Johnson Information and its use in Multivariate Problems and Time-Series Analysis

Fabián, Zdeněk
2007

Institute of Computer Science
Academy of Sciences of the Czech Republic

Johnson Information and Its Use in Multivariate Problems and Time-Series Analysis

Zdeněk Fabián

Technical report No. 1010

November 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

Johnson Information and Its Use in Multivariate Problems and Time-Series Analysis [1]

Zdeněk Fabián

Technical report No. 1010

November 2007

Abstract:

The author introduced for regular continuous distributions a scalar inference function, Johnson score $S(x)$ [Fabián 2006]. Function $S^2(x)$ has properties of an information function of a distribution and its mean value can be interpreted as the Fisher information for a 'center' of the distribution. In this report we discuss the possible use of this information function in multivariate problems and time series analysis.

Keywords:
Johnson score, correlation, regression, spectral density

[1] The research reported in this paper has been supported by the grant No. ME701 of the Czech Ministry for Education, and partially supported by the Institutional Research Plan AVOZ10300504.

1 Johnson score

It was shown in [2] and [3] that a continuous probability distribution F with interval support $\mathcal{X} \in \mathbb{R}$ can be characterized, besides the distribution function $F(x)$ and density $f(x)$, by its Johnson score, defined as follows. Define mapping $\eta : \mathcal{X} \to \mathbb{R}$ by

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

(1.1) is the *Johnson transformation* [4] adapted for arbitrary interval support $\mathcal{X} = (a, b) \in \mathbb{R}$. Further, define *Johnson score* of F with f continuously differentiable according to $x \in \mathcal{X}$ by

$$T(x) = \frac{1}{f(x)} \frac{d}{dx} \left(-\frac{1}{\eta'(x)} f(x) \right). \tag{1.2}$$

The philosophy behind this definition is the following. Any distribution F with interval support $\mathcal{X} \neq \mathbb{R}$ is viewed as a transformed '*prototype*' G with support $\mathbb{R}$, that is, its distribution function is

$$F(x) = G(\eta(x)) \tag{1.3}$$

where η is given by (1.1). Denoting by g the density of G, the density of F is

$$f(x) = g(\eta(x))\eta'(x), \qquad x \in \mathcal{X}, \tag{1.4}$$

where $\eta'(x) = d\eta(x)/dx$ is the Jacobian of the transformation. Let us denote the score function of G by Q,

$$Q(y) = -\frac{g'(y)}{g(y)}. \tag{1.5}$$

By setting $y = \eta(x)$, we obtain from (1.2) and (1.4)

$$T(x) = \frac{1}{g(y)\eta'(x)} \frac{d}{dy}(-g(y)) \frac{dy}{dx} = Q(\eta(x)). \tag{1.6}$$

The Johnson score of distribution F with 'partial support' is the transformed score function of its prototype.

It follows from (1.2) and (1.1) that the Johnson score of distributions supported by $\mathbb{R}$ is the score function (1.5). Consider now a particular case of a distribution $F_t(x)$ having prototype $G(y - \mu)$ with location parameter $\mu \in \mathbb{R}$, the parameter t of which is given by

$$t = \eta^{-1}(\mu). \tag{1.7}$$

Let us call the 'Johnson image' t of the location of the prototype a *Johnson parameter*. Denote by $f(x; t)$ the density and $T(x; t)$ the Johnson score of such F_t. It was proven (Theorem 1 in [1]) that

$$\frac{\partial}{\partial t} \log f(x; t) = \eta'(t) T(x; t), \tag{1.8}$$

i.e., the Johnson score of a distribution with Johnson parameter is proportional to the likelihood score $\frac{\partial}{\partial t} \log f(x; t)$ for this parameter. In a general case, the density of the prototype does not need to have the location parameter and the transformed distribution does not need to have the Johnson parameter. For such distributions, the Johnson score is yet an unknown function.

A unique solution x^* of equation

$$T(x) = 0 \tag{1.9}$$

will be called a *Johnson mean* of distribution F. Due to (1.6), the solution of (1.9) is unique if the prototype G of distribution F is unimodal and x^* is the 'Johnson image' of the mode of the prototype.

Confining ourselves to these distributions, the Johnson mean characterizes the typical value of the heavy-tailed distributions without mean, and is a value near the mean of the light-tailed distributions.

To make clear the role of the modified Johnson transformation (1.1), it is to say that, principally, a use of any increasing injective 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 can be used instead of the mean), and to a corresponding 'ξ-score' of distribution F with respect to x^*_ξ. The Johnson transformation yields the 'central points' and 'ξ-scores', which are for many currently used distributions mathematically convenient. There are, however, other supporting reasons for choosing the mapping (1.1):

i/ (1.1) 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.

On the other hand, for distributions with support $\mathcal{X} = (-\pi/2, \pi/2)$ with densities expressed by means of trigonometric functions, a more convenient transformation is $\xi(x) = \tan x$.

2 Johnson information

Let X be random variable with distribution F_t where t is the Johnson parameter, with support $\mathcal{X}$, density f and Johnson score T, and let $x \in \mathcal{X}$ be the observed value of X. We assert that the value $S^2(x)$, where

$$S(x;t) = \eta'(t)T(x;t), \tag{2.1}$$

expresses the information about the Johnson mean t, carried by x.

Reasons supporting this assumption are as follows. Let F be given by (1.3). The term $\eta'(x)$ in (1.4) is common to all distributions with support $\mathcal{X}$. All the information contained in X is thus condensed in term $g(\eta(x))$. The least informative point of distribution F is therefore the solution of $\frac{d}{dx} g(\eta(x)) = 0$. By (1.2), this solution is x^* since, by (1.4), $g(\eta(x)) = \frac{1}{\eta'(x)} f(x)$.

$S^2(x;t)$ is thus a non-negative function attaining its minimum in the least informative point of the distribution. Distributions with densities quickly decaying to zero have unbounded Johnson scores so that the value $S^2(x;t)$ for x far away from x^* is high. Distributions with heavy-tailed densities have bounded Johnson scores so that $S^2(x;t)$ has much less influence. Finally, by (2.1) and (1.8), for distributions with Johnson parameter t,

$$ES^2 = \int_{\mathcal{X}} \left(\frac{\partial}{\partial t} \log f(x;t) \right)^2 f(x;t)\, dx \tag{2.2}$$

is the Fisher information for t.

We generalize (2.2) for arbitrary F, not necessarily having the Johnson parameter.

Definition 1 *Let T be the Johnson score and x^* the Johnson mean of distribution F with support $\mathcal{X}$. Set*

$$S(x) = \eta'(x^*)T(x). \tag{2.3}$$

The value

$$I^* = ES^2 = \int_{\mathcal{X}} S^2(x) f(x)\, dx, \tag{2.4}$$

will be called a Johnson information.

For distributions with Johnson parameter t, I^* is the Fisher information for this parameter. In the general case, I^* is the mean information carried by a sample from F about its Johnson mean.

In Table 1 we present the densities and Johnson scores of some distributions together with their Johnson means and Johnson information. The support of all distributions is $\mathcal{X} = (0, \infty)$ except the 'prototype beta' (the prototype of the beta-prime distribution) with support $\mathbb{R}$.

Table 1. Johnson score, Johnson mean and Johnson information of some distributions.

distribution	$f(x)$	$T(x)$	x^*	I^*
prototype beta	$\frac{1}{B(p,q)}\frac{e^{px}}{(e^x+1)^{p+q}}$	$\frac{qe^x-p}{e^x+1}$	$\log p/q$	$\frac{pq}{p+q+1}$
lognormal	$\frac{\beta}{\sqrt{2\pi}x}e^{-\frac{1}{2}\log^2(\frac{x}{t})^\beta}$	$\beta\log(x/t)^\beta$	t	β^2/t^2
Weibull	$\frac{\beta}{x}(\frac{x}{t})^\beta e^{-(\frac{x}{t})^\beta}$	$\beta[(x/t)^\beta-1]$	t	β^2/t^2
gamma	$\frac{\gamma^\alpha}{x\Gamma(\alpha)}x^\alpha e^{-\gamma x}$	$\gamma x-\alpha$	α/γ	γ^2/α
inverse gamma	$\frac{\gamma^\alpha}{x\Gamma(\alpha)}x^{-\alpha}e^{-\gamma/x}$	$\alpha-\gamma/x$	γ/α	α^3/γ^2
beta-prime	$\frac{1}{xB(p,q)}\frac{x^p}{(x+1)^{p+q}}$	$\frac{qx-p}{x+1}$	p/q	$\frac{q^3}{p(p+q+1)}$

Fig. 1 shows the information functions of the exponential distribution (Weibull with $\beta = 1$), and of the lognormal distribution with $\beta = 1$. They are both the distributions with the Johnson parameter (here is $t = 1$). The third function is $S^2(x)$ of the beta-prime distribution, which is a distribution without Johnson parameter. We set $p = q = 1.5$ so that the Johnson variance of all three distributions is $\omega^2 = 1$. Observations near zero in the lognormal model and large observations in the exponential model carry a large amount of information about the Johnson mean $x^* = 1$.

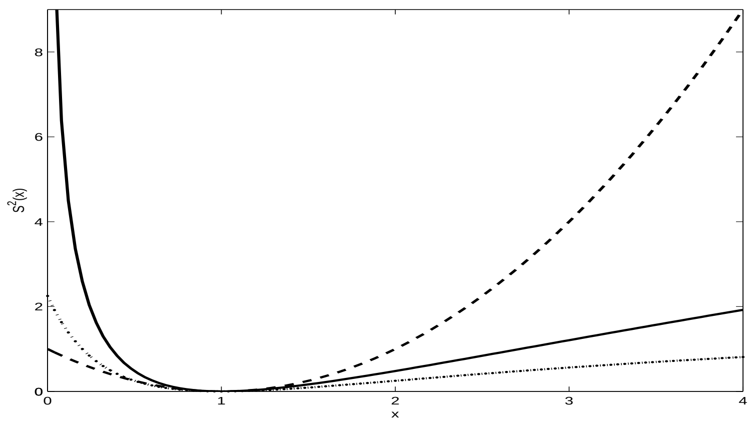

Figure 1. Information functions of the lognormal (full line), exponential (dashed line) and beta-prime (dotted line) distribution.

3 Linear regression with non-Gaussian residuals

Consider a couple (X, Y) where X is non-random and let

$$Y_i = \alpha_0 + \alpha_1 X_i + \epsilon_i, \quad i = 1, ..., n \tag{3.1}$$

where ϵ is a random variable with distribution F_ϵ. We estimate coefficients α_0 and α_1 such as to minimize the difference between the mean information of the sample and the theoretical Johnson information for F_ϵ. We obtain

$$\zeta(\epsilon) = \frac{1}{n}\sum_{i=1}^{n} T_\epsilon^2(\epsilon_i) - ET_\epsilon^2 = min. \tag{3.2}$$

where $\epsilon_i = y_i - (\alpha_0 + \alpha_1 x_i)$ and T_ϵ is the Johnson score of distribution F_ϵ. Differentiating (3.2) according the parameters α_0, α_1, one obtains a system of equations

$$\frac{\partial}{\partial \alpha_0}\zeta(\epsilon) = \sum_{i=1}^{n} T(\epsilon_i)T'(\epsilon_i) \quad = \quad 0$$

$$\frac{\partial}{\partial \alpha_1}\zeta(\epsilon) = \sum_{i=1}^{n} T(\epsilon_i)T'(\epsilon_i)x_i \quad = \quad 0.$$

Fig.2 shows points (x_i, y_i) distributed according (3.1) with ϵ distributed according the prototype beta distribution. The dashed line is the regression line under the assumption of the normal distribution, and full line under the correct assumption of the prototype beta. The latter is similar to the result of robust regression (function *robustfit*, MATLAB) marked by the crosses.

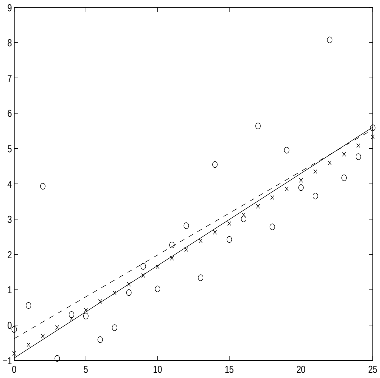

Figure 2. Linear regression with prototype beta residuals.

4 Johnson mutual information

Let T_X^2 and T_Y^2 be the information functions of random variables X and Y with densities f_X and f_Y, respectively. The function $S_X S_Y$ apparently measures a degree of an association between X and Y. Let us call it a Johnson mutual information. Having a sample $(x_i, y_i), i = 1, ..., n$ taken from (X, Y), as an empirical measure of this association one could consider a *sample Johnson mutual information*

$$\hat{I}_{XY}^* = \frac{\sum_{i=1}^{n} T_X(x_i)T_Y(y_i)}{\sqrt{\sum_{i=1}^{n} T_X^2(x_i) \sum_{i=1}^{n} T_Y^2(y_i)}}. \tag{4.1}$$

We take (4.1) as a correlation coefficient of arbitrarily distributed random variables. Since the Johnson score of the standard normal distribution is $T(x) = x$, in the case of normally distributed X and Y $\hat{I}_{XY}^*$ is the usual correlation coefficient.

In the simulation experiment, samples $(x_i, y_i), i = 1, ..., 16$ were generated from (X, Y), where $Y = 0.35X + 0.65Z$ and where X and Z were independent random variables with inverse gamma distribution. The 'mutual information coefficient' (4.1) was computed under five different assumptions on the underlying marginal distributions:

f_X, f_Y	gamma	Weibull	lognormal	beta-prime	inv.gamma
$\hat{I}_{XY}^*$	0.12	0.14	0.29	0.42	0.53

The generated samples (x_i, y_i), and $(T(x_i), T(y_i))$ where T is the Johnson score of the beta-prime distribution are plotted in Fig.3. I^* of the gamma distribution is the usual correlation coefficient; its estimated value is influenced by the 'outlier' at about (1,5.8) in the left subplot. The Johnson mutual information of the beta-prime distribution does not "see" this point as outlying (right).

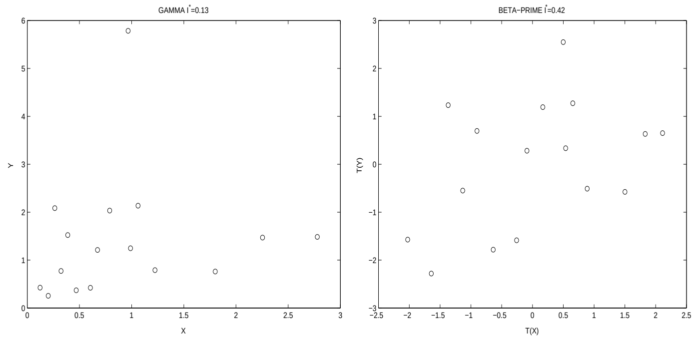

Figure 3. Beta-prime distribution: (x_i, y_i) (left) and $(T(x_i), T(y_i))$ (right).

5 Johnson score of time series

An AR time series $\{x_i\}$ with errors possessing beta-prime distribution is plotted in the upper subplot of Fig.4. Its spectral density estimate is in Fig.5 (dashed line). The Johnson score of $\{x_i\}$, the time series $\{T(x_i)\}$, where T is the Johnson score of the beta-prime distribution, is plotted in the lower subplot of Fig.4 and the corresponding spectral density estimate in Fig.5 (full line). This non-linear transformation of the original time series apparently retains the frequency content but suppresses the influence of the parts of the signal which spectral estimation methods "see" as outliers.

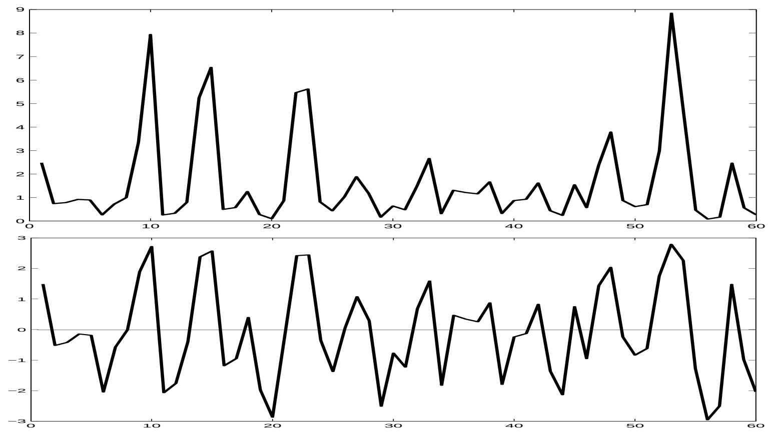

Figure 4. Time series with beta-prime distribution (top) and its Johnson score (bottom).

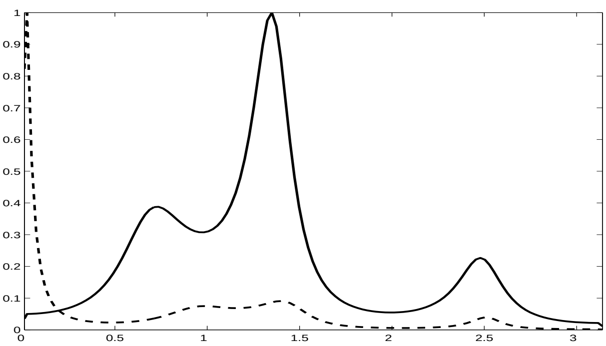

Figure 5. Spectral densities of the original time series (dashed line) and of its Johnson score (full line).

Acknowlegments. The research reported in this paper has been supported by the grant No. ME701 of the Czech Ministry for Education, and partially supported by the Institutional Research Plan AVOZ10300504.

References

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

2 Fabián Z. (2006). Johnson mean and Johnson variance. In *Proc. of Prague Stochastics'2006* (Janžura, M. and Hušková, M., Eds.) Matfyzpress 354–363.

3 Fabián Z. (2008). New measures of central tendency and variability of continuous distributions. To appear in *Commun. in Statist.-Theory Methods*, 2.

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

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