Johnson Score and Characteristics of Distributions

Fabián, Zdeněk
2006

Institute of Computer Science
Academy of Sciences of the Czech Republic

Johnson score and characteristics of distributions

Zdeněk Fabián

Technical report No. 963

February 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 score and characteristics of distributions [1]

Zdeněk Fabián

Technical report No. 963

February 2006

Abstract:

New measures of central tendency and dispersion of continuous probability distributions are introduced. They exist for common distributions and their estimates have necessary statistical properties. Their meaning is explained by means of instructive examples.

Keywords:
basic characteristics, variance, Johnson transform

[1] The work was supported by Grant Agency AS CR under grant IAA1075403.

1 Introduction

Let $\Theta \subset \mathbb{R}^m$ be a space of parameters and $\mathbf{X} = (X_1, ..., X_n)$ a vector of random variables i.i.d. according to Cramér-Rao regular continuous distribution $F_\theta(x)$ with unknown true value $\theta \in \Theta$, which is to be estimated from a realization $(x_1, ..., x_n)$ of $\mathbf{X}$. The standard problem has a standard solution: the maximum likelihood estimate $\hat{\theta}$.

However, there are some distributions, the vector θ of which has no component characterizing a 'center' and/or dispersion of distribution F_θ, so that $\hat{\theta}$ does not offer numbers describing 'center' and/or the dispersion of the observed data. The usual measures of the central tendency and dispersion, the mean and variance, do not usually exist if the distribution is heavy-tailed. Robust estimates of a 'center' and dispersion of data samples, such as MAD and median deviation, do not take into account of prior information.

In the present paper we at first define Johnson score and show that for a certain class of distributions it coincides with likelihood score for the most important parameter. This part accomplishes the approach consisting in viewing distributions with interval support as transformed 'prototypes' with 'full' support $\mathbb{R}$, introduced by Johnson (1949) and generalized by Fabián (2001).

In the rest of the paper, by means of the Johnson score, we suggest new measures of central tendency and the dispersion of continuous probability distributions. They exist for common model distributions and their estimates can be constructed by means of the maximum likelihood estimates without any additional effort.

2 Johnson score

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

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

Definition 1.

A mapping $\eta : (a, b) \to \mathbb{R}$ given by

$$\eta(x) = \begin{cases} x & \text{if} \quad (a,b) = \mathbb{R} \\ \log(x-a) & \text{if} \quad -\infty < a < b = \infty \\ \log \dfrac{(x-a)}{(b-x)} & \text{if} \quad -\infty < a < b < \infty \\ \log(b-x) & \text{if} \quad -\infty = a < b < \infty \end{cases} \tag{1}$$

will be called a *modified Johnson transformation.*

(1) is the Johnson's transformation (Johnson 1949, cf. Johnson, Kotz and Ballakrishnan, 1995) adapted to arbitrary interval support.

Definition 2.

Let $(a, b) \subseteq \mathbb{R}$ and η be given by (1). Let F be an absolutely continuous distribution supported by (a, b) with continuously differentiable density $f(x)$. Let function $T(x)$ be given by formula

$$T(x) = \frac{1}{f(x)} \frac{d}{dx} \left(-\frac{1}{\eta'(x)} f(x) \right). \tag{2}$$

Let the solution x^* of equation

$$T(x) = 0 \tag{3}$$

be unique. The value x^* will be called the *Johnson point* (J-point) and function

$$S(x) = \eta'(x^*) T(x) \tag{4}$$

the *Johnson score* (J-score) of distribution F.

If F has 'full' support $\mathbb{R}$, $\eta'(x) = 1$ and J-score $S(x) = -f'(x)/f(x)$ is the score function. To show what J-score means when a distribution has 'partial' support $(a, b) \neq \mathbb{R}$, we have to study parametric distributions with full support and with location and scale form, transformed to (a, b) by mapping η.

Definition 3.

Let G be distribution supported by $\mathbb{R}$, $\eta : (a, b) \to \mathbb{R}$ be given by (1) and $F = G\eta$ be the transformed distribution on (a, b). G will be called a *prototype* of F.

Let g be the density of G. Then the transformed distribution $F = G\eta$ has density

$$f(x) = g(\eta(x))\eta'(x). \tag{5}$$

Definition 4.

Let the density g of distribution G supported by $\mathbb{R}$ be continuously differentiable and unimodal with

$$g'(0) = 0. \tag{6}$$

Let $\mathcal{G}_{\mu,s} = \{G_{\mu,s} : \mu \in \mathbb{R}, s \in (0, \infty)\}$ be a parametrized family of distributions with parent G and densities in the form

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

Let $\eta : (a, b) \to \mathbb{R}$ be given by (1) and

$$\mathcal{F}_{t,s}^{(a,b)} = \{F_{t,s} : F_{t,s} = G_{\mu,s}\eta, t = \eta^{-1}(\mu)\} \tag{7}$$

be a family of transformed distributions with support (a, b). A set of all families in form (7) is denoted by $\Pi_{t,s}^{(a,b)}$ and called a *set of distributions on* (a, b) *with location and scale prototypes.*

(5) and (7) implies that density of $F_{t,s} \in \mathcal{F}_{t,s}^{(a,b)}$ is

$$f_{t,s}(x) = \frac{1}{s} g\left(\frac{\eta(x) - \eta(t)}{s}\right) \eta'(x) \tag{8}$$

(Proposition 5, Fabián and Vajda, 2003). Parameter

$$t = \eta^{-1}(\mu) \tag{9}$$

of distributions from $\Pi_{t,s}^{(a,b)}$ will be called the *Johnson parameter.*

Now we show that the J-score of a distribution with location and scale prototype is the likelihood score for the Johnson parameter.

Proposition 1.

Let $F_{t,s} \in \Pi_{t,s}^{(a,b)}$. Its J-point is $x^* = t$ and the J-score

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

Proof. Let us set

$$u = \frac{\eta(x) - \eta(t)}{s}.$$

By (4), (2), (8) and the chain rule for differentiation

$$S_{t,s}(x) = \eta'(t)T_{t,s}(x) = \frac{\eta'(t)}{f_{t,s}(x)}\frac{d}{dx}\left(-\frac{1}{\eta'(x)}f_{t,s}(x)\right)$$
$$= \frac{\eta'(t)}{g(u)\eta'(x)}\frac{d(-g(u))}{du}\frac{\partial u}{\partial x},$$

so that

$$S_{t,s}(x) = -\frac{\eta'(t)}{s}\frac{g'(u)}{g(u)}. \tag{10}$$

From (10) and (6) is follows that $S_{t,s}(t) = 0$. The second assertion follows from the fact that, using (8),

$$\frac{\partial}{\partial t}\log f_{t,s}(x) = \frac{1}{f_{t,s}(x)}\frac{\partial}{\partial t}f_{t,s}(x) = \frac{1}{g(u)}\frac{dg(u)}{du}\frac{\partial u}{\partial t} = -\frac{\eta'(t)}{s}\frac{g'(u)}{g(u)}$$

equals to (10).

The J-scores of distributions either i/ supported by $\mathbb{R}$ or ii/ from $\Pi_{t,s}^{(a,b)}$ are thus well-known important functions. J-scores of distributions without Johnson parameter are unknown. We suppose that they could be of similar importance as i/ and ii/ are.

Remark.

The modified Johnson transformation is preferred since i/prototype of the lognormal distribution is the normal one, ii/ score function of the uniform distribution on $(0, 1)$ is linear, iii/ $(\eta(x) - \eta(t))$ is continuous when $b \to \infty$. However, for some distributions, simpler formulas can be obtained by the use of other $\eta : (a, b) \to \mathbb{R}$; for instance, for distributions described by means of trigonometric functions on $(-\pi/, \pi/2)$, a more suitable mapping is $\eta(x) = \tan(x)$.

3 Numerical characteristics of distributions

The J-point of a distribution can be taken as a measure of its centrality.

Proposition 2.

Let F satisfy the assumptions of Definition 2. Then its J-point exists.

Proof. Let $G = F\eta^{-1}$ be a prototype of F and $S_G(y) = -g'(y)/g(y)$. By (2) and (5),

$$T(x) = \frac{1}{g(\eta(x))\eta'(x)}\frac{d}{dx}(-g(\eta(x))) = -\frac{g'(\eta(x))}{g(\eta(x))},$$

that is,

$$T(x) = S_G(\eta(x)). \tag{11}$$

By assumptions, S_G is continuous and for large $y_0 > 0$ it holds that $-g'(-y_0) < 0$ and $-g'(y_0) > 0$, which accomplishes the proof.

Since $\eta'(x) > 0$, the J-point can be defined instead of (3) by equation $S(x^*) = 0$. x^* is unique if the density of the prototype is unimodal. Function $T(x)$ was introduced by relation (11) under the name core function in Fabián (2001).

Now we introduce a number which can serve as a measure of dispersion of a distribution around its J-point.

Definition 5.

Let the assumptions of Definition 2 hold for F with J-score S and let the second J-score moment,

$$I_S = ES^2 = \int_a^b S^2(x)f(x)\,dx, \tag{12}$$

be finite. The value $\sigma_S^2 = I_S^{-1}$ will be called the *Johnson variance* (J-variance) of distribution F.

Taking into account (12) and Proposition 1, for distributions from $\Pi_{t,s}^{(a,b)}$, σ_S is the reciprocal value of Fisher information for the Johnson parameter.

Proposition 3.

Let $F_{t,s} \in \Pi_{t,s}^{(a,b)}$ and $S_{t,s}$ be the corresponding J-score. Let $G_{\mu,s}$, the prototype of $F_{t,s}$, have parent G with density g, score function $S_G(u) = -g'(u)/g(u)$ and Fisher information

$$I_{S_G} = \int_{-\infty}^{\infty} S_G^2(u)g(u)du. \tag{13}$$

Then the J-variance of $F_{t,s}$ is given by

$$\sigma_S^2 = \frac{s^2}{[\eta'(t)]^2 I_{S_G}}. \tag{14}$$

Proof. (14) follows immediately from the expectation of the square of (10).

Corollary.

The J-variance of distributions from $\Pi_{t,s}^{(0,\infty)}$ is $\sigma_S^2 = t^2s^2/I_{S_G}$.

The J-variance of a distribution on $(0,\infty)$ is thus the squared product of the scale of the prototype and of the coordinate of its 'central point', which seems to be a reasonable measure of dispersion of the distribution. The square root of the J-variance we call a standard Johnson deviation.

4 Examples

Example 1.

Let $g(z) = e^z e^{-e^z}$ be density of the parent of Gumbel family $\mathcal{G}_{\mu,s}$. $G_{\mu,s} \in \mathcal{G}_{\mu,s}$ have density in a location and scale form

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

By (8) the densities of transformed distribution $F_{t,s}$ on $(0,\infty)$ are

$$f_{t,s}(x) = \frac{1}{s} g\left(\frac{\ln x - \ln t}{s}\right) \frac{1}{x} = \frac{\beta}{x}\left(\frac{x}{t}\right)^{\beta} e^{-(x/t)^{\beta}} \tag{15}$$

where we denoted $\beta = 1/s$. The transformed family (15) is the Weibull family with Johnson parameter $t = e^{\mu}$. By (4), the J-score of $F_{t,s}$ is

$$S_{t,s}(x) = t^{-1}((x/t)^{\beta} - 1), \tag{16}$$

which equals to the likelihood score for t. The J-point of $F_{t,s}$ is $x^* = t$, $I_{S_G} = 1$ and $\sigma_S^2 = t^2/\beta^2$. We add that other members of $\Pi_{t,s}^{(0,\infty)}$ are for instance the lognormal, Rayleigh, Maxwell, log-logistic, Fréchet, log-Cauchy and the generalized inverse Gaussian distributions (cf. Johnson, Kotz and Ballakrishnan, 1994, 1995).

Example 2.

An example of a prototype distribution without location and scale parameters is a distribution supported by $\mathbb{R}$ with density

$$f_{R_{p,q}}(x) = \frac{1}{B(p,q)} \frac{e^{px}}{(e^x+1)^{p+q}} \tag{17}$$

with shape parameters $p > 0, q > 0$. Let us call it a prototype beta. The densities of the transformed distributions on $(0,\infty)$ and $(0,1)$ are

$$\begin{aligned} f_{(0,\infty)}(x) &= \frac{1}{B(p,q)} \frac{x^{p-1}}{(x+1)^{p+q}} \\ f_{(0,1)}(x) &= \frac{1}{B(p,q)} x^{p-1}(1-x)^{q-1}, \end{aligned}$$

which are densities of the beta-prime and beta distribution, respectively. The mean $\tau = \int_a^b x f_{(a,b)}(x)\ dx$ and variance $\sigma = \int_a^b (x-\tau)^2 f_{(a,b)}(x)\ dx$ of all three beta distributions are given in Table 1a, ψ denotes the logarithmic derivative of gamma function. Both the mean and variance of the beta-prime distribution exist only in the limited range of parameter q.

Table 1 *Mean and variance of beta distributions.*

distribution	*density*	τ	σ^2
prototype beta	f_R	$\psi(p) - \psi(q)$	$\psi'(p) + \psi'(q)$
beta-prime	$f_{(0,\infty)}$	$\frac{p}{q-1}, q > 1$	$\frac{p(p+q+1)}{(q-1)^2(q-2)}, q > 2$
beta	$f_{(0,1)}$	$\frac{p}{p+q}$	$\frac{pq}{(p+q)^2(p+q+1)}$

The score function of prototype beta is

$$S_R(x) = -\frac{f_R'(x)}{f_R(x)} = \frac{qe^x - p}{e^x + 1}. \tag{18}$$

From (2) or from (11) one obtains

$$\begin{aligned} T_{(0,\infty)}(x) &= \frac{qx-p}{x+1} \\ T_{(0,1)}(x) &= (p+q)x - p. \end{aligned}$$

The J-points are thus $x^*_{(0,\infty)} = p/q$ and $x^*_{(0,1)} = p/(p+q)$. Making use of (4) and (12) we obtained the J-variances in Table 2.

Table 2 *J-point and J-variance of beta distributions.*

distribution	*density*	x^*	σ_S^2
prototype beta	f_R	$\ln\frac{p}{q}$	$\frac{p+q+1}{pq}$
beta-prime	$f_{(0,\infty)}$	$\frac{p}{q}$	$\frac{p(p+q+1)}{q^3}$
beta	$f_{(0,1)}$	$\frac{p}{p+q}$	$\frac{p(p+q+1)}{q(p+q)^2}$

σ_S^2 of the beta-prime distribution looks like variance with 'corrected' denominator. The J-point of the beta distribution equals to the mean (as a consequence of a linear J-score), σ_S and σ are different. If $p = q$ and $p \to 0$, it holds that $\sigma \to 1/2$ whereas σ_S of U-shaped beta distributions grows to infinity. By the same procedure one obtains J-points and J-variances of any distribution which is not a member of $\Pi_{t,s}^{(0,\infty)}$, such as the general gamma distribution (Klugmann et al. 1998), gamma, Fisher-Snedecor, Pareto, Lomax, Burr III, Burr XII, Wald, inverse Gaussian, Student and Gompertz distributions (Johnson, Kotz and Ballakrishnan, 1994, 1995).

Example 3.

Remind that for distributions G with support $\mathbb{R}$ the J-score is $S_G(x) = -g'(x)/g(x)$ and I_{S_G} is the Fisher information of the distribution (Cover and Thomas, 1993, p. 494). For normal distribution $\mathcal{N}(\mu, s)$,

$$S_G(x) = \frac{x-\mu}{s} \tag{19}$$

and its $\sigma_S = s$. In case of the Cauchy distribution $\sigma_S = s\sqrt{2}$, for logistic distribution $\sigma_S = s\sqrt{3}$.

Let us find a symmetric prototype beta distribution (17) with $\sigma_S = 1$. The solution of equation $1 = p(p+q+1)/q^3$ for $p = q$ gives value $p = 1+\sqrt{2} = 2.414$. In Fig.1, a surprising coincidence of $f_{R_{1+\sqrt{2},1+\sqrt{2}}}$ with density of the standard normal is apparent.

Let S_1 and S_2 be J-scores of distributions F_1, F_2. A distance between F_1, F_2 was introduced by Fabián and Vajda (2003) as (in terminology of the present paper) the mean square difference of corresponding J-scores,

$$CD(F_1, F_2) = (I_{S_1})^{-1} E_{f_1}(S_1 - S_2)^2. \tag{20}$$

Distance (20) between the prototype beta $F_{R_{p,p}}$ and the standard normal distribution is, by (19) and (18),

$$CD(\Phi, F_{R_{p,p}}) = \frac{1}{2\sqrt{\pi}} \int_{-\infty}^{\infty} \left(x - p\frac{e^x - 1}{e^x + 1} \right)^2 e^{-x^2/2}\, dx. \tag{21}$$

Minimizing (21), one obtains $p = 2.382$, which is in a good agreement with the value obtained above.

Example 4.

Figure 2 shows comparison of σ and σ_S as functions of $1/\beta$ of the Fréchet distribution (16) (Fig. 2a) and of $1/q$ of the beta-prime distribution (17) with $p = q$ (Fig 2b). σ is typically growing exponentially to the limit of the range of its existence, whereas σ_S is increasing linearly or nearly linearly. By dotted lines, average values of the median absolute deviation MAD = median($|x_i -$ median$(x_j)|$) are plotted. They were estimated in a simulation experiment with samples of size 50. Standard deviation of heavy-tailed distributions is of no use even if it exists, whereas σ_S is comparable with MAD in both cases.

5 Characteristics of data samples

Let us denote the J-score of $F_\theta(x), \theta \in \Theta$ by $S_\theta(x)$. We propose to characterize a sample $\mathbf{X}$ taken from F_θ, θ unknown, by the estimate of the J-point of $F_\theta(x)$, that is, by the solution of equation $S_{\hat{\theta}}(\hat{x}^*) = 0$ and by the estimate of the J-variance of $F_\theta(x)$ by $\hat{\sigma}_S^2 = I_S(\hat{\theta})^{-1}$, where $\hat{\theta}$ is the maximum likelihood estimate of θ.

It is well known that if $\hat{\theta}$ is asymptotically normal $(AN(\theta, n^{-1}\Sigma))$, and if h is continuously differentiable at θ, then $h(\hat{\theta})$ is an estimate of $h(\theta)$, consistent and $AN(h(\theta), n^{-1}D\Sigma D')$ where $D = (\partial h(\theta)/\partial\theta_1, ..., \partial h(\theta)/\partial\theta_m)$ (e.g. Corollary to Theorem A, Serfling 1980, pp.122). Since S_θ^{-1} and $1/I_S(\theta)$ are by assumptions continuous and $I_S(\theta) > 0$, we conclude that having the maximum likelihood estimates of parameters of the distribution, one can obtain, without any further efforts, the numbers characterizing the 'center' and dispersion of the data sample together with their statistical properties.

Aknowledgements

The author is indebted to I. Vajda for valuable critical remarks and suggestions.

References

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

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

Fabián Z., Vajda I. (2003), Core functions and core divergences of regular distributions. *Kybernetika* **39**, 1, 29-42

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

Johnson, N.L., Kotz, S., Balakrishnan, N. (1994, 95), *Continuous univariate distributions 1, 2* (Wiley, New York).

Klugmann, S.A., Panjer, H.H. and Willmott, G.E. (1998). *Loss models. From data to decisions.* (Wiley, New York).

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

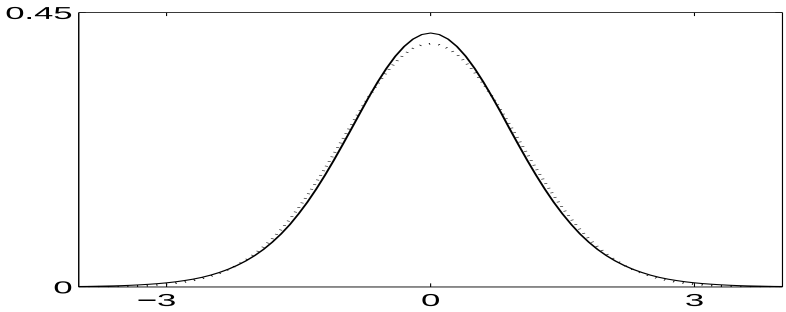

Figure 1. Densities of prototype beta ($p = q = 1 + \sqrt{2}$, full line) and standard normal (dotted line) with $\sigma_S = 1$.

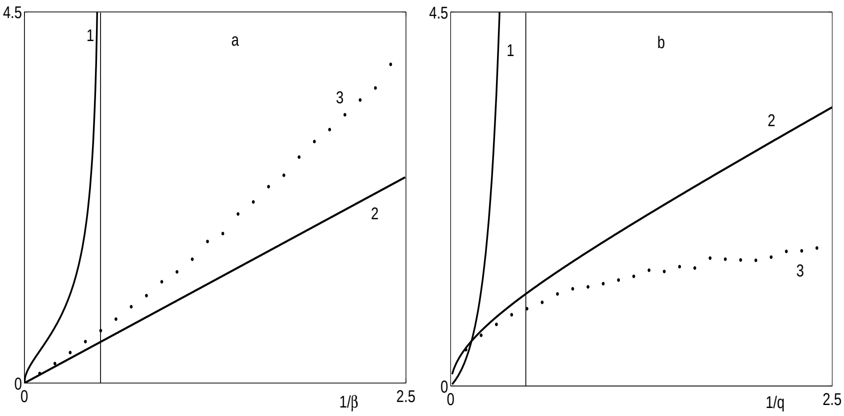

Figure 2. Deviances of the Fréchet (a) and beta-prime (b) distributions 1 - σ, 2 - σ_J, 3 - MAD.