New Measures of Central Tendency and Variability of Continuous Distributions

Fabián, Zdeněk
2007

Institute of Computer Science
Academy of Sciences of the Czech Republic

New measures of central tendency and variability of continuous distributions

Zdeněk Fabián

Technical report No. 996

May 2007

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

New measures of central tendency and variability of continuous distributions [1]

Zdeněk Fabián

Technical report No. 996

May 2007

Abstract:

A scalar inference function introduced in Fabián (2001) is generalized for a larger class of continuous distributions. Its first two moments are used for introduction of measures of the central tendency and the variability of the distribution. The number of examples shows that the new measures are plausible for continuous distribution, even for such for which the mean and/or the variance do not exist. They can be estimated from the data through the maximum likelihood estimates of the parameters; the estimates are expressed in particular cases by algebraic formulas without need to estimate the parameters.

Keywords:
description of distributions; basic statistics; score function; point estimates

[1] The work was supported by the Grant Agency AS CR under grant IS 1ET400300513 and the Institutional Research Plan AV0Z10300504. Final version of report 978, accepted in Communication in Statistics, Theory Methods.

1. INTRODUCTION

Let F be a distribution function of a continuous probability distribution with density

$$f(x) \begin{cases} > 0 & \text{if} \quad x \in \mathcal{X} \\ = 0 & \text{if} \quad x \in \mathbb{R} \setminus \mathcal{X} \end{cases}$$

where $\mathcal{X} \subseteq \mathbb{R}$ is an open interval support. The commonly used numerical characteristics of F are the moments

$$\nu_k = \int_{\mathcal{X}} x^k \, dF(x), \quad k = 1, 2, ... \tag{0.1}$$

Particularly, the mean ν_1 is taken as a measure of the central tendency of F and variance $\sigma^2 = \nu_2 - \nu_1^2$ as a measure of the variability of the values around the mean. However, for many simple and frequently used distributions the integrals (0.1) are infinite (c.f. Kendall and Stuart, 1977). An often quoted example is the Cauchy distribution, but there is a large amount of heavy-tailed parametric distributions with support $\mathcal{X} = (0, \infty)$, for which (0.1) converge in a limited range of parameters only. The admissible range of parameters can be slightly extended when using L-moments, the expectations of certain linear combinations of order statistics (Hosking, 1990). However, L-moments are defined for random variables whose mean does exist so that the problem of characterizing the central tendency and variability of heavy-tailed distributions still remains. The sample mean and sample variance (or, in the case of L-moments, the sample mean and one half of the Gini's mean difference statistics) of data samples taken from these distributions characterize neither their 'center' nor the dispersion.

Let m be an integer and $\Theta \subset \mathbb{R}^m$ a space of parameters and $F_\theta, \theta \in \Theta$ a parametric distribution with density $f_\theta(x) = dF_\theta(x)/dx$. Neither moments nor the L-moments can be used for the estimation of parameters of heavy-tailed distributions since it is usually unknown whether the values of the parameters to be estimated lie in the admissible range. On the other hand, the structure of the vector of parameters of parametric families is the consequence of historical development and, quite often, it do not contain any component which could characterize a 'center' and/or the variability of the distribution. The classical inference function, the vector of partial likelihood scores

$$\mathbf{U}(\theta) = \left[\frac{\partial}{\partial \theta_1} \log f_\theta(x), ..., \frac{\partial}{\partial \theta_m} \log f_\theta(x) \right],$$

is a good tool for estimating the parameters from the observed data, but too complicated to offer simple characteristics of their central tendency and variability.

We find that central tendency and dispersion of a large class of continuous distributions can be characterized by the first two moments of a scalar inference function, which we call Johnson score. A main component of Johnson score is the core function, introduced by Fabián (2001) and briefly re-introduced (in a more appropriate notation) in the next section. In the rest we show that the first two moments of the Johnson score exist and provide a meaningful description of distributions, and that their sample versions provide a meaningful description of the observed data.

2. CORE FUNCTION

Let G be distribution with support $\mathbb{R}$ and density g continuously differentiable according to the variable and regular in the sense that

$$\int_{-\infty}^{\infty} \frac{(g'(y))^2}{g(y)} \, dx < \infty. \tag{0.2}$$

By relation

$$Q(y) = -\frac{g'(y)}{g(y)} \tag{0.3}$$

is defined the score function of G with properties

$$EQ = 0 \tag{0.4}$$

and, by (0.2), $EQ^2 < \infty$.

Let $\Theta = \mathbb{R} \times \Theta^{m-1}$ and $\mathcal{G}_\mu$ be a set of distributions G_μ for which $\theta = (\mu, \tilde{\theta})$ where $\mu \in \mathbb{R}$ is the location parameter and $\tilde{\theta} \in \Theta^{m-1}$. The location parameter is the most important parameter of $G_\mu \equiv G_{(\mu,\tilde{\theta})}$ expressing its central tendency. Let the density of G_μ be

$$g_\mu(y) = g(y - \mu).$$

The score function of G_μ,

$$Q_\mu(y) = -\frac{1}{g(y-\mu)} \frac{dg(y-\mu)}{dy} = \frac{\partial}{\partial \mu} \log g_\mu(y),$$

is equal to the likelihood score for location. For distributions with 'full support' $\mathbb{R}$ the score function $Q(y)$ appears to be a suitable scalar inference function, describing the sensitivity of a construction of the 'central point' to the value y.

However, the score function cannot be considered to be an inference function of distributions with support $\mathcal{X} \neq \mathbb{R}$ (with 'partial support'). For instance, the score function of exponential distribution $Q(x) = 1$ and of uniform distribution $Q(x) = 0$. It may be thought that a suitable scalar inference function could be the likelihood score for the most important parameter, but this is not a good idea since it is not clear which of the parameters, if any, of distributions with partial support could represent a measure of their central tendency.

Based on the fifty-year-old idea of Johnson (1949), Fabián (2001) suggested to view any distribution F with partial support $\mathcal{X} = (a, b)$ as transformed '*prototype*', that is, as if it is in form

$$F(x) = G(\eta(x)), \qquad x \in \mathcal{X}, \tag{0.5}$$

where G is supported by $\mathbb{R}$ (a prototype) and $\eta^{-1} : \mathbb{R} \to (a, b)$ is a suitable mapping. It appeared that for many model distributions suits the inverse of the Johnson transformation (Johnson, 1949) adapted for arbitrary support interval,

$$\eta(x) = \begin{cases} x & \text{when} \quad (a,b) = \mathbb{R} \\ \log(x-a) & \text{when} \quad -\infty < a < b = \infty \\ \log \dfrac{(x-a)}{(b-x)} & \text{when} \quad -\infty < a < b < \infty \\ \log(b-x) & \text{when} \quad -\infty = a < b < \infty. \end{cases} \tag{0.6}$$

An interesting characteristic of F was shown to be the transformed score function of the prototype,

$$T(x) = Q(\eta(x)), \qquad x \in \mathcal{X}, \tag{0.7}$$

termed the *core function*. From (0.7) and relation

$$f(x) = g(\eta(x))\eta'(x), \tag{0.8}$$

following from (0.5), a formula

$$T(x) = \frac{1}{f(x)} \frac{d}{dx} \left(-\frac{1}{\eta'(x)} f(x) \right) \tag{0.9}$$

was derived showing that the core function can be determined without reference to its prototype by a special type of differentiating the density according to the variable.

An unusual feature of the core function is that it is 'support-dependent', since $\eta(x)$ is specific for a given support. It follows from (0.7) and (0.6) that core functions of distributions with the most frequently encountered supports are

$$T(x) = \begin{cases} -f'(x)/f(x) & \text{when} \quad \mathcal{X} = \mathbb{R} \\ -1 - xf'(x)/f(x) & \text{when} \quad \mathcal{X} = (0, \infty) \\ -1 + 2x - x(1-x)f'(x)/f(x) & \text{when} \quad \mathcal{X} = (0, 1). \end{cases} \tag{0.10}$$

To show the sense of the core function in a particular case, let us consider a prototype $G_\mu \in \mathcal{G}_\mu$ transformed to $\mathcal{X} = (0, \infty)$. By (0.8), the density of $F_\mu(x) = G_\mu(\eta(x))$ is $f_\mu(x) = g(\eta(x) - \mu)\eta'(x)$. By setting

$$t = \eta^{-1}(\mu), \tag{0.11}$$

f_μ can be written as

$$f_{\eta(t)}(x) = g(u)\eta'(x)$$

where $u = \eta(x) - \eta(t)$. Writing f_t instead of $f_{\eta(t)}$,

$$\frac{\partial}{\partial t} \log f_t(x) = \frac{g'(u)}{g(u)} \frac{\partial u}{\partial t},$$

so that

$$\frac{1}{\eta'(t)} \frac{\partial}{\partial t} \log f_t(x) = Q(u).$$

However, $T_t(x) = Q(u)$ is the core function of $F_t \equiv F_{\eta(t)}$.

The 'Johnson image' of the location of the prototype (0.11) will be called a *Johnson parameter* and considered as expressing the central tendency of the transformed distribution. Function

$$S(x) = \eta'(t)T_t(x) \tag{0.12}$$

is the likelihood score of the transformed distribution for its most important parameter and, perhaps, could be a suitable scalar inference function of distributions with partial support.

EXAMPLE 1

G_μ with support $\mathbb{R}$ and density $g(y - \mu) = e^{y-\mu}e^{-e^{y-\mu}}$ is the prototype of distribution with density

$$f_t(x) = g(\log x - \log t)\frac{1}{x} = \frac{1}{t}e^{-\frac{x}{t}},$$

the exponential distribution. By (0.9), the core function of the induced distribution is

$$T_t(x) = \frac{1}{f_t(x)} \frac{d}{dx} (-x f_t(x)) = \frac{x}{t} - 1,$$

and $S(x) = t^{-1}T_t(x)$ is the likelihood score for t, obtained by a special type of differentiating the density according to the variable.

However, the preceding conclusion does not hold generally since the density of distribution G with support $\mathbb{R}$ need not have the location parameter so that the transformed distribution need not have the Johnson parameter.

EXAMPLE 2

Distribution $G_{p,q}$ with support $\mathcal{X} = \mathbb{R}$ and density

$$g_{p,q}(y) = \frac{1}{B(p,q)} \frac{e^{py}}{(e^y + 1)^{p+q}} \tag{0.13}$$

where B is the beta function, in (Johnson, Kotz and Ballakrishnan, 1995) called a generalized logistic distribution of the fourth type, has parameters $p > 0, q > 0$, neither of which is a location. Its score function is $Q(y) = (qe^y - p)/(e^y + 1)$. $G_{p,q}$ is the prototype of distribution transformed to $\mathcal{X} = (0, \infty)$ with density

$$f_{p,q}(x) = \frac{1}{xB(p,q)} \frac{x^p}{(x+1)^{p+q}}, \tag{0.14}$$

which is the standard form of Pearson type VI distribution, sometimes called the beta-prime distribution (Johnson, Kotz and Ballakrishnan, 1995). None of the parameters p and q is the Johnson parameter. By (0.10), its core function is

$$T(x) = -1 - x\frac{f'_{p,q}(x)}{f_{p,q}(x)} = \frac{qx - p}{x + 1}. \tag{0.15}$$

Our task is to generalize (0.12) for distributions without the Johnson parameter. The solution is given by Definition 1 in the next section.

3. JOHNSON SCORE, JOHNSON MEAN AND JOHNSON VARIANCE

We realized that t (the 'Johnson image' of the location of the prototype) in term $\eta'(t)$ at (0.12) is for a given F_t the *value* of the Johnson parameter for which $T_t(t) = 0$. We thus generalize (0.12) by replacing t by the zero of the core function, i.e., by the 'Johnson image' of the *mode* of the prototype distribution.

Definition 1 *Let F be distribution with interval support $\mathcal{X} \subseteq \mathbb{R}$ and density f continuously differentiable according to the variable. Let $\eta : \mathcal{X} \to \mathbb{R}$ be given by (0.6), $T(x)$ be core function (0.9) and the solution x^* of equation*

$$T(x) = 0 \tag{0.16}$$

be unique. Function

$$S(x) = \eta'(x^*)T(x) \tag{0.17}$$

will be called the Johnson score of distribution F.

Due to (0.7), the solution of (0.16) is unique if the prototype G is unimodal. In cases of multimodal distributions, x^* could be the 'Johnson image' of the coordinate of the 'most important peak' of the density of the prototype - we do not follow this line here. Johnson score is either the usual score function for distributions supported by $\mathbb{R}$ or the likelihood score for the Johnson parameter for distributions with this parameter or a new function in other cases. A meaning of this new function is similar as in previous cases: for a given $x \in \mathcal{X}$, the value $S(x)$ describes the sensitivity of the construction of the measure of central tendency x^* of F to the value x.

EXAMPLE 3

By (0.15), Johnson mean of the beta-prime distribution is $x^* = p/q$, by (0.6 $\eta'(x^*) = 1/x^*$ and the Johnson score is

$$S(x) = \frac{1}{x^*}T(x) = \frac{q}{p}\frac{qx - p}{x + 1}. \tag{0.18}$$

Definition 2 *For integer k and random variable X with distribution F and Johnson score S we define the k-th Johnson score moment by*

$$ES^k = \int_{\mathcal{X}} S^k(x)\, dF(x). \tag{0.19}$$

In the present paper we study only the first two moments (0.19).

Proposition 1 *Let F with Johnson score S have prototype G with score function Q. Then $ES^k = (\eta'(x^*))^k EQ^k$.*

Proof. By (0.17), $ES^k = (\eta'(x^*))^k ET^k$. By (0.7), (0.8) and (0.3),

$$\begin{aligned} ET^k &= \int_a^b T^k(x)f(x)\, dx = \int_a^b Q^k(\eta(x))g(\eta(x))\eta'(x)\, dx \\ &= \int_{-\infty}^{\infty} Q^k(y)g(y)\, dy = EQ^k. \end{aligned}$$

□

Since $\eta'(x^*) > 0$, by Proposition 1 and (0.2)-(0.4), $ES = 0$ and $ES^2 < \infty$. The solution x^* of

$$S(x^*) = 0 \tag{0.20}$$

will be called a *Johnson mean.* Definition 1 thus assigns to any F with differentiable density a number characterizing its central tendency. Johnson score is thus the score for Johnson mean, which may or may not be a parameter of F.

Due to (0.12), for any F_t with Johnson parameter, the mean square of the Johnson score is the Fisher information for this parameter. Cover and Thomas (1991, p.494) termed the value EQ^2 the Fisher information of a distribution, but they in fact consider only distributions supported by $\mathbb{R}$. The value ES^2 generalizes their concept for any regular continuous F.

Definition 3 *Let the assumptions of Definition 1 hold for distribution F with Johnson score S. The value*

$$\omega^2 = (ES^2)^{-1} \tag{0.21}$$

will be called a Johnson variance and its square root ω a Johnson deviation of distribution F.

Note that for distributions with support $\mathcal{X} = (0, \infty)$ by Proposition 1, $ES^2 = \eta'(x^*)ET^2$ and by (0.6), $\eta'(x^*) = 1/x^*$ so that $\omega^2 = (x^*)^2/ET^2$.

By (0.20) and (0.21) we define alternative measures of the central tendency and variability of continuous distributions.

To make clear the role of the modified Johnson transformation (0.6) in Definition 1, it is to say that, principally, a use of any (strictly monotonous continuous) mapping $\xi : \mathcal{X} \to \mathbb{R}$ leads, by using the procedure described above, to a 'central point' $x^*_\xi = \xi^{-1}$(mode of G) which exists, and to a corresponding 'ξ-score' of distribution F with respect to x^*_ξ. Johnson transformation yields 'central points' and 'ξ-scores', which are for many currently used distributions expressed by simple formulas. There are other supporting reasons for choosing the mapping (0.6):

i/ (0.6) is the only transformation under which the prototype of the lognormal distribution is a normal distribution,

ii/ the 'η-score' of the uniform distribution is linear (see the next section).

4. EXAMPLES

Here we present expressions for the Johnson mean and Johnson variance of some frequently used distributions and show that they can serve as plausible measures of central tendency and dispersion of the values around the Johnson mean not only of distributions, the mean and the variance of which may not exist, but of distributions having regular mean and variance as well.

Johnson mean and Johnson variance of distributions from $\mathcal{G}_\mu$ (i.e., with support $\mathcal{X} = \mathbb{R}$ and location μ) is $x^* = \mu$ and the reciprocal Fisher information for μ. As an example, normal distribution $N(\mu, s)$ has score function $Q(x) = (x - \mu)/s^2$ and $\omega^2 = 1/EQ^2 = s^2$, so that its Johnson mean and Johnson variance are the mean and variance. As a less trivial example, consider a generalized Student distribution with density

$$f_{\mu,s,\nu}(x) = \frac{\lambda^{\lambda/2}}{sB(1/2, \lambda/2)} \frac{1}{(\lambda + \xi^2)^{\frac{\lambda+1}{2}}} \tag{0.22}$$

where $\xi = (x - \mu)/s$ (particularly, $f_{\mu,s,1}$ is the Cauchy distribution, having neither a mean nor variance and $f_{0,1,n}$ the Student distribution with n degrees of freedom). Its mean $\nu_1 = 0$ exist if $n > 1$ and variance $\sigma^2 = n/(n-2)$ if $n > 2$, which is particularly in the case of the mean difficult to understand. The distribution has score function $Q(x) = \frac{\lambda+1}{s} \frac{\xi}{\lambda+\xi^2}$ so that the Johnson mean $x^* = \mu$ and the Johnson variance

$$\omega^2 = \frac{1}{EQ^2} = \frac{\lambda + 3}{\lambda + 1} s^2.$$

Fig.1 shows densities of distributions (0.22) with $\mu = 0$, $\lambda = 1, 1.5, 3$ and with s such that $\omega^2 = 3$ for all three distributions. Variances of distributions with $\lambda = 1$ and $\lambda = 1.5$ do not exist, variance of the distribution with $\lambda = 3$ equals to the Johnson variance.

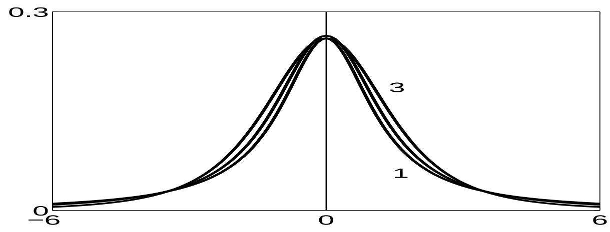

Figure 1. Densities of generalized Student distributions with Johnson variance $\omega^2 = 3$. 1: $\lambda = 1$, 3: $\lambda = 3$, without number: $\lambda = 1.5$.

A generalized logistic distribution with density $g_{p,q}$ given by (0.13) has the second score moment

$$EQ^2 = \frac{pq}{p+q+1}. \qquad (0.23)$$

By (0.23), distribution $g_{k,k}$ with $k = (1+\pi/\sqrt{3})/\pi^2$ has Johnson deviation $\omega = 1/(EQ^2)^{1/2} = \pi/\sqrt{3}$. In Fig.2, $g_{k,k}$ is compared with $g_{1,1}$, the standard deviation of which is $\sigma = \pi/\sqrt{3}$ as well as with density of the standard normal distribution with $\sigma = \pi/\sqrt{3}$. It is apparent that it is the Johnson variance and not the variance of the generalized logistic distribution, which corresponds to the variance of the normal distribution.

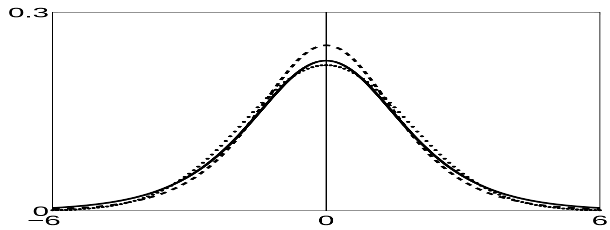

Figure 2. Densities of the generalized logistic distribution with $\omega = \pi/\sqrt{3}$ (full line) and $\sigma = \pi/\sqrt{3}$ (dashed line), and the density of the standard normal with $\sigma = \pi/\sqrt{3}$ (dotted line).

Distributions with support $\mathcal{X} = (0,\infty)$ and Johnson parameter t have Johnson mean $x^* = t$, Johnson score $S(x) = t^{-1}T_t(x)$ equal to the likelihood score for t and Johnson variance

$$\omega^2 = \frac{t^2}{c^2 ET^2},$$

where $c = 1/s$ is the reciprocal scale of the prototype. Some examples are given in Table 1.

Table I: Core function, Johnson mean and Johnson variance of distributions on $(0,\infty)$ with Johnson parameter.

Distribution	$f(x)$	$T(x)$	x^*	ω^2
lognormal	$\frac{c}{\sqrt{2\pi}x}e^{-\frac{1}{2}\log^2(\frac{x}{t})^c}$	$c\log(\frac{x}{t})^c$	t	t^2/c^2
hyperbolic	$\frac{1}{2K_0(\alpha)x}e^{-\frac{\alpha}{2}(\frac{x}{t}+\frac{t}{x})}$	$\frac{\alpha}{2}(\frac{x}{t}-\frac{t}{x})$	t	$K(\alpha)t^2$
Weibull	$\frac{c}{x}(\frac{x}{t})^c e^{-(\frac{x}{t})^c}$	$c((\frac{x}{t})^c - 1)$	t	t^2/c^2
Fréchet	$\frac{c}{x}(\frac{t}{x})^c e^{-(\frac{t}{x})^c}$	$c(1-(\frac{t}{x})^c))$	t	t^2/c^2
log-logistic	$\frac{c}{x}\frac{(x/t)^c}{((x/t)^c+1)^2}$	$c\frac{(x/t)^c-1}{(x/t)^c+1}$	t	$3t^2/c^2$

$K(\alpha) = (\frac{K_2(\alpha)}{K_0(\alpha)} - 1)$, K_ν is the McDonald function.

Fig.3 shows densities and Johnson scores of Weibull distributions with $c = 1$ (exponential distribution), $c = 2$ (Rayleigh distribution) and $c = 3$ (Maxwell distribution). Johnson mean of all these distributions is $x^* = 1$. The means (denoted by stars) are near, in case $c = 1$ equal to x^*.

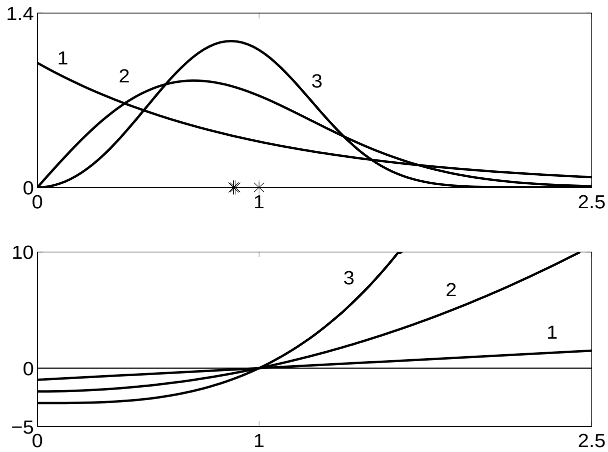

Figure 3. Densities and Johnson scores of Weibull distributions with $t = 1$, $c = 1, 2, 3$. The means $\nu_1(c)$ are denoted by stars, $\nu_1(1) = 1, \nu_1(2) = 0.885, \nu_1(3) = 0.893$.

Fig.4 shows densities of Fréchet distributions with various Johnson means. The variability of values around the Johnson mean is apparently similar to all four distributions. Actually, they have the same Johnson variance $\omega^2 = 1$.

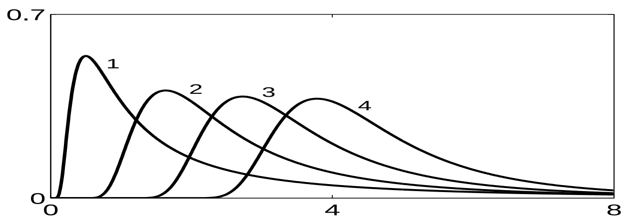

Figure 4. Densities of Fréchet distributions, $t = 1, 2, 3, 4$, $\omega = 1$.

Examples of distributions with support $(0, \infty)$ without the Johnson parameter are given in Table 2.

Table II. Core function, Johnson mean and Johnson variance of distributions $(0, \infty)$ without Johnson parameter.

Distribution	$f(x)$	$T(x)$	x^*	ω^2
gamma	$\frac{\gamma^\alpha}{x\Gamma(\alpha)} x^\alpha e^{-\gamma x}$	$\gamma x - \alpha$	α/γ	α/γ^2
inverse gamma	$\frac{\gamma^\alpha}{x\Gamma(\alpha)} x^{-\alpha} e^{-\gamma/x}$	$\alpha - \gamma/x$	γ/α	γ^2/α^3
beta-prime	$\frac{1}{B(p,q)} \frac{x^{p-1}}{(x+1)^{p+q}}$	$\frac{qx-p}{x+1}$	p/q	$\frac{p(p+q+1)}{q^3}$
Fisher-Snedecor	$\frac{(p/q)^p}{B(p,q)} \frac{x^{p-1}}{(\frac{p}{q}x+1)^{p+q}}$	$\frac{q(x-1)}{x+q/p}$	1	$\frac{p+q+1}{pq}$
Burr XII	$\frac{kcx^{c-1}}{(x^c+1)^{k+1}}$	$\frac{kx^c-1}{x^c+1}$	$(\frac{1}{k})^{1/c}$	$\frac{k+2}{c^2k^{1+2/c}}$

Γ is the gamma function, B the beta function.

Due to the linear core function, the Johnson mean and Johnson variance of the gamma distribution are equal to the mean and variance.

Fig.5 shows densities and Johnson scores of Burr XII distributions. All the plotted distributions have Johnson mean $x^* = 1$ and equal Johnson variances as well. The means, given by $\nu_1(c) = kB(1+1/c, k-1/c)$ for $ck > 1$ are denoted in Fig.5 by stars for $k = 1$. $\nu(1)$ does not exist and other two means do not provide a reasonable description of the central tendency of the plotted distributions as well. vskip2mm

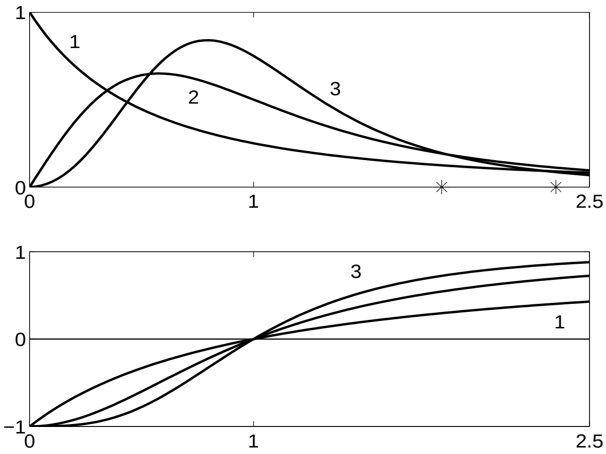

Figure 5. Densities and Johnson scores of Burr XII distributions with $k = 1$, $c = 1, 2, 3$. The means $\nu_1(c)$ are denoted by stars, $\nu_1(1)$ does not exist.

The mean $\nu_1 = p/(q-1)$ and variance

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

of the beta-prime distribution do not exist if $q \leq 1$ and $q \leq 2$, respectively. The Johnson mean and Johnson variance in Table 2 look like the mean and variance 'with corrected denominator'.

Fig.6 shows standard deviation and Johnson deviation of the beta-prime distribution as functions of $1/q$. Whereas σ blows up at $1/q = 1/2$, ω is comparable with the simulated median absolute deviation (Hampel et al., 1986), plotted by the dotted curve.

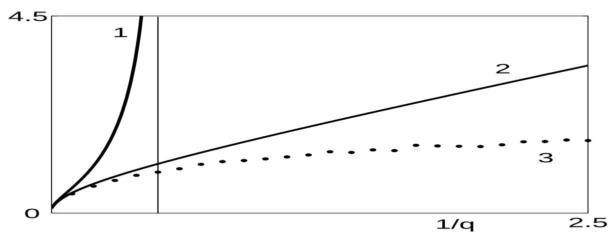

Figure 6. Deviances of the beta-prime distribution. 1 - σ, 2 - ω, 3 - simulated MAD.

Let us examine a more complex case. The Pareto distribution has support $\mathcal{X} = (a, \infty)$ and density

$$f(x) = ca^c/x^{c+1}. \tag{0.24}$$

Its mean $\nu_1 = ca/(c-1)$ and variance $\sigma^2 = ca^2/(c-1)^2(c-2)$ do not exist when $c \leq 1$ and $c \leq 2$, respectively. By (0.6), $\eta'(x) = 1/(x-a)$ so that by (0.9) the core function is

$$T(x) = -1 - (x-a)\frac{f'(x)}{f(x)} = (c+1)\frac{x-a}{x} - 1. \tag{0.25}$$

The Johnson mean is thus

$$x^* = a(c+1)/c. \tag{0.26}$$

Since $ET^2 = c/(c+2)$, $ES^2 = (x^* - a)^2 c/(c+2)$ and the Johnson variance is $\omega^2 = a^2(c+2)/c^3$. The generalized Pareto distribution (Hosking and Wallis, 1987) has density

$$f(x) = k^{-1}(1 - cx/k)^{1/c-1}, \qquad c \neq 0,$$

and $f(x) = k^{-1}e^{-x/k}$ when $c = 0$.

If $c < 0$, the support of the distribution is $\mathcal{X} = (0, \infty)$ so that the core function is

$$T(x) = 1 - xf'(x)/f(x) = \frac{x/k - 1}{-cx/k + 1},$$

Johnson mean $x^* = k,\ ET^2 = 1/(-2c+1)$ and $\omega^2 = k^2(1-2c)$.

If $c > 0$, the support of the distribution is $\mathcal{X} = (0, b)$ where $b = k/c$. Since by (0.6) $\eta'(x) = b/[x(b-x)]$, by (0.9)

$$T(x) = \frac{1}{bf(x)}\frac{d}{dx}[-x(b-x)f(x)] = \frac{x}{k/(c+1)} - 1$$

so that $x^* = k/(c+1)$. Since $ET^2 = 1/(2c+1)$,

$$\omega^2 = \frac{b}{[x^*(b-x^*)]^2 ET^2} = \frac{k^2(1+2c)}{(c+1)^4}. \tag{0.27}$$

In Fig.7 we compare (0.27) for $k = 1$ with the standard deviation $\sigma = k/[(1+c)\sqrt{1+2c}]$ for $c > -\frac{1}{2}$ (Johnson, Kotz and Ballakrishnan, 1995) and with $2l_2$ where $l_2 = k/[(1+c)(2+c)]$ if $c > -1$ is, by Hosking (1990), the 2-nd L-moment. Whereas both 'empirical moments' are increasing to infinity when c approaches to the boundary of the range of validity of the formulas, Johnson deviation increases linearly, behaving like a scale parameter.

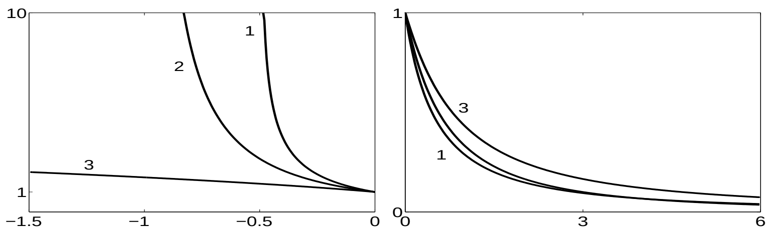

Figure 7. Deviances of the generalized Pareto distribution. 1 - σ, 2 - $2\lambda_2$, 3 - ω.

To the end, consider the beta distribution with support $\mathcal{X} = (0,1)$ and density $f_{p,q}(x) = \frac{1}{B(p,q)}x^{p-1}(1-x)^{q-1}$, which has common prototype (0.13) with the beta-prime distribution. By (0.10), the core function of the beta distribution is

$$T(x) = -1 + 2x - x(1-x)\left[\frac{p-1}{x} - \frac{q-1}{1-x}\right] = (p+q)x - p,$$

from which $x^* = p/(p+q)$. The Johnson score of the beta distribution is thus a linear function bounded on the support and Johnson score of the uniform distribution on $(0,1)$ is $S(x) = 2x - 1$. The Johnson mean of the beta distribution is equal to the mean. By (0.23) and Proposition 1, the Johnson variance

$$\omega^2 = \frac{[(x^*(1-x^*)]^2}{EQ^2} = \frac{pq(p+q+1)}{(p+q)^4}$$

is different from the variance $\sigma^2 = pq/[(p+q+1)(p+q)^2]$. If $p = q \to 0$, $\sigma^2 \to 1/4$, whereas ω^2 grows to infinity, giving a large 'weight' to observations from the ends of the support of U-shaped beta distributions.

ESTIMATES

Let $X_1, ..., X_n$ be random variables i.i.d. according to $F_\theta, \theta \in \Theta, \Theta \subset \mathbb{R}^m$ with unknown θ and $x_1, ..., x_n$ their observed values. The Johnson score of F_θ will be denoted by $S(x;\theta)$. Both the Johnson mean x^* : $S(x^*, \theta) = 0$ and the Johnson variance $\omega^2 = 1/ES^2(\theta)$ are functions of θ and can be constructed from the maximum likelihood estimate $\hat{\theta}_{ML}$ of θ. In what follows AN means 'asymptotically normal'. Since $ES^2(\theta) > 0$, numbers $\hat{x}^*_{ML} = x^*(\hat{\theta}_{ML})$ and $\hat{\omega}^2_{ML} = \omega^2(\hat{\theta}_{ML})$ characterize the 'center' and dispersion of the sample. Their asymptotic behavior can be easily established by using delta method theorem, saying that if $\hat{\theta}$ is $AN(\theta, \sigma^2)$ and $\varphi(\theta)$ is differentiable at θ with $\varphi'(\theta) \neq 0$, $\varphi(\hat{\theta})$ is $AN(\varphi(\theta), [\varphi'(\theta)]^2\sigma^2)$ (Corollary to Theorem A, Serfling, 1980, pp.122).

EXAMPLE 4

Let F be Pareto distribution (0.24) with $a = 1$ and $\hat{c}_{ML}$ be $AN(c, \sigma_c^2)$. By (0.26), the Johnson mean is $x^* = 1 + 1/c$ so that $\hat{x}^*_{ML}$ is $AN(1 + 1/\hat{c}_{ML}, \sigma_c^2/\hat{c}^4_{ML})$.

Unlike the usual moments or the L-moments, the sample versions of Johnson score moments cannot be determined without an assumption about the underlying distribution family. On the other hand, by substituting the empirical distribution function into (0.19), a system of equations

$$\frac{1}{n}\sum_{i=1}^{n} S^k(x_i;\theta) = ES^k(\theta), \qquad k = 1, ..., m, \tag{0.28}$$

appears to be an alternative to the system of maximum likelihood equations for estimation of θ in the whole range of parameters. The estimates $\hat{\theta}_n$ from (0.28) are the 'core moment estimates' (Fabián, 2001), shown to be consistent, asymptotically normal and, in cases of families with bounded core functions, robust and with relative efficiencies near to one.

In the rest we show that the first or two first equations of system (0.28) give for particular distributions simple estimates of the Johnson mean or of both Johnson characteristics. Let us call the estimate $\hat{x}^*_n$ of x^* based on observations $x_1, ..., x_n$ a *sample Johnson mean* and the estimate $\hat{\omega}^2_n$ of ω^2 a *sample Johnson variance.*

EXAMPLE 5

The sample Johnson mean of distributions with Johnson parameter is the maximum likelihood estimate of this parameter. For the distributions from Table 1, $\hat{x}^*_n$ of the lognormal distribution is the geometric mean, $\hat{x}^*_n$ in the case of the Weibull distribution with fixed c is the c-th mean (for exponential distribution with $c = 1$ the arithmetic mean $\bar{x}$) and in the case of the Fréchet distribution with a given c, $\hat{x}^*_n = \hat{t}_n = 1/(n^{-1}\sum_{i=1}^n 1/x_i^c)^{1/c}$, which is for $c = 1$ the harmonic mean $\bar{x}_H$, say. The sample Johnson mean of the hyperbolic distribution is $\hat{x}^*_n = (\bar{x} + \bar{x}_H)^{1/2}$.

We obtain new results when considering distributions without the Johnson parameter. If the structure of the parameters is such that $x^*(\theta)$ is only a function of one parameter or a function of the ratio of two parameters, it is often possible to express $x^*(\theta)$ from the first equation of (0.28) uniquely as a function of observed values. In such a case one can treat x^* as a parameter and write $S(x;x^*)$ instead of $S(x;\theta)$ so that the first equation of (0.28) turns into

$$\sum_{i=1}^{n} S(x_i;x^*) = 0. \tag{0.29}$$

Proposition 2 *Let F_θ satisfy conditions of Definition 1 and let the first equation of (0.28) be written in form (0.29). The solution $\hat{x}^*_n$ of (0.29) is $AN(x^*, \omega^2)$, where ω^2 is given by (0.21).*

Proof. $\hat{x}^*_n$ is a consistent estimate of x^* since $S(x^*, x^*) = 0$ and S is continuous. Random variables $S(X)$ have zero mean and finite variance so that $\hat{x}^*_n$ is $AN(x^*, 1/ES^2)$ according to the Lindeberg-Lévy central limit theorem. The assertion then follows from (0.21). □

EXAMPLE 6

The first equation of (0.28) for the gamma distribution is

$$\sum_{i=1}^{n}(\gamma x_i - \alpha) = 0.$$

Since $x^* = \alpha/\gamma$, $\hat{x}_n^* = \bar{x}$. Similarly, the sample Johnson mean of the inverse gamma distribution is $\bar{x}_H$. From the first equation of (0.28) for the beta-prime distribution,

$$\sum_{i=1}^{n} \frac{qx_i - p}{x_i + 1} = 0,$$

one obtains (since $x^* = p/q$)

$$\hat{x}_n^* = \frac{\sum_{i=1}^{n} \frac{x_i}{1+x_i}}{\sum_{i=1}^{n} \frac{1}{1+x_i}}. \tag{0.30}$$

By (0.25), the first equation of (0.28) for the Pareto distribution is

$$\frac{1}{n}\sum_{i=1}^{n}(1 - a/x_i) = 1/(c+1)$$

so that by (0.26) $\hat{x}_n^* = \bar{x}_H$. The sample mean of the beta distribution is $\hat{x}_n^* = \bar{x}$.

By Proposition 2, Johnson variance is a characteristic of the variability of the underlying distribution and, simultaneously, the asymptotic variance of the estimate of the Johnson mean.

EXAMPLE 7

The sample Johnson variance of the lognormal distribution (Table 2) is $\hat{\omega}_n^2 = \hat{x}_n^2/\hat{c}^2$ where, from the second equation of (0.28), $\hat{c}^2 = n/\sum\limits_{i=1}^{n} \log^2(x_i/\hat{x}_n^*)$. The sample Johnson variance of the gamma distribution is the sample variance. The second equation of (0.28) for the inverse gamma distribution is

$$\frac{1}{n}\sum_{i=1}^{n} (\alpha - \gamma/x_i)^2 = \alpha$$

so that

$$\hat{\omega}_n^2 = \bar{x}_H^2 \, \frac{\bar{x}_H^2 - \bar{x}_{2H}}{\bar{x}_{2H}}$$

where $\bar{x}_{2H} = n/\sum\limits_{i=1}^{n} 1/x_i^2$. The second equation of (0.28) for the beta-prime distribution is

$$\frac{1}{n}\sum_{i=1}^{n} \left(\frac{qx_i - p}{x_i + 1}\right)^2 = \frac{pq}{p+q+1}. \tag{0.31}$$

Multiplying (0.31) by $1/pq$, substituting $p = q\hat{x}_n^*$ and using Table 2,

$$\hat{\omega}_n^2 = \frac{\hat{\rho}\hat{x}_n^*(1+\hat{x}_n^*)^2}{(\hat{\rho}-1)^2}$$

where $\hat{x}_n^*$ is given by (0.30) and

$$\frac{n}{\hat{\rho}} = \frac{1}{\hat{x}_n^*}\sum_{i=1}^{n} \frac{x_i^2}{(x_i+1)^2} - 2\sum_{i=1}^{n} \frac{x_i}{(x_i+1)^2} + \hat{x}_n^* \sum_{i=1}^{n} \frac{1}{(x_i+1)^2}.$$

In a general case, however, it is to estimate the parameters and construct $\hat{\omega}_n^2 = \omega^2(\hat{\theta}_n)$.

In a simulation study, samples of length 100 were generated consecutively from each distribution listed in rows of Table 3, each with values of θ giving $x^*(\theta) = 1$ and $\omega^2(\theta) = 1.2$. Both x^* and ω were estimated under the assumption of either distribution listed in headlines of columns in Table 3. The estimates were obtained according to the simple formulas discussed above except for the Weibull distribution, for which parameters were estimated by the maximum likelihood method. Table 3 summarizes average values of the estimated Johnson means and the Johnson variances over 5000 samples.

Table III: Estimates of Johnson mean and Johnson deviance of some distributions. The true values are $x^* = 1, \omega = 1.118$.

$\hat{x}^*$	gamma	Weibull	lognormal	beta-prime	inv.gamma
gamma	1.000	0.94	0.60	0.49	0.12
Weibull	1.06	1.005	0.64	0.53	0.15
lognormal	1.66	1.66	1.01	1.010	0.63
beta-prime	2.00	1.77	1.008	1.01	0.54
inv.gamma	84.4	4.71	1.70	2.13	1.022
$\hat{\omega}$					
gamma	1.094	1.06	0.81	0.72	0.31
Weibull	1.17	1.108	0.83	0.75	0.39
lognormal	2.04	1.62	1.082	1.09	0.74
beta-prime	3.52	2.00	1.11	1.113	0.82
inv.gamma	187.	8.52	2.32	3.23	1.117

It is apparent from Table 3 that erroneous assumptions often lead to unacceptable estimates (note, however, the similar results obtained under assumptions of the lognormal and beta-prime distributions). By the use of the estimates of the Johnson mean and Johnson variance, it is easy to compare the parametrized by arbitrary ways.

AKNOWLEDGEMENTS

The work was supported by GA ASCR under grant No. 1ET 400300513. The author is very grateful to Igor Vajda for valuable discussions. The author also thanks the associate editor and the referees for helpful suggestions improving the manuscript.

BIBLIOGRAPHY

Cover, T. M. and Thomas, J. A. (1991). *Elements of information theory.* Wiley, New York.

Fabián, Z. (2001). Induced cores and their use in robust parametric estimation. *Commun. Stat., Theory Methods*, **30**, 3, 537-556.

Fabián, Z. (2006). Johnson point and Johnson variance. *Proc. Prague Stochastics 2006,* (eds. Hušková and Janžura), Matfyzpress, 354-363.

Hampel F. R., Rousseew P. J., Ronchetti E. M. and Stahel W. A. (1986). *Robust statistic. The approach based on influence functions.* Wiley, New York.

Hosking J. R. M. and Wallis, J. R. (1987). Parameter and quantile estimation for the generalized Pareto distribution. *Technometrics*, **29**, 339-349.

Hosking J.R.M. (1990). L-moments: Analysis and estimation of distributions using linear combinations of order statistics. *J.R.Statist.Soc.B*, **53**, 105-114.

Johnson, N.L. (1949). Systems of frequency curves generated by methods of translations. *Biometrika*, **36**:, 149-176.

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

Kendall, M. and Stuart, A. (1977). *The advanced theory of statistics.* Fourth edition, Ch. Griffin, London.

Serfling, R. J. (1980). *Approximation theorems of mathematical statistics.* Wiley, New York.