New Measures of Central Tendency and Variability of Continuous Distributions

Fabián, Zdeněk

2006

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. 978

October 2006

Pod Vodárenskou věží 2, 182 07 Prague 8, phone: +4202 6605 3210, 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

Zdeněk Fabián

Technical report No. 978

October 2006

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.

Mathematics Subject Classification Primary 62A01; Secondary 62F01.

Keywords:
Description of distributions; Basic statistics; Score function; Core function; Point estimates

1 Introduction and statement of the problem

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

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

where $\mathcal{X}\subseteq\mathbb{R}$ is an open interval (support). Its commonly used numerical characteristics are the moments

$$\nu_k=\int_{\mathcal{X}} x^k\,dF(x), \quad k=2,... \tag{1.1}$$

Particularly, the mean $m=\nu_1$ 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 (1.1) are infinite. An often quoted example is the Cauchy distribution having neither mean nor variance, but there are many parametric distributions with support $\mathcal{X}=(0,\infty)$ for which (1.1) converge in a limited range of parameters only, for instance the log-logistic, Fréchet, Pareto, log-Cauchy, beta-prime, Fisher-Snedecor and Burr XII distributions. The sample mean and sample variance of data samples taken from these distributions characterize neither a 'center' nor the dispersion of the data.

On the other hand, 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 $f_\theta(x)=dF_\theta(x)/dx$. The classical inference function, the vector of likelihood scores

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

is too complicated to offer simple numerical characteristics. Vector of estimates of the parameters do not often contain any component which could characterize the 'center' and/or the dispersion of the data. We find that the central tendency and dispersion of continuous distributions can be characterized by the first and second moment of a scalar inference function $S(x)$, the moments of which

$$ES^k=\int_{\mathcal{X}} S^k(x)\,dF(x), \quad k=1,2,... \tag{1.2}$$

exist independently of the speed with which the density approaches to zero.

A function of these properties is well-known for distributions supported by $\mathbb{R}$. Let G be such distribution with density g continuously differentiable according to the variable and let

$$Q(x)=-\frac{g'(x)}{g(x)} \tag{1.3}$$

be its score function. Let $\Theta=\mathbb{R}\times\Theta^{m-1}$ and $\mathcal{G}_\mu$ be the set of distributions G_μ for which $\theta=(\mu,\tilde{\theta})$ where $\mu\in\mathbb{R}$ is the location parameter and $\tilde{\theta}\in\Theta^{m-1}$. Here and in the sequel, we did not explicitly indicate a possible dependance on $\tilde{\theta}$. Let G_μ have density

$$g_\mu(x)=g(x-\mu). \tag{1.4}$$

Its score function

$$Q_\mu(x)=-\frac{1}{g(x-\mu)}\frac{dg(x-\mu)}{dx}=\frac{\partial}{\partial\mu}\log g_\mu(x) \tag{1.5}$$

is equal to the likelihood score for location, which is the most important parameter of G_μ expressing its central tendency. We conclude that for distributions with 'full support' $\mathbb{R}$ function

$$S(x)=Q(x) \tag{1.6}$$

appears to be a suitable scalar inference function. Obviously,

$$EQ=0. \tag{1.7}$$

Let us further assume that

$$EQ^2 = \int_{-\infty}^{\infty} \frac{(g'(x))^2}{g(x)}\, dx < \infty. \tag{1.8}$$

Condition (1.8) is weak and corresponds to the usual conditions of regularity.

Relation (1.6) cannot be used, however, in cases of distributions supported by $\mathcal{X} \neq \mathbb{R}$ (with 'partial support'). For instance, for exponential distribution $Q(x) = 1$ and for uniform distribution $Q(x) = 0$. It may be thought that a suitable S could be the likelihood score for the most important parameter, but this is not a plausible idea since it is not clear which of the parameters of distribution with partial support could represent a measure of the central tendency.

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

$$F(x) = G(\eta(x)), \tag{1.9}$$

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

$$\eta(x) = \begin{cases} x & \text{if} \quad (a,b) = \mathbb{R} \\ \log(x-a) & \text{if} \quad -\infty < a < b = \infty \\ \log\dfrac{(x-a)}{(b-x)} & \text{if} \quad -\infty < a < b < \infty \\ \log(b-x) & \text{if} \quad -\infty = a < b < \infty. \end{cases} \tag{1.10}$$

An interesting characteristic of F was shown by Fabián (2001) to be the transformed score function of the prototype

$$T(x) = Q(\eta(x)), \tag{1.11}$$

termed a *core function.* From (1.11) and relation

$$f(x) = g(\eta(x))\eta'(x) \tag{1.12}$$

following from (1.9), a formula

$$T(x) = \frac{1}{f(x)} \frac{d}{dx} \left(-\frac{1}{\eta'(x)} f(x) \right) \tag{1.13}$$

was derived showing that the core function of distribution with differentiable density can be determined without reference to its prototype by - somewhat sophisticated - differentiating of 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 (1.11) and (1.10) that the core functions of the distributions with the most frequent supports are

$$T(x) = \begin{cases} -f'(x)/f(x) & \text{if} \quad \mathcal{X} = \mathcal{R} \\ -1 - xf'(x)/f(x) & \text{if} \quad \mathcal{X} = (0,\infty) \\ -1 + 2x - x(1-x)f'(x)/f(x) & \text{if} \quad \mathcal{X} = (0,1). \end{cases} \tag{1.14}$$

For other (continuous and strictly monotone) mappings $\eta : \mathbb{R} \to \mathcal{X}$ one obtains different 'core functions'. For instance, for distribution with support $\mathcal{X} = (-\pi/2, \pi/2)$ and density

$$f(x) = \frac{1}{\sqrt{2\pi}\cos^2 x} e^{-\frac{1}{2}\tan^2 x},$$

the use of $\eta(x) = \tan x$ leads to a 'core function' given by a simple formula. However, the logarithmic transformation (1.10) is suitable for many frequently used model distributions. There are other supporting reasons: under (1.10) the prototype of the lognormal distribution is the normal distribution and the core function of the uniform distribution is linear.

Consider a distribution with partial support $\mathcal{X} \neq \mathbb{R}$ and with prototype $G_\mu \in \mathcal{G}_\mu$. Let us denote the image of the location of the prototype on $\mathcal{X}$ by

$$t = \eta^{-1}(\mu) \tag{1.15}$$

and call it a *Johnson parameter*. Denote this distribution with parameter space $\Theta' = \mathcal{X} \times \Theta^{m-1}$ by F_t and let $\mathcal{F}_t(\mathcal{X})$ be the set of such distributions. A distributions $F_t \in \mathcal{F}_t(\mathcal{X})$ has the following important property: Let us set $u = \eta(x) - \eta(t)$. By (1.12), (1.4) and (1.15), the density of $F_t \in \mathcal{F}_t(\mathcal{X})$ is $f_t(x) = h(u)\eta'(x)$. Denoting by $T_t(x) = Q_{\eta(t)}(u)$, we have

$$\frac{\partial}{\partial t} \log f_t(x) = \frac{\partial}{\partial t} \log h(u)\eta'(x) = -\frac{1}{h(u)} \frac{dh(u)}{du} \frac{\partial u}{\partial t}$$

so that

$$\frac{\partial}{\partial t} \log f_t(x) = \eta'(t) T_t(x). \tag{1.16}$$

Function

$$S(x) = \eta'(t) T_t(x) \tag{1.17}$$

of distribution $F_t \in \mathcal{F}(\mathcal{X})$ equals to the likelihood score for parameter t. Since t is the image of the location of the prototype, it can be considered as expressing the central tendency of distributions from $\mathcal{F}(\mathcal{X})$.

Example 1.1 Gumbel distribution with support $\mathbb{R}$, density

$$g_\mu(x) = e^{x-\mu} e^{-e^{x-\mu}}$$

and score function $Q_\mu(x) = e^{x-\mu} - 1$ is the prototype of distribution with density

$$f_t(x) = \frac{x}{t} e^{-\frac{x}{t}} \frac{1}{x} = \frac{1}{t} e^{-\frac{x}{t}} \tag{1.18}$$

which is exponential distribution. By (1.11), core function of (1.18) is $T_t(x) = x/t - 1$ and $S(x) = \frac{1}{t} T_t(x)$ equals to the likelihood score for t.

However, the density of distribution G with full support $\mathbb{R}$ does not need to have the location parameter and the transformed distribution $F(x) = G(\eta(x))$ thus does not need to have the Johnson parameter. Actually, many distributions supported by $\mathcal{X} = (0, \infty)$ are not members of the set $\mathcal{F}_t(0, \infty)$.

Example 1.2. Distribution G^{PB} with support $\mathcal{X} = \mathbb{R}$, density

$$g^{PB}(x) = \frac{1}{B(p,q)} \frac{e^{px}}{(e^x + 1)^{p+q}} \tag{1.19}$$

where B is the beta function and with score function

$$Q(x) = \frac{qe^x - p}{e^x + 1} \tag{1.20}$$

has parameters $p > 0, q > 0$, neither of which is the location. (1.19) is the prototype of distribution with support $\mathcal{X} = (0, \infty)$ and density

$$f(x) = \frac{1}{xB(p,q)} \frac{x^p}{(x+1)^{p+q}}, \tag{1.21}$$

which is the standard form of the Pearson Type VI distribution, sometimes called the beta-prime or beta II distribution. Neither of the parameters of (1.21) appears to be the Johnson parameter.

The problem how to generalize (1.17) for distributions without the Johnson parameter is solved by Definition 1 in the next section.

2 Main result: Definition of the Johnson score, Johnson mean and Johnson variance

We realized that t (the 'image' of the location of the prototype) in the term $\eta'(t)$ in (1.17) is for concrete F_t the *value* of the Johnson parameter for which $T_t(t) = 0$. (1.17) can be thus generalized by replacing t by the the zero of the core function, which is the 'image' of the *mode* of the prototype distribution.

Definition 2.1. *Let F be distribution with interval support $\mathcal{X} \subseteq \mathbb{R}$ and density f continuously differentiable according to the variable except possibly a finite number of point. Let $\eta : \mathcal{X} \to \mathbb{R}$ be given by (1.10), $T(x)$ be core function (1.13) and the solution x^* of equation*

$$T(x) = 0 \tag{2.1}$$

be unique. A Johnson score of distribution F is defined by

$$S(x) = \eta'(x^*)T(x). \tag{2.2}$$

Definition 1 adjoins a unique scalar function $S(x)$ to any distribution F with unimodal prototype. $S(x)$ is either the usual score function for F supported by $\mathbb{R}$ or the likelihood score for the Johnson parameter for $F_t \in \mathcal{F}_t(\mathcal{X})$ or a new function in other cases. We suppose that the meaning of new functions is similar as in the previous cases: for a given $x \in \mathcal{X}$, the value $S(x)$ describes the sensitivity of the construction of a measure of central tendency of F from the observed values $(x_1, ..., x_n)$ to the value x.

Example 1.2 (continues). By (1.11), core function of the beta-prime distribution is $T(x) = (qx - p)/(x + 1)$ so that

$$x^* = p/q \tag{2.3}$$

and the Johnson score (2.2) is

$$S(x) = \frac{1}{x^*}T(x) = \frac{q}{p}\frac{qx - p}{x + 1}. \tag{2.4}$$

Since (2.4) is a bounded function, the influence of an additional value $S(x)$ to the average $\sum_{i=1}^{n} S(x_i)$ is limited and this average is not sensitive to outliers in the data.

Proposition 2.1. *Let F have prototype G satisfying condition (1.8) and Johnson score S.* Than $ES = 0$, $ES^2 < \infty$.

Proof. By (1.11), (1.12) and (1.3)

$$\begin{aligned} ES^k &= \int_a^b S^k(x)f(x)\,dx \\ &= (\eta'(x^*))^k \int_a^b Q^k(\eta(x))g(\eta(x))\eta'(x)\,dx \\ &= (\eta'(x^*))^k \int_{-\infty}^{\infty} Q^k(y)g(y)\,dy. \end{aligned}$$

The assertions then follows from (1.7) and (1.8). □

By (1.16), if $F_t \in \mathcal{F}_t(\mathcal{X})$, value ES^2 is the Fisher information for t. Value EQ^2 of distributions with support $\mathbb{R}$ is called by Cover and Thomas (1991, pp.494) the Fisher information of the distribution. Analogically, we can call ES^2 the Fisher information of distribution F (with arbitrary partial support). Furthermore, since $\eta'(x) > 0$, it follows from Definition 1 that

$$S(x^*) = 0. \tag{2.5}$$

Definition 3.2. Let the assumptions of Definition 2.1 hold for distribution F with Johnson score S. The value x^* given by (2.5) will be called a *Johnson mean* and the value

$$\omega^2 = (ES^2)^{-1} \tag{2.6}$$

a *Johnson variance* of distribution F.

Proposition 2.2. *For distributions with support $\mathcal{X} = (0, \infty)$*

$$\omega^2 = \frac{(x^*)^2}{ET^2}. \tag{2.7}$$

Proof. Clear from (2.6), (2.2) and (1.10). □

By relations (2.5) and (2.6), the alternative measures of the central tendency and variability are assigned to any distribution with regular and unimodal prototype. If the prototype is not unimodal, some subsidiary construction for determining its measure of central tendency have to be used. These cases are not considered.

3 Examples

In this section we derive expressions for the Johnson mean and Johnson variance of some frequently used distributions (see Johnson, Kotz and Balakrishnan, 1994, 1995). We show that the Johnson characteristics can serve as measures of central tendency and dispersion of the values around the Johnson mean not only of distributions, the mean and the variance of which are given by formulas valid only in certain ranges of parameters, but even for distributions having the usual mean and variance.

Normal distribution $N(\mu, s)$ has a score function

$$Q(x) = \frac{x-\mu}{s^2}$$

with $x^* = \mu$. Since $EQ^2 = 1/s^2$, Johnson mean and Johnson variance are identical with the mean and variance.

Distributions from $\mathcal{F}_t(0, \infty)$ has Johnson mean $x^* = t$, Johnson score $S(x) = t^{-1}T(x)$ equal to the likelihood score for t and Johnson variance

$$\omega^2 = t^2/\beta^2. \tag{3.1}$$

An example is the **Weibull distribution** with density

$$f(x) = \frac{\beta}{x}\,(x/t)^\beta\, e^{-(x/t)^\beta}$$

and Johnson score

$$S(x) = \frac{\beta}{t}[(x/t)^\beta - 1].$$

Fig. 1 shows densities and Johnson scores of three particular cases of the Weibull distribution with $\beta = 1$ (exponential distribution), $\beta = 2$ (Rayleigh distribution) and $\beta = 3$ (Maxwell distribution). Johnson means of all three distributions are $x^* = 1$, the means are near (and in the case $\beta = 1$ equal) to x^*.

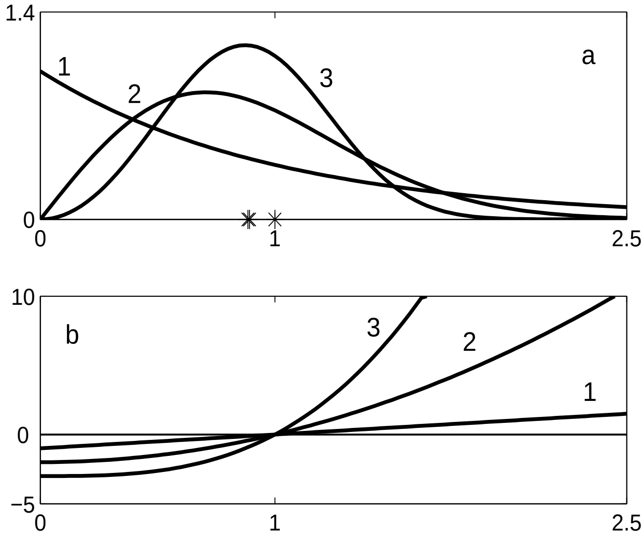

Fig. 1. Densities (a) and Johnson scores (b) of Weibull distributions with $t = 1$, $\beta = 1, 2, 3$. The means $m(\beta)$ are denoted by stars. $m(1) = 1, m(2) = 0.885, m(3) = 0.893$.

Another example of a distribution from $\mathcal{F}_t(0, \infty)$ is the **Fréchet distribution** with density

$$f(x) = \frac{\beta}{x}\,(x/t)^{-\beta}\,e^{-(x/t)^{-\beta}}.$$

The mean $m = t\Gamma(1 - 1/\beta)$ and the variance $\sigma^2 = t^2[\Gamma(1 - 2/\beta) - \Gamma^2(1 - 1/\beta)]$ of the distribution do not exist if $\beta \leq 1$ and $\beta \leq 2$, respectively. Johnson score is

$$S(x) = \frac{\beta}{t}[1 - (x/t)^{-\beta}],$$

so that $x^* = t$ and ω^2 is given by (3.1). Fig. 2 shows the standard deviation and the square root of ω^2 as functions of $1/\beta$. Whereas σ blows up at $1/\beta = 1/2$, ω is comparable with the simulated average MAD (median absolute deviation, Hampel et al. (1986), dotted curve).

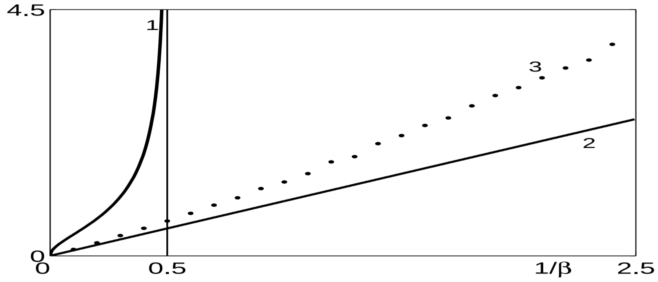

Fig. 2. Square root of Johnson variance of the Fréchet distribution. 1 - σ, 2 - $\omega = t/\beta$, 3 - MAD.

Fig. 3 shows densities of the Fréchet distributions with various Johnson means. The variability of the values around the Johnson mean is apparently similar to all four distributions. Indeed, they have the same Johnson variance $\omega^2 = 1$.

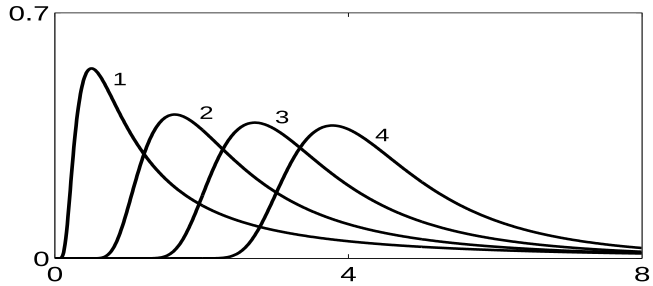

Fig. 3. Densities of Fréchet distributions, $t = 1, 2, 3, 4$, $\omega = 1$.

Let us mention two more distributions from $\mathcal{F}_t(0, \infty)$, the **lognormal** distribution with density

$$f(x) = \frac{\beta}{\sqrt{2\pi}x}\, e^{-\frac{1}{2}\ln^2(x/t)^\beta}$$

and Johnson score $S(x) = \frac{\beta}{t}\ln(x/t)^\beta$ and the **log-logistic distribution** with density

$$f(x) = \frac{\beta}{x}\frac{(x/t)^\beta}{((x/t)^\beta + 1)^2}$$

and Johnson score

$$S(x) = \frac{\beta}{t}\frac{(x/t)^\beta - 1}{(x/t)^\beta + 1}.$$

The mean $m = t[\Gamma(1 + 1/\beta)\Gamma(1 - 1/\beta]$ and variance $\sigma^2 = t^2[\Gamma(1 + 2/\beta)\Gamma(1 - 2/\beta)] - m^2$ of the log-logistic distribution do not exist if $\beta \leq 1$ and $\beta \leq 2$, respectively. Johnson mean of both distributions is $x^* = t$ and Johnson variance is given by (3.1).

Let us now present some distributions with support $(0, \infty)$ which are not members of $\mathcal{F}_t(0, \infty)$.

Gamma distribution with density

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

has parameters $\alpha > 0, \gamma > 0$ neither of which appears to be the Johnson parameter. By (1.11), its core function is $T(x) = \gamma x - \alpha$. The solution of $T(x^*) = 0$ is $x^* = \alpha/\gamma$ so that the Johnson score

$$S(x) = \frac{1}{x^*}T(x) = \gamma\left(\frac{x}{\alpha/\gamma} - 1\right). \tag{3.2}$$

(3.2) is a linear function. Since $ET^2 = \alpha$, $\omega^2 = (x^*)^2/\alpha = \alpha/\gamma^2$. Johnson mean and Johnson variance of the gamma distribution are the usual mean and variance.

Burr XII distribution has density

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

Its mean $m = \alpha B(1 + 1/\beta), \alpha - 1/\beta)$ and variance $\sigma^2 = \alpha^2 B(1 + 2/\beta, \alpha - 2/\beta) - m^2$ do not exist if $\beta\alpha \leq 1$ and $\beta\alpha \leq 2$, respectively. Since

$$T(x) = -1 - x\frac{f'(x)}{f(x)} = \beta\frac{\alpha x^\beta - 1}{x^\beta + 1},$$

Johnson mean $x^* = \alpha^{-1/\beta}$ and Johnson score

$$S(x) = \alpha^{1/\beta}\beta\frac{\alpha x^\beta - 1}{x^\beta + 1}. \tag{3.3}$$

After computing ES^2 we obtain

$$\omega^2 = (ES^2)^{-1} = \frac{1}{\beta^2 \alpha^{2/\beta}} \frac{\alpha + 2}{\alpha}. \tag{3.4}$$

Fig. 4 shows the densities and Johnson scores of the Burr XII distributions. The mean of the distribution with $\beta = 1$ does not exist, the means of other two distributions, denoted by stars, do not provide a reasonable description of their central tendency. All three distribution has the same Johnson mean $x^* = 1$.

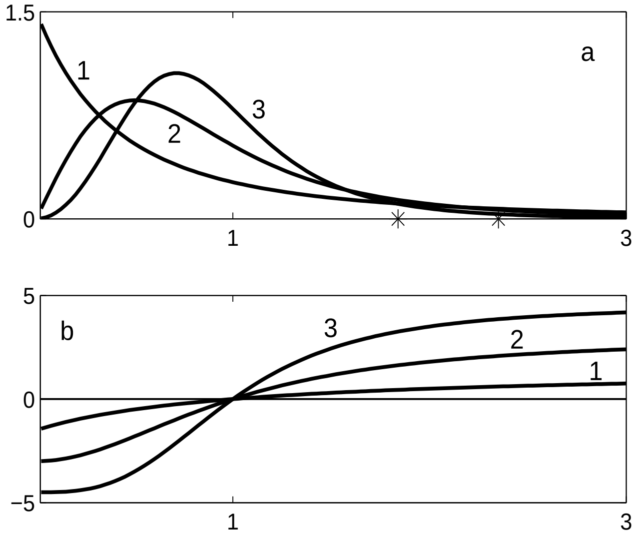

Fig. 4. Densities (a) and Johnson scores (b) of Burr XII distributions with $x^* = 1$, $\beta = 1, 2, 3$. The means $m(\beta)$ are denoted by stars. $m(1)$ does not exist.

Beta-prime distribution with support $\mathcal{X} = (0, \infty)$ and density (1.21) has mean $m = p/(q-1)$ and variance

$$\sigma^2 = \frac{p(p+q+1)}{(q-1)^2(q-2)}, \tag{3.5}$$

which do not exist if $q < 1$ and $q < 2$. Since $x^* = p/q$, its Johnson variance is by (3.10) (given below) and Proposition 1

$$\omega^2 = (x^*)^2/EQ^2 = \frac{p(p+q+1)}{q^3}. \tag{3.6}$$

(3.6) looks like (3.5) with corrected denominator.

Fig. 5 shows the standard deviation and the square root of the Johnson variance of beta-prime distribution as functions of $1/q$. Similarly, as in the case of the Fréchet distribution, standard deviation blows up at $1/q = 1/2$ whereas the Johnson 'standard deviation' is comparable with the simulated MAD.

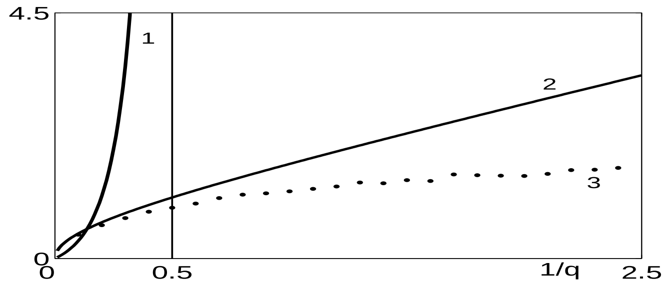

Fig. 5. Square root of Johnson variance of beta-prime distribution. 1 - σ, 2 - ω, 3 - MAD.

Let us present some distributions with other supports.

Pareto distribution has support $\mathcal{X} = (a, \infty)$ and density

$$f(x) = \frac{ca^c}{x^{c+1}}. \tag{3.7}$$

Its mean $m = ca/(c-1)$ and variance

$$\sigma^2 = \frac{ca^2}{(c-1)^2(c-2)} \tag{3.8}$$

do not exist if $c \le 1$ and $c \le 2$, respectively. By (1.10), $\eta'(x) = 1/(x-a)$. By (1.11), the core function of the Pareto distribution is

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

so that the Johnson mean is

$$x^* = a(c+1)/c \tag{3.9}$$

and Johnson score

$$S(x) = \frac{1}{x^*}T(x) = \frac{c}{a}\left[\frac{x-a}{x} - \frac{1}{c+1}\right].$$

Since

$$ET^2 = (c+1)^2 \int_0^\infty \left[\frac{c}{c+1} - \frac{a}{x}\right]^2 \frac{ca^c}{x^{c+1}}\, dx = \frac{(c+1)^2 c}{c+2},$$

Johnson variance of the Pareto distribution is

$$\omega^2 = \frac{(c+2)a^2}{c^3}$$

which, as in the case of the beta-prime distribution, looks like variance (3.8) with corrected denominator.

Prototype beta distribution G^{PB} (1.19) with support $\mathbb{R}$ and score function (1.20) has the second score moment

$$EQ^2 = \frac{pq}{p+q+1} \tag{3.10}$$

and Johnson variance $\omega^2 = (EQ^2)^{-1}$. Let us find from (3.10) a symmetric ($p = q$) prototype beta distribution with $\omega = \pi/\sqrt{3}$: the solution is $q = k = (1 + \pi/\sqrt{3})/\pi^2$). In Fig. 6 are compared $g^{PB}_{k,k}$ with $g^{PB}_{1,1}$ which has $\sigma = \pi/\sqrt{3}$ and with the density of the standard normal distribution with $\sigma = \pi/\sqrt{3}$. Johnson variance of the prototype beta distribution corresponds to the variance of the normal distribution.

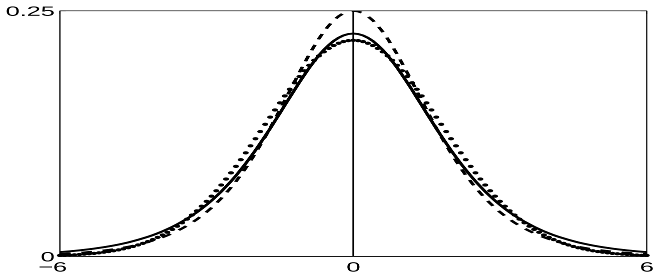

Fig. 6. Densities of prototype beta with $\omega = \pi/\sqrt{3}$ (full line), prototype beta with $\sigma = \pi/\sqrt{3}$ (dashed line) and standard normal with $\omega = \pi/\sqrt{3}$ (dotted line).

Generalized Student distribution has support $\mathbb{R}$ and density

$$f_{\mu,s,\nu}(x) = \frac{\nu^{\nu/2}}{sB(1/2,\nu/2)} \frac{1}{\left(\nu + \left(\frac{x-\mu}{s}\right)^2\right)^{\frac{\nu+1}{2}}}. \tag{3.11}$$

Particularly, $f_{\mu,s,1}$ is the Cauchy distribution, having neither mean nor variance and $f_{0,1,n}$ 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. The distribution has score function

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

Since

$$ES^2 = \left(\frac{\nu+1}{s}\right)^2 \frac{\nu^{\nu/2}}{B(1/2,\nu/2)} \int_{-\infty}^{\infty} \frac{\xi^2 d\xi}{(\nu+\xi^2)^{(\frac{\nu+1}{2}+2)}},$$

we obtain by the use of the table integral

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

the second core moment

$$ES^2 = \frac{(\nu+1)^2}{s^2} \frac{1}{\nu(\nu+3)} \frac{\nu}{\nu+1}$$

and Johnson variance

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

Fig. 7 shows densities of distributions (3.11) with $\mu = 0$, $\nu = 1, 1.5, 3$ and s such that the Johnson variance of all three distributions is $\omega^2 = 3$. Variances of the first two distributions do not exist, the variance of the distribution denoted by 3 is equal to the Johnson variance.

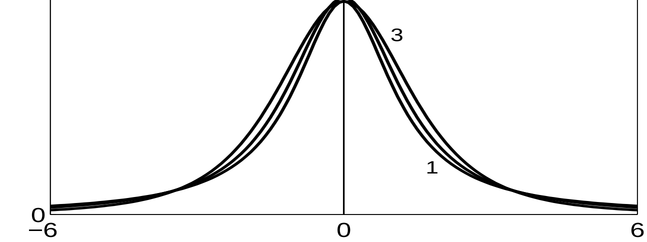

Fig. 7. Densities of generalized Student distributions with Johnson variance $\omega^2 = 3$. 1: $\nu = 1$, 3: $\nu = 3$, without number: $\nu = 1.5$.

Distributions from $\mathcal{F}_t(0,1)$ have Johnson mean $x^* = t(1-t)$, Johnson score $S(x) = (t(1-t))^{-1}T(x)$ equal to the likelihood score for t and Johnson variance

$$\omega^2 = (t(1-t))^2/\beta^2.$$

An example is the **Johnson's U_B distribution**, the prototype of which is normal distribution, with density

$$f(x) = \frac{\beta}{\sqrt{2\pi}x(1-x)} \; e^{-\frac{1}{2}\ln^2(x(1-t)/(1-x)t)^\beta}$$

and core function $T(x) = \beta \ln \left(\frac{x(1-t)}{(1-x)t}\right)^\beta$.

We mention two distributions with support $(0,1)$ which are not members of $\mathcal{F}_t(0,1)$.

Beta distribution with density

$$f_{p,q}(x) = \frac{1}{B(p,q)} x^{p-1}(1-x)^{q-1} \tag{3.12}$$

has common prototype (1.19) with beta-prime distribution. The core function of distribution $T(x) = (p+q)x - p$ has zero $x^* = p/(p+q)$ so that the Johnson score is

$$S(x) = \frac{1}{x^*(1-x^*)} T(x) = \frac{(p+q)^3}{pq}\left(x - \frac{p}{p+q}\right), \tag{3.13}$$

which is a linear function bounded on the support. Johnson mean x^* is equal to the mean, but the Johnson variance

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

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

Triangular distribution with density

$$f(x) = \begin{cases} \frac{2x}{t} & \text{if} \quad 0 < x \le t \\ \frac{2(1-x)}{1-t} & \text{if} \quad t < x < 1. \end{cases}$$

has Johnson mean $x^* = t$ and its core function is, by (1.14),

$$T(x) = \begin{cases} 3x - 2 & \text{if} \quad 0 < x \le t \\ 3x - 1 & \text{if} \quad t < x < 1. \end{cases} \tag{3.14}$$

The Fisher information of triangular distribution is given in Fabián (1997).

4 Estimates

Consider a realization $\mathbf{x} = (x_1, ..., x_n)$ of independent random variables $X_1, ..., X_n$ identically distributed (i.i.d.) according to F with unknown Johnson point x^* and unknown Johnson variance ω^2. Both $x^* = x^*(\theta)$ and $\omega^2 = \omega^2(\theta)$ are functions of θ and can thus be constructed from the maximum likelihood estimate $\hat{\theta}_{ML}$ of θ. Let $\hat{\theta}_{ML}$ be asymptotically normal $AN(\theta, n^{-1}\Sigma^2)$ and $h(\theta)$ be a function continuously differentiable at θ. $h(\hat{\theta})$ is the estimate of $h(\theta)$ consistent and $AN(h(\theta), n^{-1}D\Sigma D')$ where $D = (\partial h(\theta)/\partial\theta_1, ..., \partial h(\theta)/\partial\theta_m)$, cf. Serfling (1980, pp.122). Since $ES^2(\theta) > 0$, the 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 data sample $\mathbf{x}$, the statistical properties of which can be easily determined.

Example 4.1. Let F be Pareto distribution (3.7) with $a = 1$ and $\hat{c}_{ML}$ be $AN(c, \sigma_c^2/n)$. By (3.9), its Johnson mean is $x^* = 1 + 1/c$ so that $\hat{x}^*_{ML}$ is $AN(1 + 1/\hat{c}_{ML}, \sigma_c^2/n\hat{c}^4_{ML})$.

An alternative to the maximum likelihood method is the generalized moment method, i.e., estimating θ as the solution of system

$$\frac{1}{n}\sum_{i=1}^{n} S^k(x_i;\theta) = ES^k, \qquad k = 1, ..., m, \tag{4.1}$$

where S is the Johnson score of distribution F. The method was introduced and partly studied in Fabián (2001). In the rest of the section we show that system (4.1) gives, for particular distributions, estimates of the Johnson mean or of the both Johnson characteristics as algebraic expressions, which are not too worse as $\hat{x}^*_{ML}$ and $\hat{\omega}^2_{ML}$.

Writing Johnson score $S(x)$ of distribution F in form $S(x;x^*)$, let us now study estimates of x^* obtained from the first equation of (4.1),

$$\sum_{i=1}^{n} S(x_i;x^*) = 0. \tag{4.2}$$

Proposition 4.1 *Sample Johnson mean $\hat{x}^*_n$ estimated from equation (4.2) is consistent and asymptotically normal $AN(x^*, \omega^2/n)$.*

Proof. Since S is assumed to be continuous, the consistence of $\hat{x}^*$ is obvious. Random variables $S(x_i, x^*)$ are i.i.d. with zero mean and finite variance ES^2. According to the Lindeberg and Lévy Central limit theorem, $\hat{x}^*_n$ is $AN(x^*, (ES^2)^{-1}/n) = AN(x^*, \omega^2/n)$ by (2.6). □

For distributions $F \in \mathcal{F}_t(\mathcal{X})$, equation (4.2) is by (1.16) identical to the maximum likelihood equation for t. The sample Johnson mean of the lognormal distribution is thus $\hat{x}^* = \hat{t} = (x_1 x_2 \cdots x_n)^{1/n}$, which is the geometric mean, the sample Johnson mean of the Weibull distribution for a given $\beta = p$ is

$$\hat{x}^* = \hat{t} = (\frac{1}{n}\sum_{i=1}^{n} x_i^p)^{1/p},$$

which is the pth mean (for exponential distribution the arithmetic mean), and the sample Johnson mean of the Fréchet distribution for a given $\beta = p$ is

$$\hat{x}^* = \hat{t} = 1/(\frac{1}{n}\sum_{i=1}^{n} 1/x_i^p)^{1/p},$$

which is in case $p = 1$ a harmonic mean. The sample Johnson mean of the log-logistic and other distributions from $\mathcal{F}_t(0, \infty)$ are to be found from (4.2) iteratively.

Let us consider some distributions which are not members of $\mathcal{F}_t$.

The sample Johnson mean of the **beta-prime distribution** (1.21) is given by

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

Since by (2.3) $x^* = p/q$,

$$\hat{x}^* = \frac{\sum_{i=1}^{n} \frac{x_i}{1+x_i}}{\sum_{i=1}^{n} \frac{1}{1+x_i}}. \tag{4.3}$$

In Table 1, average values of $\hat{x}^*$ estimated from 5000 samples randomly generated from the distribution are compared with average values of $\hat{x}^*_{ML} = \hat{p}_{ML}/\hat{q}_{ML}$.

n	$\bar{x}^*_{ML}$	$\bar{x}^*$	$\sigma(\bar{x}^*_{ML})$	$\sigma(\bar{x}^*)$
15	3.208	3.194	0.971	0.975
30	3.104	3.097	0.638	0.644
50	3.050	3.045	0.468	0.475

Table 1. Estimates of the Johnson mean of the beta-prime distribution.

Both estimates are practically the same and have similar variances.

Equation (4.2) for **Pareto distribution** (3.7) is

$$\frac{1}{n}\sum_{i=1}^{n}\left(\frac{c}{c+1}-\frac{a}{x_i}\right)=0. \tag{4.4}$$

Since $x^* = a(c+1)/c$, it follows from 4.4 that

$$\hat{x}^* = (\frac{1}{n}\sum_{i=1}^{n} 1/x_i)^{-1}. \tag{4.5}$$

The Johnson mean of the Pareto distribution is a harmonic mean. Let us compare $\hat{x}^*$ computed from (4.5) with $\hat{x}^*_{ML} = a(\hat{c}_{ML}+1)/\hat{c}_{ML}$, where $\hat{c}_{ML} = n(\sum_{i=1}^{n}\log(x_i/a))^{-1}$. Average values of both estimates for $a = 1$ are given in Table 2.

c	x^*	$\bar{x}^*_{ML}$	$\bar{x}^*$	$\sigma(\bar{x}^*_{ML})$	$\sigma(\bar{x}^*)$
2	1.5	1.498	1.504	0.091	0.098
1	2	2.001	2.025	0.182	0.222
0.5	3	2.997	3.084	0.366	0.541

Table 2. Estimates of the Johnson mean of the Pareto distribution, $a = 1, n = 30$.

The maximum likelihood estimates are somewhat better.

The first two equations (4.1) for **gamma distribution** are

$$\begin{aligned}\sum_{i=1}^{n}(\gamma x_i-\alpha) &= 0\\ \frac{1}{n}\sum_{i=1}^{n}(\gamma x_i-\alpha)^2 &= \alpha,\end{aligned}$$

from which $\hat{x}^* = \alpha/\gamma = \frac{1}{n}\sum_{i=1}^{n} x_i = \bar{x}$ and

$$\hat{\omega}^2 = \alpha/\gamma^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - \bar{x}^2. \tag{4.6}$$

Both the sample Johnson mean and sample Johnson variance can thus be estimated directly from the data without estimating the parameters. In Table 3, average values $\bar{\omega}^2$ estimated by using (4.6) are compared with $\bar{\omega}^2_{ML} = \hat{\alpha}_{ML}/\hat{\gamma}^2_{ML}$ ($\hat{x}^*$ and $\hat{x}^*_{ML}$ are identical).

α	γ	ω^2	$\bar{\omega}^2_{ML}$	$\bar{\omega}^2$	$\sigma(\bar{\omega}^2_{ML})$	$\sigma(\bar{\omega}^2)$
2	2	0.5	0.489	0.483	0.033	0.036
1	1	1	0.996	0.968	0.080	0.091
0.5	0.5	2	2.049	1.907	0.212	0.229

Table 3. Estimates of the Johnson variance of the gamma distribution, n=30.

Both estimates are similar.

The first two equations (4.1) for the **beta distribution** are by (3.13) and Proposition 1

$$\begin{aligned}\sum_{i=1}^{n}(x_i-x^*) &= 0\\ \frac{1}{n}\sum_{i=1}^{n}(x_i-x^*)^2 &= \frac{pq}{(p+q+1)(p+q)^2}\end{aligned}$$

so that $\hat{x}^* = \bar{x}$ and $\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - \bar{x}^2$ ($\hat{\omega}^2$ is expressed by means of the estimates of parameters).

Equation (4.2) for the **triangular distribution** with core function (3.14) is

$$\sum_{i=1}^{k}(3x_i - 2) + \sum_{i=k+1}^{n}(3x_i - 1) = 0$$

from which we obtain

$$3\sum_{i=1}^{n} x_i = 2k + (n - k)$$

and finally

$$\hat{t} = \frac{k}{n} = 3\bar{x} - 1,$$

which seems to be a reasonable estimate of the central tendency.

In a general case, estimates (4.1) appear to be alternatives to the maximum likelihood estimates. According to Fabián (2001), they are slightly worse but often much simpler.

Acknowledgement

The author is indebted to I.Vajda for many helpful discussions. The work was supported by Grant Agency AS CR under grant number IAA1075403.

References

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

Fabián, Z. (1997). Information and entropy of continuous random variables. *IEEE Trans. on Information Theory*, 43: 1080-1083.

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., Rousseeuw, P. J., Ronchetti, E. M. and Stahel, W. A. (1986). *Robust Statistic. The Approach Based on Influence Functions,* Wiley, New York.

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.

Serfling, R. J. (1980). *Approximation Theorems of Mathematical Statistics.* Wiley, New York.