. Densities of the prototypes of distributions from Fig.1a. Dotted: density of standard normal, dashed: density of the beta prototype with $\sigma = 1$.

Example 5.4 Consider distribution with support $\mathbb{R}$ and density

$$f_{\mu,s,\nu}(x) = \frac{\nu^{\nu/2}}{sB(1/2,\nu/2)} \frac{1}{\left(\nu + \left(\frac{x-\mu}{s}\right)^2\right)^{\frac{\nu+1}{2}}}. \tag{5.4}$$

Particularly, $f_{\mu,s,1}$ is the Cauchy distribution, having neither mean nor variance. Further, $f_{0,1,n}$ is the Student distribution with n degrees of freedom. Its mean $m = 0$ and variance $\sigma^2 = n/(n-2)$ exist only if $n > 1$ and $n > 2$, respectively.

Distribution (5.4) has score function

$$S(x) = \frac{\nu+1}{s} \frac{(x-\mu)/s}{\nu + \left(\frac{x-\mu}{s}\right)^2}.$$

Its second score moment is

$$\begin{aligned} ES^2 &= \int_{-\infty}^{\infty} S^2(x) f_{\mu,s,\nu}(x)\, dx \\ &= \left(\frac{\nu+1}{s}\right)^2 \frac{\nu^{\nu/2}}{B(1/2,\nu/2)} \int_{-\infty}^{\infty} \frac{\xi^2 d\xi}{(\nu+\xi^2)^{\frac{\nu+1}{2}+2}}. \end{aligned} \tag{5.5}$$

Since

$$\int_{-\infty}^{\infty} \frac{x^2 dx}{(\nu+x^2)^{\lambda}} = \frac{1}{2(\lambda-1)} \int_{-\infty}^{\infty} \frac{dx}{(\nu+x^2)^{\lambda-1}} = \frac{\nu^{1/2} B(1/2, \lambda-3/2)}{2(\lambda-1)\nu^{\lambda-1}},$$

we obtain

$$\begin{aligned} ES^2 &= \left(\frac{\nu+1}{s}\right)^2 \frac{\nu^{\nu/2}\nu^{1/2}}{2(\lambda-1)\nu^{\lambda-1}} \frac{B(1/2,\lambda-3/2)}{B(1/2,\nu/2)} \\ &= \frac{(\nu+1)^2}{s^2}\frac{1}{\nu(\nu+3)}\frac{\nu}{\nu+1} = \frac{\nu+1}{\nu+3}\frac{1}{s^2}. \end{aligned}$$

Johnson variance of the generalized Student distribution (5.4) is thus

$$\omega^2 = \frac{\nu+3}{\nu+1}s^2. \tag{5.6}$$

In Fig.3. there are plotted densities of three distributions (5.4) with values of parameters $\mu = 0$, $\nu = 1, 1.5, 3$ and different s such that the Johnson variances (5.6) are $\omega^2 = 3$. The distribution with $\nu = 1$ is the Cauchy distribution. All the three plotted distributions with equal Johnson variances are similar each other, but the variances of the first two distributions do not exist, whereas the variance of the third distribution, the Student distribution with three degrees of freedom, equals to the Johnson variance. For comparison, the density of the normal distribution with $\sigma = 3$ is plotted by the dotted curve. We see that in cases, at which the classical variance fails, the Johnson variance can characterize the variability of heavy-tailed distributions similarly as the classical variance characterizes variability of the normal distribution.

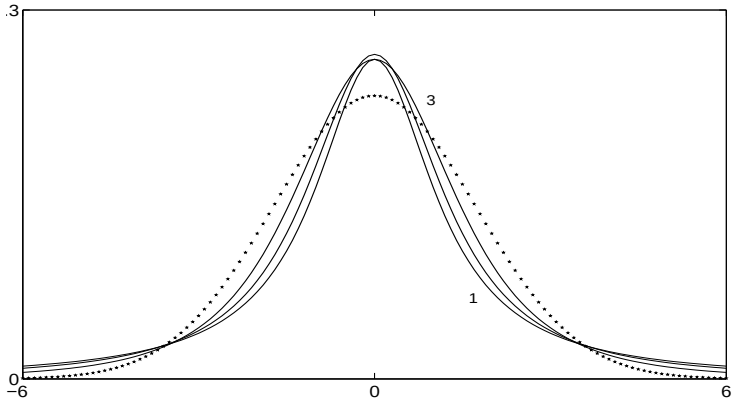

Figure 3. Densities of the generalized Student distributions with Johnson variance $\omega^2 = 3$. 1: $\nu = 1$, 3: $\nu = 3$, without number: $\nu = 1.5$. Dotted: normal with $\sigma^2 = 3$.

6 Characteristics of data samples

Let $\Theta \subset \mathbb{R}^m$. Denote by S_θ the Johnson score of distribution $F_\theta, \theta \in \Theta$. In [1], a suggestion was proposed to characterize the 'central tendency' of a sample $X_1, ..., X_n$ from F_θ with θ unknown by the estimate $\hat{x}^*$ of the Johnson point of F_θ, that is, by the solution of equation

$$S_{\hat{\theta}}(x) = 0 \tag{6.1}$$

and the dispersion of the sample around $\hat{x}^*$ by the estimate of the Johnson variance

$$\hat{\sigma}_S^2 = (ES_{\hat{\theta}}^2)^{-1}, \tag{6.2}$$

where $\hat{\theta}$ in (6.1) and (6.2) is the maximum likelihood estimate of θ. If h is a real function continuously differentiable at θ, the statistical properties of $h(\hat{\theta})$ can be derived from the properties of $\hat{\theta}$ (e.g., Theorem A, [6], pp.122). By assumptions of Definition 5.1, S_θ is continuously differentiable. Having the maximum likelihood estimates of parameters, it is easy to obtain without any additional efforts the numbers characterizing the central tendency and dispersion of the data sample and their asymptotic properties.

Example 6.1 Sample Johnson points (commas) and sample standard Johnson deviations (SJD) of eight samples of size 9 from the Fréchet distribution (3.10) with $t = 1$ and $\beta = 0.95$ (i.e., from a distribution having neither mean nor variance) are compared with the medians (black circles) and median absolute deviations MAD = median($|x_i - \text{median}(x_j)|$) in Fig.4. It is apparent that the results are comparable and that the estimates of the Johnson point and Johnson variance of heavy-tailed Fréchet distribution are robust.

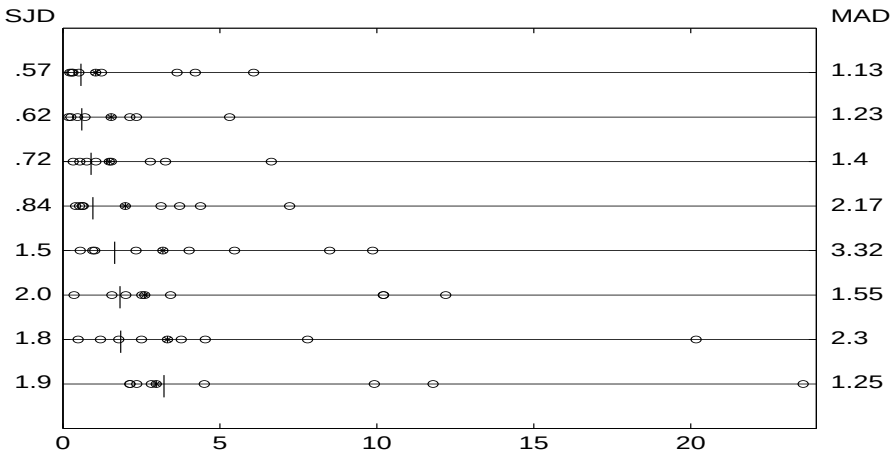

Figure 4. Comparison of descriptive statistics of samples from the Fréchet distribution: | $\hat{x}^*$, • median

Bibliography

[1] Z. Fabián. Johnson score and characteristics of distributions. *Statist. and Prob. Letters*, 2006 (submitted).

[2] Z. Fabián. Induced cores and their use in robust parametric estimation. *Commun. in Statist.-Theory Meth.*, 30: 537–556, 2001.

[3] N. L. Johnson. Systems of frequency curves generated by methods of translations. *Biometrika*, 36: 149–176, 1949.

[4] Z. Fabián, and I. Vajda. Core functions and core divergences of regular distributions. *Kybernetika*, 39: 29–42, 2003.

[5] N. L. Johnson, S. Kotz, and M. Balakrishnan. *Continuous univariate distributions 1, 2*, Wiley, New York, 1994, 1995.

[6] R. J. Serfling. *Approximation Theorems of Mathematical Statistics*, Wiley, New York, 1980.