Scalar Score Function and Score Correlation

Fabián, Zdeněk
2010

Institute of Computer Science
Academy of Sciences of the Czech Republic

SCALAR SCORE FUNCTION AND SCORE CORRELATION

Zdeněk Fabián

Technical report No. 1077

August 2010

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

SCALAR SCORE FUNCTION AND SCORE CORRELATION[1]

Zdeněk Fabián

Technical report No. 1077

August 2010

Abstract:

After a short and clear re-introduction of the recent concept of the scalar score, we introduce and study a distribution-dependent correlation coefficient based on it. Properties of the new measure of association of continuous random variables are compared with those of the Pearson, Kendall and Spearman correlation coefficients.

Keywords:
score function; correlation; rank correlation coefficient; heavy tails;

[1]The research presented in the paper was supported by projects AV0Z10300504 and GACR 205/09/1079.

1 INTRODUCTION

The basic inference function of classical statistics is the score function, the generic form of the maximum likelihood score statistic. In cases of vector parameter it is a vector function, too complex to be simply used as an inference function in statistical tasks other than the estimation of the parameters. For instance, the Pearson correlation coefficient of random variables X and Y, constructed from 'pure' data not adapted to marginal distributions of X and Y, is unable to make clear which part of the dependence of X and Y stems from the real dependence, and which part stems from the properties of marginals.

The scalar score function is not as a detailed description of the distribution as the score function, but it is reflecting main features of the distribution and, being scalar even in cases of vector parameter space, it is easily applicable in various inference problems. For particular classes of distributions with parameter expressing the central tendency, the new function equals to the score function for this parameter. In other cases it is a yet unknown function which has the sense of the 'score function for the center of the distribution'.

The introduction of the scalar score function, its properties and examples of its use, including definitions of new measures of central tendency and variability of probability distributions and methods of estimation of their sample counterparts, are described in three papers Fabián (2001)-Fabián (2009) published in this journal. However, in these papers we successively used somewhat different notation and terminology, reflecting author's increasing understanding of the problem and suggestions of reviewers. To make clear the basic ideas of the present paper, as well as of the foregoing ones, we describe shortly but completely the whole procedure of the construction of the scalar score function (Section 2) and add a short summary of the main results obtained up to now (Section 3). Section 4 contains an illustrative example. A new parametric measure of association between two random variables based on the scalar score, the score correlation coefficient suggested by Fabián (2009b), is introduced in Section 5, together with comparisons of the new measure with the Pearson correlation coefficient and with Kendall and Spearman rank correlation coefficients by means of simulation examples for various distributions. At the end we discuss interesting results obtained in cases of heavy-tailed distributions.

2 INTRODUCING THE SCALAR SCORE

Let G be a location distribution with support $\mathcal{X} = \mathbb{R}$ and density in the form $g(x-\mu)$ with location parameter $\mu \in \mathbb{R}$. If g is unimodal, μ indicates the position of the mode of the density. The score function for μ is

$$\frac{\partial}{\partial \mu} \log g(x-\mu) = S_G(x-\mu), \tag{2.1}$$

where

$$S_G(x) = -\frac{g'(x)}{g(x)} \tag{2.2}$$

is a function obtained by differentiating the density with respect to the variable. Given data $\mathbf{x} = (x_1, ..., x_n)$, $\sum_{i=1}^{n} S_G(x_i - \mu)$ is the likelihood score for location.

While the score for location is the basic inference function, function S_G is usually not studied; the reason is that for distributions with support $\mathcal{X} \neq \mathbb{R}$ it exhibits a not acceptable behavior. We call function (2.2) the *scalar score* on $\mathcal{X} = \mathbb{R}$. Its general parametric version $S_G(x;\theta)$ we suggest as an inference function for distributions with support $\mathbb{R}$. Since the solution x^* of equation $S_G(x;\theta) = 0$ is for unimodal distributions the mode, S_G can be viewed as the score function for the mode.

Let Y be random variable with location distribution G. Set $\eta(x) = \log x$. The 'log-location distribution' (Marshall and Olkin, 2007) F of random variable $X = \eta^{-1}(Y)$ with support $\mathcal{X} = (0, \infty)$ has density

$$f(x;\tau) = g(u)\eta'(x), \tag{2.3}$$

where g is the density of the 'prototype' G and $u = \eta(x) - \eta(\tau)$, where

$$\tau = \eta^{-1}(\mu) \tag{2.4}$$

is called a 'log-location' parameter. By (2.3) and the chain rule for differentiation,

$$\frac{\partial}{\partial \tau} \log f(x;\tau) = \frac{1}{g(u)\eta'(x)} \frac{\partial}{\partial \tau}(g(u)\eta'(x)) = S_G(u)\eta'(\tau). \tag{2.5}$$

S_G with transformed variable can be rewritten using (2.3) and (2.2) as

$$S_G(u) = -\frac{\eta'(x)}{f(x;\tau)} \frac{d}{dx} \left(\frac{1}{\eta'(x)} f(x;\tau) \right) \left(\frac{du}{dx} \right)^{-1}. \tag{2.6}$$

The score function for τ is thus

$$\frac{\partial}{\partial \tau} \log f(x;\tau) = T(x;\tau)\eta'(\tau), \tag{2.7}$$

where $T(x;\tau)$ is the 'log-location' version of function

$$T(x) = -\frac{1}{f(x)} \frac{d}{dx} \left(\frac{1}{\eta'(x)} f(x) \right), \tag{2.8}$$

called the *transformation-based score* or shortly the *t-score*. The score function for τ is thus decomposed into product of two terms obtained without need of differentiating with respect to the parameter. The likelihood score for τ is $\eta'(\tau) \sum_{i=1}^{n} T(x_i;\tau)$.

Example 2.1 The standard exponential distribution with density $f(x;\tau) = \frac{1}{\tau} e^{-x/\tau}$ has t-score $T(x;\tau) = x/\tau - 1$. By (2.7), the score function for τ is $\frac{1}{\tau}(x/\tau - 1)$.

Relation (2.7) can be generalized for distributions with general interval support $\mathcal{X} \subseteq \mathbb{R}$ and various one-to-one mappings $\eta : \mathcal{X} \to \mathbb{R}$. For comparison of t-scores of different distributions, it is necessary to use consistently one concrete mapping for a given $\mathcal{X}$. To be consistent with the class of log-location distributions, we set

$$\eta(x) = \begin{cases} \log(x-a) & \text{if} \quad \mathcal{X} = (a,\infty) \\ \log \dfrac{(x-a)}{(b-x)} & \text{if} \quad \mathcal{X} = (a,b). \end{cases} \tag{2.9}$$

Under the term t-score we thus understand (2.8) with η given by (2.9).

The t-score of a general (and in obvious sense regular) distribution $F(x;\theta)$ is function

$$T(x;\theta) = -\frac{1}{f(x;\theta)} \frac{d}{dx} \left(\frac{1}{\eta'(x)} f(x;\theta) \right). \tag{2.10}$$

However, relation (2.7) holds true only if $\theta = (\tau, \theta_2, ..., \theta_m)$ where τ is the log-location parameter. Referring to Example 1, τ is usually taken as the scale parameter, but, from our point of view, it is the image of the location of the prototype distribution expresses the central tendency of F. Since $T(\tau;\tau) = S_G(0) = 0$, we realized that the important quantity in (2.7) is not the value of a concrete parameter, but the zero of the t-score. This is the reason for introducing new statistical concepts:

Definition 1. *Let t-score of distribution F_θ with support $\mathcal{X} \subseteq \mathbb{R}$ be given by (2.10) with η given by (2.9). The solution $x^* = x^*(\theta)$ of equation*

$$T(x;\theta) = 0 \tag{2.11}$$

is called the transformation-based mean or shortly the t-mean.

Actually, the t-mean is the transformed mode of the 'prototype' distribution. Relation (2.7) was consequently generalized in the following way.

Definition 2. *Let T be the t-score and x^* the t-mean of distribution F_θ with support $\mathcal{X} \subseteq \mathbb{R}$. Function*

$$S(x;\theta) = \eta'(x^*)T(x;\theta) \tag{2.12}$$

is called the scalar score of distribution F_θ.

Scalar score is 'the score function for t-mean', describing the relative influence of x for a construction of the t-mean.

Example 2.2. The gamma distribution with density $f(x;\alpha,\gamma) = \frac{\gamma^\alpha}{\Gamma(\alpha)}x^{\alpha-1}e^{-\gamma x}$ has t-score $T(x;\alpha,\gamma) = \gamma x - \alpha$ so that the t-mean is $x^* = \alpha/\gamma$. Its scalar score $S(x;\alpha,\gamma) = \gamma(x/x^* - 1)$ is the score function for the ratio α/γ.

Other examples of scalar scores of various continuous distributions are given in Fabián (2008).

3 BASIC PROPERTIES OF SCALAR SCORES

Function $S^2(x;\theta)$ attains its minimum at x^*. By analogy to log-location distributions, $ES^2(\tau)$ of which is the Fisher information for τ, value ES^2 can be interpreted as the scalar Fisher information (or as the information of the distribution). Since x^* is the least informative point of the distribution (Fabián, 2010), $S^2(x)$ can be thought of as the information function, expressing relative information contained in observation x.

The score moments

$$M_k(\theta) = E_\theta S^k(X) = \int_{\mathcal{X}} S^k(x;\theta)f(x;\theta)\,dx, \qquad k = 1,2,... \tag{3.1}$$

can be used as numeric characteristics of distributions. It follows from (2.2) that if $g(x) = O(e^{-x})$ for $x \to \pm\infty$, then $S_G(x) \sim O(1)$. Since mapping (2.9) retains the properties of S_G on boundaries of the support, the scalar scores of heavy-tailed distributions are bounded and the score moments exist. Furthermore, $ES = 0$ (the scalar score is centered around the t-mean). The reciprocal value of the scalar Fisher information,

$$\omega^2 = \frac{1}{E_\theta S^2} = \frac{1}{[\eta'(x^*)]^2 E_\theta T^2}, \tag{3.2}$$

appeared to be a good measure of the variability of distributions, particularly in cases in which the usual variance does not exist (Fabián, 2009). We call it now the *score variance*. For distributions with support $(0,\infty)$, (3.2) sounds

$$\omega^2 = \frac{(x^*)^2}{E_\theta T^2}. \tag{3.3}$$

Let the observed data $\mathbf{x}$, realizations of random variables $X_1,...,X_n$, be iid according to some F from parametric family $\{F_\theta, \theta \in \Theta\}$ and let $S(x;\theta)$ be the corresponding scalar scores. The sample characteristics, the sample t-mean $\hat{x}^* = x^*(\hat{\theta})$ and the sample score variance $\hat{\omega}^2 = \omega^2(\hat{\theta})$ can be obtained as functions of the estimated parameters. By using the new data characteristics it is easy to compare results of the estimation in different models (Fabián, 2008).

The scalar score equations for estimation of θ, derived from (3.1) using the substitution principle, are the generalized moment equations

$$\hat{\theta}_M: \qquad \frac{1}{n}\sum_{i=1}^{n} T^k(x_i;\theta) = \mathrm{E}_\theta T^k, \qquad k = 1,...,m, \tag{3.4}$$

where scalar scores are replaced by t-scores due to (2.12)). Since the score (t-score) moments are often expressed by elementary functions of parameters, (Fabián, 2010b), and scalar scores (t-scores) of heavy-tailed distributions are bounded and estimators (3.4) are in these cases robust (Fabián and Stehlík, 2009, Fabián, 2010b).

In some cases, the first equation of system (3.4) can be written in the form

$$\sum_{i=1}^{n} T(x_i; x^*) = 0. \tag{3.5}$$

Then, by (Fabián, 2009, the estimate of the t-mean is asymptotically normal, $\hat{x}^* \sim AN(x^*, \sigma_*^2)$, with

$$\sigma_*^2 = \frac{E_\theta T^2}{(E_\theta T'_*)^2},$$

and where $T'_* = \frac{d}{dx^*} T(x; x^*)$. According to (3.3), the square root of the measure of variability ω^2 is $\omega = x^*/\sqrt{(E_\theta T^2)}$ so that the asymptotic variance of the 'sample score deviance' $\hat{\omega}$ is

$$\sigma^2(\hat{\omega}) = \frac{1}{(E_\theta T'_*)^2}.$$

For testing $H_0 : x^* = x_0^*$ versus $H_1 : x^* \neq x_0^*$, it is natural to use the score test with scalar score instead of the vector score function. A simpler alternative test and the corresponding confidence intervals are described in Fabián (2009).

According to (3.2), the sample score variance is in these cases given by

$$\hat{\omega}^2 = \frac{(\hat{x}^*)^2}{\frac{1}{n}\sum_{i=1}^{n} T^2(x_i; \hat{x}^*)}. \tag{3.6}$$

In a general case, however, the sample characteristics are to be determined as $\hat{x}^* = x^*(\hat{\theta}_{SM})$ and $\hat{\omega}^2 = \omega^2(\hat{\theta}_{SM})$.

4 EXAMPLE: THE BETA-PRIME DISTRIBUTION

The beta-prime distribution (beta distribution of the second kind) with support $(0, \infty)$ and density

$$f(x; p, q) = \frac{1}{B(p, q)} \frac{x^{p-1}}{(x+1)^{p+q}} \qquad p, q > 0$$

is an example of a heavy-tailed distribution. Neither of parameters is the log-location, the mean and variance exist only if $q > 1$ and $q > 2$, respectively. By (2.8), the t-score is

$$T(x; p, q) = \frac{qx - p}{x + 1},$$

a simple bounded function different from both partial scores for p and q. The t-mean $x^* = p/q$, so that the scalar score function is given by

$$S(x; x^*, q) = \frac{q^2}{p} \frac{x - x^*}{x + 1}.$$

Since $ET^2 = pq/(p+q+1)$, the variability of the distribution is described by the score variance (3.3),

$$\omega^2 = \frac{p(p+q+1)}{q^3}. \tag{4.1}$$

The scalar score estimation equations (3.4) are

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

from which one obtains the sample t-mean $\hat{x}^*$ and $\hat{q}$ in closed formulas. Further, $\hat{p} = \hat{x}^* \hat{q}$ and $\hat{\omega}^2$ is obtained from (4.1). Since $ET'_* = q^2/(p+q)$, the standard deviations of the estimates are

$$\sigma(\hat{x}^*) = \frac{p^{1/2}(p+q)}{q^{3/2}(p+q+1)}, \qquad \sigma(\hat{\omega}) = \frac{p+q}{q^2}.$$

Fig. 1a shows the average estimates of the t-mean $\hat{x}^* = x^*(\hat{p}, \hat{q})$ and $\hat{\omega} = \omega(\hat{p}, \hat{q})$, where $\hat{p}$ and $\hat{q}$ are either the maximum likelihood or the score moment estimates of parameters of samples of length $n = 100$, randomly generated from the beta-prime distribution with increasing variability ω. Average values are computed after 2000 replications. The estimates of ω constructed from the maximum likelihood estimates are increasingly biased with increasing ω, since they are influenced by 'outliers', the values far from the bulk of the data (and generated in accordance with the distribution). The estimates based on the score moment estimates are robust. The average standard deviations of the estimates are shown in Fig. 1b and 1c. Standard deviations of robust score moment estimates of ω roughly follow the theoretical values whereas standard deviations of the maximum likelihood estimates of ω (Fig. 1c) are biased to higher values due to 'outliers'.

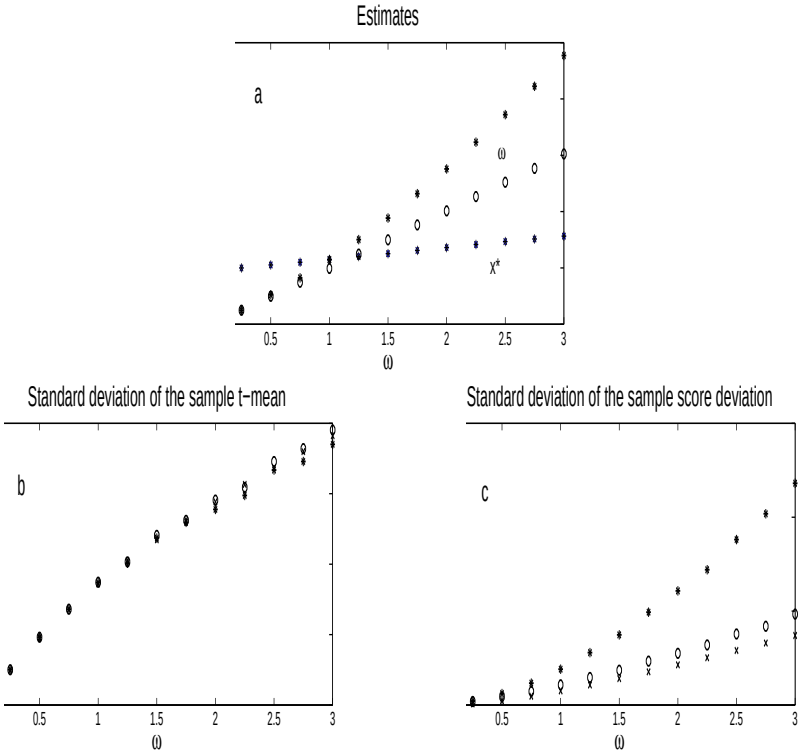

Figure 1. (a) Average maximum likelihood ($*$) and score moment (o) estimates of x^* and ω of the beta-prime distribution. (b) Average standard deviations of $\hat{x}^*$, (c) average standard deviations of $\hat{\omega}$. Theoretical values are marked by x.

5 SCORE CORRELATION COEFFICIENT

Let S_X, S_Y, respectively, be the scalar scores of random variables X and Y with supports $\mathcal{X}_X, \mathcal{X}_Y$ and joint distribution $f(x, y)$. The joint score moment of X and Y is

$$ES_X S_Y = \int_{\mathcal{X}_X} \int_{\mathcal{X}_Y} S_X(x) S_Y(y) f(x, y)\, dxdy. \tag{5.1}$$

For normally distributed X and Y, (5.1) is the ordinary covariance coefficient.

Definition 1. Define the score covariance coefficient of random variables X, Y with distributions F_X, F_Y and scalar scores S_X, S_Y, respectively, by

$$Cov_{score}(X, Y) = ES_X S_Y. \tag{5.2}$$

Definition 2. Define the score correlation coefficient of random variables X, Y from Definition 1 by

$$\rho_{score}(X, Y) = \rho(S_X(X), S_Y(Y)) = \frac{ES_X S_Y}{\sqrt{ES_X^2 ES_Y^2}}, \tag{5.3}$$

where ρ is the Pearson correlation coefficient.

It is apparent that $-1 \leq \rho_{score} \leq 1$ and that if X and Y are independent, $\rho_{score}(X, Y) = 0$. Moreover, by (2.12), the score correlation coefficient is expressed by means of t-scores, $\rho_{score}(X, Y) = \frac{ET_X T_Y}{\sqrt{ET_X^2 ET_Y^2}}$.

The formula for the sample score correlation coefficient is straightforward. In simulation experiments we generated couples (X, Z) using independently generated random samples of X and Z from distributions from Table 1, and set

$$Y = \alpha X + (1 - \alpha)Z. \tag{5.4}$$

The theoretical value of $r \equiv \rho(X, Y; \alpha)$ is $r = \alpha/\sqrt{2\alpha^2 - 2\alpha + 1}$. The correlation coefficients were estimated from samples of length 75 with 2000 replications.

Figure 2 shows average values of estimates of the Pearson ρ, Kendall's $\tau(x, y)$, Spearman $\rho_S(x, y)$ and the score $\rho_{score}(x, y)$ correlation coefficients as functions of increasing variability of distributions, described by the square root ω of the score variance (3.2). Relations between ω and parameters of distributions used for simulation experiments are given in Table I.

Table I. Scalar score and score variances of some distributions.

Distribution	$F(x)$	$f(x)$	$S(x)$	ω^2
exponential	$1 - e^{-x/\tau}$	$\frac{1}{\tau} e^{-x/\tau}$	$\frac{1}{\tau}(\frac{x}{\tau} - 1)$	τ^2
Weibull	$1 - e^{-x^c}$	$cx^{c-1}e^{-x^c}$	$c(x^c - 1)$	$1/c^2$
Pareto $(1, \infty)$	$1 - x^{-c}$	$\frac{c}{x^{c+1}}$	$c - \frac{c+1}{x}$	$\frac{c+2}{c^3}$
Fréchet	$1 - e^{-x^{-c}}$	$cx^{-(c+1)}e^{-x^{-c}}$	$c(1 - x^{-c})$	$1/c^2$
log-logistic	$1 - \frac{1}{(1+x)^q}$	$\frac{1}{(1+x)^{1+q}}$	$\frac{qx-1}{x+1}$	$(q + 2)/q^3$

For each sample, the parameters of marginal densities were estimated by procedure described in the Section 4 before estimating ρ_{score}. The Kendall and Spearman correlation coefficients were computed by means of code *corr* from the MATLAB library. We expected an increase of correlation coefficients with increasing ω due to increasing number of values far from the bulk of the data in the generated samples.

In case of the exponential distribution with linear scalar score, $\rho_{score} = \rho$ and all average values of the sample correlation coefficients are roughly constant with increasing ω. In case of the Weibull

distribution (a light-tailed distribution with a non-linear scalar score), ρ seems to be the best choice for small $\omega < 1.5$. However, for $\omega \geq 1.5$ are standard deviations of ρ too high. For $r = 0.2$, ρ_{score} behaves similarly as Kendall's τ, for $r = 0.4$ has ρ high standard deviations and ρ_{score} seems to be the best choice.

In cases of heavy-tailed distributions (Pareto, Fréchet, log-logistic), on the other hand, the Pearson correlation coefficient loses any meaning, as documented by the plot of standard deviations of the formal estimates of ρ for Pareto distribution with $r = 0.2$. ρ_{score} of Pareto and log-logistic distributions are closed to the Spearman estimate, in case of the Fréchet distribution, ρ_{score} is obviously the best one.

6. SUMMARY

Scalar score is a new inference function, constructed in accordance with the well-known statistical concepts and reflecting the properties of the assumed parametric model. The function made it to introduce new measures of central tendency and variability of probability distributions, which exist in cases of heavy-tailed distributions. Their sample counterparts, the sample t-mean and sample score variance, can be constructed from the estimates of parameters and enables comparison of results of estimation in differently parametrized models.

We used the scalar score for definition of a distribution-dependent score covariance and score correlation coefficient. Our conclusion based on simulation experiments is that although increasingly biased with increasing variability (score variance) of the distribution, the score correlation coefficient can detect an association of random variables having heavy-tailed distributions, taking into account the properties of marginal distributions.

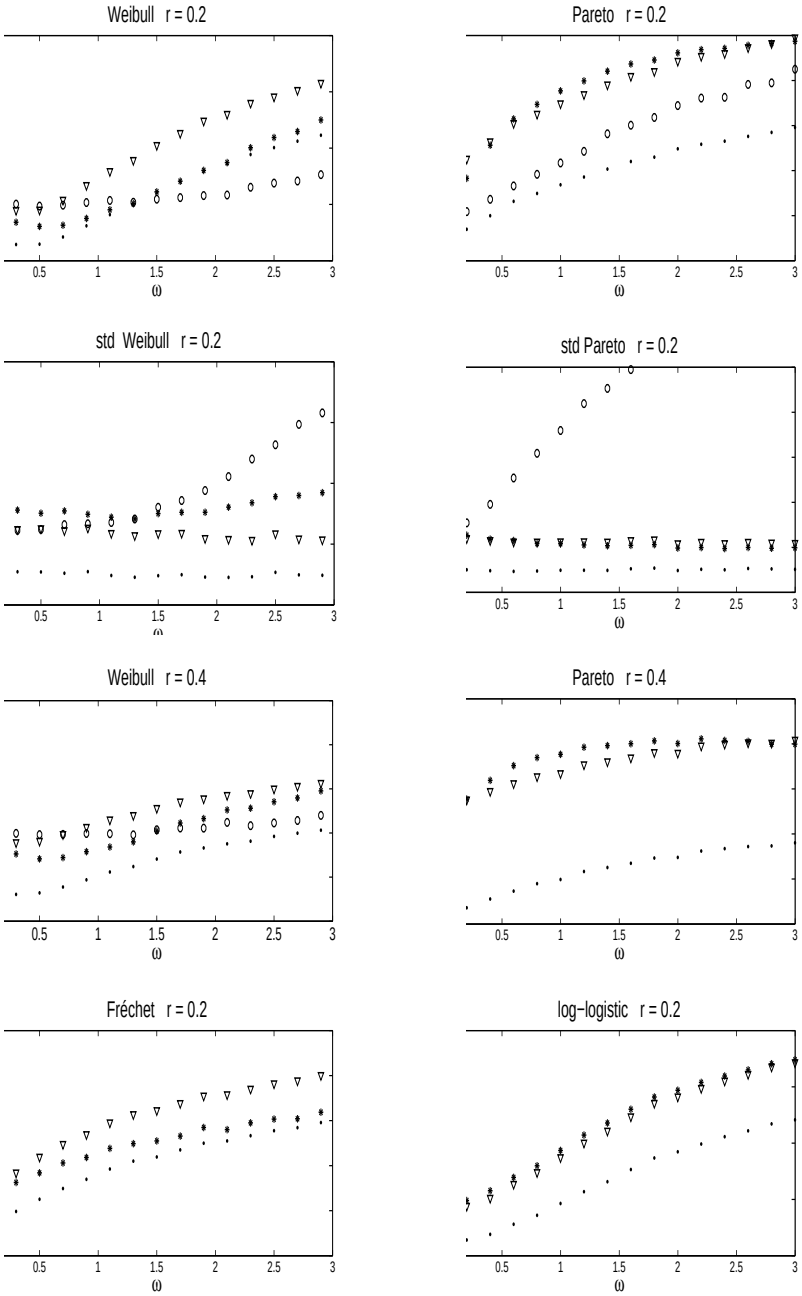

Figure 2. Average values of the sample correlation coefficients as functions of increasing variability ω of some distributions.
o Pearson . Kendall ∇ Spearman $*$ scalar score

Acknowledgements. The research presented in the paper was supported by projects AV0Z10300504 and GACR 205/09/1079.

BIBLIOGRAPHY

Balakrishnan, N., Nevzorov, V. B. (2003). *A Primer of Statistical Distributions.* Hoboken: Wiley.

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

Fabián, Z. (2007). Estimation of simple characteristics of samples from skewed and heavy-tailed distribution. In C. Skiadas, ed, *Recent Advances in Stochastic Modeling and Data Analysis.* Singapore: World Scientific.

Fabián, Z. (2008). New measures of central tendency and variability of continuous distributions, *Comm. Statist. Theory Methods* 37: 159–174.

Fabián, Z., Stehlík, M. (2008). A note on favorable estimation when data is contaminated. *Comm. Dependability and Quality Management* 11: 36–43.

Fabián, Z. (2009). Confidence intervals for a new characteristic of central tendency of distributions. *Comm. Statist. Theory Methods* 38: 1804–1814.

Fabián, Z. (2009b). The t-information and its use in multivariate problems and time series analysis. *J. Statist. Planning and Inference* 139:3773–3778.

Fabián, Z. (2010). Uncertainty of random variables. Proc. of conf. ASMDA 2010.

Fabián, Z. (2010b). Score moment estimators. Proc. of conf. COMPSTAT 2010.

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

Johnson, N. L., Kotz, S., Balakrishnan, N. (1995). *Continuous univariate distributions 2.* Hoboken: Wiley.

Marshall A. W., and Olkin I. (2007). *Life distributions. Structure of nonparametric, semiparametric and parametric families.* Springer.

Mori, D. D., Kotz, S. (2001). *Correlation and dependence.* London: Imperial College Press.