Description of Continuous Distributions and Data Samples by Means of Score Functions of Distribution

Fabián, Zdeněk
2014

Institute of Computer Science
Academy of Sciences of the Czech Republic

Description of Continuous Distributions and Data Samples by Means of Score Functions of Distribution

Z. Fabián

Technical report No. 1216

December 2014

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

Description of Continuous Distributions and Data Samples by Means of Score Functions of Distribution[1]

Z. Fabián

Technical report No. 1216

December 2014

Abstract:

In this paper we elucidate and further develop the concept of the score function of distribution, which could be a new mathematical tool for description of univariate continuous probability distributions and for estimation of characteristics of data samples generated from them. Moreover, the function appeared to be the basis for generation of other functions characterizing distributions and of new numerical characteristics, finite even in cases of heavy-tailed distributions. Given a model, the sample counterparts of these numerical characteristics suitably describe the data samples.

Keywords:
score function; score mean; score variance; generalized Fisher information; data characteristics

[1]The work was done by institutional support RVO:67985807. AMS Subject Classification: 62E10

1 Introduction

Having observed data $\mathbf{X}_n = (x_1, ..., x_n)$, realizations of random variables $X_1, ..., X_n$ iid according to some F, it is supposed in parametric approach that F is a member of a parametric family $\mathcal{F}_{\mathcal{X},\theta} = \{F_\theta : \theta \in \Theta \subseteq \mathcal{R}^m\}$. The unknown θ is estimated by 'treating' observed values in accordance with the model: the data are inserted into a score function describing relative influence of observations with respect to θ. Thus, there are not the data themselves entering into inference procedures, but 'latent' values of parametric score functions. However, if $m > 1$ the vector-valued Fisher (maximum likelihood) score functions of classical statistics do not enable consistent use of this point of view at more complicated problems, and scalar-valued score functions of robust statistics are often in a loose relation to the assumed model.

The simplest parametric model is the location model. Let $\theta = \mu \in \mathcal{R}$ be a location parameter and F_μ a distribution supported by $\mathcal{X} = \mathcal{R}$ with density $f(x-\mu)$. The identity

$$\frac{\partial}{\partial\mu}\log f(x-\mu) = -\frac{1}{f(x-\mu)}\frac{d}{dx}f(x-\mu) \equiv S_F(x-\mu) \tag{1.1}$$

shows that the classical Fisher (likelihood) score for location can be obtained by differentiating $-\log f(x-\mu)$ with respect to the variable. By setting $\mu = 0$, Hampel et al. (1986) have pointed out that, loosely speaking, the relative rate of the change of f, function

$$S_F(x) = -\frac{f'(x)}{f(x)}, \tag{1.2}$$

can be interpreted as describing relative influence of $x \in \mathcal{R}$ with respect to the 'most probable' value, the center point (= mode) of the distribution. Consistently with this observation, Cover and Thomas (1991, pp.494) treated ES_F^2 as the Fisher information of distribution F. Function (1.2), mentioned sometimes (Sen et al. 2009, pp.59) as a generalized score function, is called by Jurečková (2012) the *score function of distribution*. Such concept encompasses even parametric distributions with support $\mathcal{R}$ without location parameter.

Taking $S_F(X)$ as 'treated X', or, in other words, as the random variable associated to X, it holds under mild regularity conditions that:

i) Moments $ES_F^k(X)$ relevant to X exist, $ES_F(X) = 0$.

ii) As a central value x^* of F can be taken the solution of $S_F(x) = 0$.

Having a model family $F_{\mathcal{X},\theta}$, it may seem rational to study the data through the associated random variable $S_F(X;\theta)$ where

$$S_F(x;\theta) = -\frac{1}{f(x;\theta)}\frac{d}{dx}f(x;\theta). \tag{1.3}$$

However, this approach is used neither in the probability theory nor in statistics due to the fact that the behavior of function (1.2) is often peculiar. For instance, $S_F(x)$ of the uniform distribution is identically equal to zero.

Fabián (2001) proposed an idea, which can be outlined as follows. There are two different types of continuous probability distributions. Under the 'prototypes' he means a basic set of distributions with support $\mathcal{R}$ and with a 'simple' form of density (the term will be clarified later). These are the distributions which are sufficiently characterized by (1.2) or (1.3). Other distributions can be viewed as 'transformed distributions': the prototypes transformed by some $\eta^{-1} : \mathcal{R} \to \mathcal{X} \subseteq \mathcal{R}$. Particularly, any distribution with support $\mathcal{X} \neq \mathcal{R}$ can be viewed as a transformed distribution. During the time it appeared that the mapping η is not an arbitrary one: it is the function, the derivative of which is the inner part of the density formula (Fabián 2013). As a result, a uniquely defined scalar score function of distribution with arbitrary support has been obtained as a counterpart of the density. In parametric cases it is a parametric function, derived, however, by means of differentiating according the variable.

The plan of the present paper is the following. The concept of the score function of distribution (1.2) is extended to 'transformed distributions' in the next section. Section 3 introduces new numerical characteristics of distributions and new functions describing distributions based on it. In the last section we try to demonstrate that the new concepts are not purposeless and discuss their possible use in statistical tasks.

2 Sfd of transformed distributions

Let $\mathcal{X} = (a, b)$ be a finite or infinite interval of the real line $\mathcal{R}$. Random variable X with support $\mathcal{X}$ and distribution F is described by distribution function $F(x) = P(X \leq x)$ and density $f(x) = dF(x)/dx$, $f(x) > 0$ for $x \in \mathcal{X}$, $f(x) = 0$ if $x \in \mathcal{R}\backslash\mathcal{X}$. By $\Pi_{\mathcal{X}}$ we denote the set of distributions with support $\mathcal{X}$, regular in the usual sense with one addition specified later.

2.1 Definition

Definition 1. *Let $\mathcal{X} = (a, b) \subset \mathcal{R}$. We say that $\eta : \mathcal{X} \to \mathcal{R}$ is a Johnson mapping if*

$$\eta(x) = \begin{cases} \log(x-a) & \text{when} \quad \mathcal{X} = (a, \infty) \\ \log\dfrac{(x-a)}{(b-x)} & \text{when} \quad \mathcal{X} = (a, b). \end{cases} \tag{2.1}$$

Johnson mappings (2.1) are Johnson's (1949) transformations to change the support, reduced to be parameter-free.

Definition 2. *Let $\mathcal{X} \subseteq \mathcal{R}$ and $f(x)$ be the density of distribution $F \in \Pi_{\mathcal{X}}$. Let $\eta : \mathcal{X} \to \mathcal{R}$ be a smooth strictly increasing mapping. Let g be the density of $G \in \Pi_{\mathcal{R}}$ and let $F = G \circ \eta$ so that the density of F is*

$$f(x) = g(\eta(x))\eta'(x), \tag{2.2}$$

where $\eta'(x) = d\eta(x)/dx$. G is called the prototype of F.

In the following text we consistently denote prototypes by G, g and transformed distributions by F, f. A transformed distribution can be a distribution from $\Pi_{\mathcal{R}}$ as well. For instance, a distribution with the density

$$f(x) = \frac{1}{1+x^2}\frac{e^{\tanh^{-1} x}}{(1+e^{\tanh^{-1} x})^2} \tag{2.3}$$

is the transformed standard logistic prototype $g(y) = e^y/(1+e^y)^2$ with (easily identifiable) mapping $\eta : \mathcal{R} \to \mathcal{R}$ in the form $\eta(x) = \tanh^{-1} x$.

For distributions with half-line or finite interval support the retrieval of the 'true' $\eta(x)$, which we call the *innate mapping*, need not be apparent at first sight. The density $f(x) = e^{-x}$ of the exponential distribution has certainly a 'hidden' innate mapping $\eta(x) = \log x$, by means of which it can be written as $f(x) = xe^{-x}\frac{1}{x}$. The reason why to use Johnson mapping in cases in which the innate $\eta(x)$ or $\eta'(x)$ is not evident from the density formula is mainly the principle of parsimony (mappings (2.1) are the often considered simple mappings), but in concrete situations there can be other supporting reasons, discussed at 2.3.

Definition 3. *Let $F \in \Pi_{\mathcal{X}}$ has differentiable density $f(x)$ and $\eta : \mathcal{X} \to \mathcal{R}$ be its innate mapping. Function*

$$T_F(x) = -\frac{1}{f(x)}\frac{d}{dx}\left[\frac{1}{\eta'(x)}f(x)\right] \tag{2.4}$$

is the t-score of distribution F. Let the solution x^* to the equation

$$T_F(x) = 0 \tag{2.5}$$

be unique. x^* is called the score mean and function

$$S_F(x) = \eta'(x^*)T_F(x) \tag{2.6}$$

is the score function of distribution (abbreviated by sfd of F).

The concept of the t-score is explained as follows. If random variable Y has distribution $G \in \Pi_{\mathcal{R}}$, random variable $X = \eta^{-1}(Y)$ has distribution $F = G \circ \eta$ with density (2.2). The first term on the right hand side of (2.2) contains probabilistic information about X, whereas the Jacobian term $\eta'(x)$ is common for many members of $\Pi_{\mathcal{X}}$, masking the statistical content of $f(x)$. The trick is to remove the Jacobian from the density formula before differentiation with respect to the variable.

By Definition 3, for computation of the sfd of any $F \in \Pi_{\mathcal{X}}$ one does not need to know its prototype G. However, knowing the prototype, one can immediately write the formula for the sfd of the transformed distribution.

Proposition 1. *Let $G \in \Pi_{\mathcal{R}}$ and $F = G \circ \eta \in \Pi_{\mathcal{X}}$. Then,*

$$T_F(x) = S_G(\eta(x)). \tag{2.7}$$

Proof. By (2.2), $\frac{1}{\eta'(x)}f(x) = g(\eta(x))$. By (2.4)

$$T_F(x) = -\frac{1}{g(\eta(x))\eta'(x)}\frac{d}{dx}[g(\eta(x)] = S_G(\eta(x)).$$

□

Actually, the behavior of sfds of prototypes has been known since Hampel et al. (1986). In particular, $S_G(y)$ of light-tailed distributions are unbounded, whereas sfds of heavy-tailed distributions are bounded. Sfds of some distributions (as the extreme value one) are semi-bounded (in this case bounded on the right, unbounded on the left). By Proposition 1, the t-scores of transformed distributions preserve at the boundaries of the support the behavior of sfds of their prototypes (in contrast to the tail properties).

2.2 Sfd of transformed parametric distributions

We distinguished two kinds of transformed distributions: with and without a central parameter.

Definition 4. *Let $G_\mu \in \Pi_\mathcal{R}$ be a location distribution. Let $F_\tau = G_\mu \circ \eta \in \Pi_\mathcal{X}$ be the transformed distribution with parameter*

$$\tau = \eta^{-1}(\mu). \tag{2.8}$$

τ is called a transformed location parameter or t-location.

Under transformed distributions of the first kind we understand the distributions with t-location parameter. A subclass of distributions with support $\mathcal{R}^+ = (0, \infty)$ and t-location parameter are the log-location distributions studied by Lawless (2003).

The next theorem, proven by Fabián (2001), shows that the sfd of distributions with t-location parameter is identical with the Fisher (maximum likelihood) score for the t-location. The proof is reproduced in the Appendix.

Theorem 1. Let F_τ be a transformed distribution with t-location parameter. Then

$$S_F(x;\tau) = \frac{\partial}{\partial\tau}\log f(x;\tau). \tag{2.9}$$

A general form of the density of location prototypes is $g(y;\mu,\theta_2,...,\theta_m)$. Consider a particular case of the location and scale prototype $G_{\mu,\sigma}$ with density $\sigma^{-1}g((y-\mu)/\sigma)$. By (2.2), the density of the transformed distribution $F = G_{\mu,\sigma} \circ \log \in \Pi_{\mathcal{R}^+}$ is

$$f(x) = \frac{1}{\sigma}g\left(\frac{\log x - \mu}{\sigma}\right)\frac{1}{x} = cg\left(\log\left(\frac{x}{\tau}\right)^c\right)\frac{1}{x}$$

where $\tau = e^\mu$ and $c = 1/\sigma$. Examples are given in the upper part of Table 1. The t-location parameter τ, usually taken as the scale, for distributions from $\Pi_{\mathcal{R}^+}$ appears to be the central parameter. Furthermore, the parameter c, generally taken as shape parameter, can be better explained in cases of distributions from $\Pi_{\mathcal{R}^+}$ as a reciprocal scale.

Transformed distributions of the second kind are those without a central (t-location) parameter. Their sfds, created in accordance with sfds of distributions of the first kind by means of Definition 3, were yet unknown functions. They can be understood as Fisher scores for the score mean, which is not a value of any parameter of the distribution. Examples are given in the lower part of Table 1.

Table 1. Characteristics of some distributions.

F	$f(x)$	$S_F(x)$	x^*	ω^2	$\bar{x}_S$
lognormal	$\frac{c}{\sqrt{2\pi}x}e^{-\frac{1}{2}\log^2(\frac{x}{\tau})^c}$	$\frac{c}{\tau}\log(\frac{x}{\tau})^c$	τ	$\frac{\tau^2}{c^2}$	$\bar{x}_G$
Weibull	$\frac{c}{x}(\frac{x}{\tau})^c e^{-(\frac{x}{\tau})^c}$	$\frac{c}{\tau}[(\frac{x}{\tau})^c - 1]$	τ	$\frac{\tau^2}{c^2}$	$(\frac{1}{n}\sum x_i^c)^{1/c}$
Fréchet	$\frac{c}{x}(\frac{x}{\tau})^{-c} e^{-(\frac{x}{\tau})^{-c}}$	$\frac{c}{\tau}[1 - (\frac{x}{\tau})^{-c}]$	τ	$\frac{\tau^2}{c^2}$	$(\frac{1}{n}\sum \frac{1}{x_i^c})^{-1/c}$
GIG	$\frac{1}{Kx}e^{-\frac{1}{2\tau}[(\frac{x}{\tau})+(\frac{\tau}{x})]}$	$\frac{1}{2\tau}[\frac{x}{\tau} - \frac{\tau}{x}]$	τ	$\frac{\tau^2}{1.43}$	$(\bar{x}\bar{x}_H)^{1/2}$
loglogistic	$\frac{c}{x}\frac{(x/\tau)^c}{(1+(x/\tau)^c)^2}$	$\frac{c}{\tau}\frac{(x/\tau)^c-1}{(x/\tau)^c+1}$	τ	$\frac{\tau^2}{c^2}$	–
gamma	$\frac{\gamma^\alpha}{x\Gamma(\alpha)}x^\alpha e^{-\gamma x}$	$\frac{\gamma^2}{\alpha}(x - x^*)$	$\frac{\alpha}{\gamma}$	$\frac{\alpha}{\gamma^2}$	$\bar{x}$
inv. gamma	$\frac{\gamma^\alpha}{x\Gamma(\alpha)}x^{-\alpha} e^{-\gamma/x}$	$\frac{\alpha^2}{\gamma}(1 - \frac{x^*}{x})$	$\frac{\gamma}{\alpha}$	$\frac{\gamma}{\alpha^2}$	$\bar{x}_H$
beta-prime	$\frac{1}{B(p,q)}\frac{x^{p-1}}{(x+1)^{p+q}}$	$\frac{q^2}{p}\frac{x-x^*}{x+1}$	$\frac{p}{q}$	$\frac{p(p+q+1)}{q^3}$	$\frac{\sum_{i=1}^n \frac{x_i}{x_i+1}}{\sum_{i=1}^n \frac{1}{x_i+1}}$
log-gamma	$\frac{c^\alpha}{\Gamma(\alpha)}\frac{(\log x)^{\alpha-1}}{x^{c+1}}$	$\rho\log\frac{x}{x^*}$	$e^{\alpha/c}$	$\frac{1}{\alpha}e^{2\alpha/c}$	$\bar{x}_G$
Pareto	c/x^{c+1}	$c^2(1 - \frac{x^*}{x})$	$\frac{c+1}{c}$	$\frac{c+2}{c^3}$	$\bar{x}_H$
beta	$\frac{x^{p-1}(1-x)^{q-1}}{B(p,q)}$	$\frac{(p+q)^2}{pq}(x - x^*)$	$\frac{p}{p+q}$	$\frac{pq(p+q+1)}{(p+q)^3}$	$\bar{x}$

Except the Pareto and log-gama distributions from $\Pi_{(1,\infty)}$ and beta from $\Pi_{(0,1)}$, distributions in the table are members of $\Pi_{\mathcal{R}^+}$ with $T_F(x) = x^* S_F(x)$. $\Gamma(\cdot)$ is the gamma function, $B(\cdot,\cdot)$ the beta function, $\rho = \frac{c^2}{\alpha}e^{-\alpha/c}$, $K = K_0(1)$ is the Bessel function of the third kind. $\bar{x}$ is the arithmetic, $\bar{x}_G$ geometric and $\bar{x}_H$ harmonic mean.

2.3 Uniqueness of the t-score

In this paragraph we present some examples of 'detection' of the innate mapping of transformed distributions.

Distributions from $\Pi_{\mathcal{R}}$ are often prototypes, as *normal* $\mathcal{N}(\mu,\sigma)$ with sfd $S_F(x) = \frac{x-\mu}{\sigma^2}$, *logistic* with sfd $S_G(y) = (e^y - 1)/(e^y + 1)$ or the *skew normal* (Azzalini 1985) with density $g(y;\lambda) = 2\phi(y)\Phi(\lambda y)$, where ϕ and Φ are the density and distribution function of the normal distribution. Its sfd (1.2) is clearly

$$S_G(y) = y - \lambda\frac{\phi(\lambda y)}{\Phi(\lambda y)}. \tag{2.10}$$

Transformed distributions from $\Pi_{\mathcal{R}}$ have easily identifiable innate mapping $\eta : \mathcal{R} \to \mathcal{R}$, see e.g. (2.3) or, for instance, the *sinh-arcsinh* distribution (Jones and Pewsey, 2009) with density

$$f(x) = \frac{1}{2\pi}\frac{\delta\cosh(\varepsilon + \delta\sinh^{-1}x)}{\sqrt{1+x^2}}e^{-\frac{1}{2}\sinh^2(\varepsilon+\delta\sinh^{-1}x)}, \tag{2.11}$$

which is clearly the standard normal distribution transformed by $\eta(x) = \sinh(\varepsilon + \delta\sinh^{-1}x)$. Densities (2.11) and sfds $S_F(x) = \eta(x)$ (much simpler than would be functions (1.2)) are for $\delta = 1$ and some values of ε given in Fig. 1.

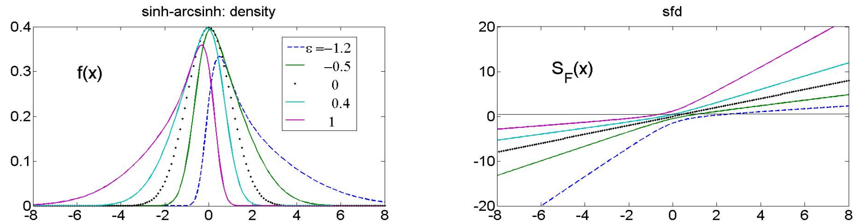

Fig. 1. Densities and sfds of the sinh-arcsinh distribution (2.11).

The innate mapping of distributions from $\Pi_{\mathcal{R}^+}$ is mostly the logarithmic mapping, and t-scores have a form

$$T_F(x) = -\frac{1}{f(x;\theta)}\frac{d}{dx}[xf(x;\theta)] = -1 - x\frac{f'(x;\theta)}{f(x;\theta)}.$$

Hence, sfd of the exponential distribution is $S_F(x) = x - 1$, a reasonable result since a function describing the 'distribution without memory' should be linear. By using Proposition 1, the sfd of the *skew-lognormal* distribution is, using (2.10), $S_F(x) = \log x - \lambda\phi(\lambda \log x)/\Phi(\lambda \log x)$.

Other possible more complex transformations, as for instance $\eta(x) = \sinh(\log x)$, recommended by Jones (2014), are from the forms of the density easily detectable.

Let us consider distributions with semi-infinite support $\mathcal{X} = (1, \infty)$. The density of the *log-gamma* distribution

$$f(x) = \frac{c^\alpha}{\Gamma(\alpha)}(\log x)^{\alpha-1}\frac{1}{x^{c+1}} = \frac{c^\alpha}{\Gamma(\alpha)}(\log x)^{\alpha}\frac{1}{x^c}\cdot\frac{1}{x\log x}$$

has an obvious innate mapping $\eta(x) = \log\log x$. By Definition 3

$$T_F(x) = -\frac{x^{c+1}}{\log^{\alpha-1} x}\frac{d}{dx}[(\log x)^\alpha \frac{1}{x^c}] = c\log x - \alpha$$

with $x^* = e^{\alpha/c}$ and $S_F(x) = \frac{1}{x^*\log x^*}T_F(x)$. However, the *Pareto* distribution with density

$$f(x) = \frac{c}{x^{c+1}} \tag{2.12}$$

does not contain any 'visible' Jacobian of any transformation. Trying to use the transform $\eta(x) = \log\log x$ one obtains

$$T_F(x) = -\frac{1}{f(x)}\frac{d}{dx}[x\log x f(x)] = c\log x - 1,$$

which is the Fisher score for c. Using the Johnson mapping one obtains

$$T_F(x) = -\frac{1}{f(x)}\frac{d}{dx}[(x-1)f(x)] = c - (c+1)/x = c(1 - x^*/x) \tag{2.13}$$

with score mean

$$x^* = (c+1)/c. \tag{2.14}$$

Since the 'shifted Pareto' from $\Pi_{\mathcal{R}^+}$ with density $f(x) = \frac{c}{(x+1)^{c+1}}$ is a particular case of the *beta-prime* distribution (Table 1) with bounded sfd, the t-score of the Pareto distribution should be the latter one. Then, $S_F(x) = \frac{1}{x^*-1} T_F(x) = c^2(1 - x^*/x)$.

The Johnson mapping for distributions from $\Pi_{(0,1)}$ is $\eta(x) = \log \frac{x}{1-x}$ with $\eta'(x) = \frac{1}{x(1-x)}$, so that the corresponding t-score is

$$T_F(x) = -\frac{1}{f(x)} \frac{d}{dx}[x(1-x)f(x)] = -1 + 2x - x(1-x)\frac{f'(x)}{f(x)}.$$

In current use are Johnson's U_B distribution with normal prototype and the *beta* distribution with density

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

and linear t-score $T_F(x) = (p+q)x - p$. The *uniform* distribution, a member of the beta distribution with $p = q = 1$, has a linear sfd $S_F(x) = 8(x - 1/2)$. Any other intended choice of $\eta(x)$ would lead to an unacceptable nonlinear sfd.

Johnson mapping is the innate mapping of distributions recently discussed by Kotz and Dorp (2006). The *generalized two-sided power* distribution, for instance, has density

$$f(x) = \begin{cases} n(\frac{x}{\theta})^{n-1} & \text{when } \; 0 < x \leq \theta \\ \frac{n-1}{\theta}(\frac{x}{\theta})^{n-2} & \text{when } \; \theta < x < 1 \end{cases}$$

and t-score $T_f(x) = -\frac{1}{f(x)}\frac{d}{dx}[x(1-x)f(x)]$, that is,

$$T_F(x) = \begin{cases} (n+1)x - n & \text{when } \; 0 < x \leq \theta \\ (n+2)x - (n+1) & \text{when } \; \theta < x < 1. \end{cases}$$

The *Topp-Leone* distribution with density

$$f(x) = \beta(2-2x)(2x - x^2)^{\beta-1} = \frac{2\beta}{x(1-x)}(1-x)^2 x^\beta (2-x^2)^{\beta-1}$$

has t-score

$$T_F(x) = (\beta+2)x - 2x^2 - \beta + 2(\beta-1)\frac{x^2(1-x)}{2-x^2},$$

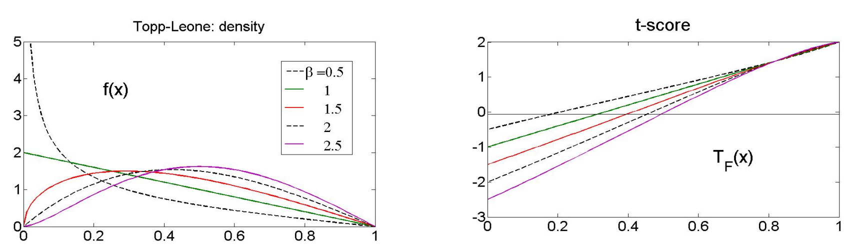

Fig. 2. Densities and t-scores of the Topp-Leone distribution.

without explicitly expressed sfd. Densities and t-scores of the distribution for some values β are shown in Fig 2.

The density of the *Kumaraswamy* distribution with resiliance parameter φ (de Pascoa et al., 2014) is $f(x) = \lambda\varphi x^{\lambda-1}(1-x^\lambda)^{\varphi-1}$. The attempt to set $\eta'(x) = 1/x(1-x^\lambda)$ was not successful; it leads to a simple 't-score' $T_F(x) = \lambda((\varphi-1)x^\lambda - 1)$, which, however, does not reduce if $\varphi = 1$ to the t-score of the power distribution. The t-score utilizing the Johnson mapping as the innate mapping is more complicated,

$$T_F(x) = -\frac{1}{f(x)}\frac{d}{dx}[x(1-x)f(x)] = (1+\lambda)x - \lambda + \lambda(\varphi-1)\frac{(1-x)x^\lambda}{1-x^\lambda},$$

but without the drawback of the previous attempt (and, unfortunately, without an explicit formula for the sfd as well).

Distributions with densities described by goniometric functions, such as Burr V or Burr XI (Johnson et al., 1994) have innate mapping $\eta(x) = \tan x$. They are shortly discussed in Fabián (2013). We must note, however, that a quantitative comparison of sfds of different distributions from any $\Pi_\mathcal{X}$ is of interest only within the class of distributions with the same innate mapping.

We arrived to the basic theorem.

Theorem 2. *For any $\mathcal{X} \neq \mathcal{R}$, the decomposition of the density $f(x)$ of any $F \in \Pi_\mathcal{X}$ into the form $g(\eta(x))\eta'(x)$ is unique.*

Proof. Either $\eta(x)$ and/or $\eta'(x)$ in the density formula (2.2) is clearly identifiable, or $f(x)$ is to be written in the form

$$f(x) = \frac{1}{\eta'(x)}f(x)\eta'(x) \tag{2.15}$$

where $\eta(x)$ is the Johnson mapping corresponding to the given support. □

3 Distribution characteristics

Based on the concept of the sfd, we introduce new numerical characteristics and functions describing continuous distributions.

Let $F_\theta \in \Pi_\mathcal{X}$ with density $f(x)$, t-score $T_F(x)$ and sfd $S_F(x)$ have a prototype $G \in \Pi_\mathcal{R}$ with density $g(y)$ and sfd $S_G(y)$. By (1.2), $S_G(y) = 0$. Let F, either a parent distribution without parameters, or a member of parametric family $\mathcal{F}_{\mathcal{X},\theta}$ satisfy the following *regularity assumptions:*

Both G and F are supposed to be absolutely continuous with respect to the Lebesgue measure, g and f are continuously differentiable a.e. according to the variable, g is unimodal and $ES_G^2 < \infty$. The last condition is in accordance with the usual regularity requirements.

3.1 Central point

We presume that any distribution could have its central point (typical value). Such a value is certainly not the mean, although usually considered to be, as the mean of some heavy-tailed distributions is infinite.

Suggestion 1. *The central point of F is its score mean, the solution x^* to the equation* (2.5).

The score mean y^* of a prototype distribution G is the mode. By Proposition 2, $T_F(x^*) = S_G(\eta(x^*))$ so that $x^* = \eta^{-1}(y^*)$. The score mean of a transformed distribution with unimodal prototype thus exists and can be explained as the *image of the mode of the prototype*. The score mean of t-location distributions is the value of the t-location parameter, the score mean of distributions of the second kind (e.g. distributions in the lower part of Table 1) is expressed by functions of parameters. Notice that, for for example, the mean $m = p/(q-1)$ of the heavy-tailed beta-prime distribution does not exist if $q < 1$, but the score mean, given by to the certain extent analogous expression $x^* = p/q$, is finite.

3.2 Score moments

Definition 5. *The k-th moment of the sfd,*

$$ES_F^k(X) = \int_{\mathcal{X}} S_F^k(x) f(x)\, dx, \qquad k = 1, 2, ... \tag{3.1}$$

is called the k-th score moment of distribution F.

Proposition 2. *Let $F = G \circ \eta \in \Pi_{\mathcal{X}}$ and x^* be the score mean. Then $ES_F^2 = [\eta'(x^*)]^2 ES_G^2$, $ES_F = 0$ and $0 < ES_F^2 < \infty$.*

Proof follows from (2.6) and Proposition 1 since x^* exists as $g(y)$ is unimodal and $\eta'(x^*)$ is a constant. □

ES_F^2 is the Fisher information for score mean. It has been shown by Fabián (2013) that ES_F^3 expresses a certain form of skewness (if $ES_F^3 = 0$ distributions with $\mathcal{X} = \mathcal{R}^+$ are 'log-symmetric'), and ES_F^4 expresses the flatness (as opposite to kurtosis).

3.3 Information function

The continuous equivalent of Shannon's entropy, the differential entropy

$$h(f) = \int_{\mathcal{X}} -\log f(x) f(x) dx = E(-\log f(x)), \tag{3.2}$$

is often considered to be the mean uncertainty contained in distribution F. However, (3.2) can be negative (the differential entropy $h(f) = 1 + \log \tau$ of the exponential distribution $f(x) = \frac{1}{\tau} e^{-x/\tau}$ is negative if $\tau < e^{-1}$). On the other hand, Fisher information (without such a defect) is defined with respect to parameters of parametric distributions only.

Function $S_F^2(x)$ increases from the least informative point $x^* : S_F^2(x^*) = 0$ to both ends of the support, being unbounded in cases of thin-tailed distributions and bounded in cases of heavy-tailed distributions.

Suggestion 2. *Function $S_F^2(x)$ is an information function of distribution F, describing relative information about the score mean contained at $x \in \mathcal{X}$.*

For prototypes and t-location transformed distributions, ES_F^2 is Fisher information with respect to the central parameter (location or t-location). In general, ES_F^2 is the (extended) Fisher information with respect to the score mean. We judge that it can be taken as the mean information of the distribution.

3.4 Measure of variability

Similarly to the mean, the classical variance is not a proper measure of variability of a distribution since it can be infinite for some heavy-tailed distributions. A measure of variability without this defect was suggested by Fabián (2007).

Definition 6. *The reciprocal mean information of F, the score variance*

$$\omega_F^2 = \frac{1}{ES_F^2}, \tag{3.3}$$

is a measure of variability of distribution F.

Fisher information of location and scale prototypes $G_{\mu,\sigma}$ is proportional to $1/\sigma^2$. The score variance $ES^2_{G_{\mu,\sigma}}$ of the normal and Gumbel distribution with density

$$g(u) = \sigma^{-1} e^u e^{-e^u}, \tag{3.4}$$

where $u = (y - \mu)/\sigma$, is $\omega^2 = \sigma^2$, Cauchy distribution with $g(u) = 1/\pi(1 + u^2)$ has score variance $\omega^2 = 2\sigma^2$ and logistic $\omega^2 = 3\sigma^2$. A score variance of a transformed $F = G \circ \eta \in \Pi_{R^+}$ is, by (3.3) and (2.6),

$$\omega_F^2 = (x^*)^2 \omega_G^2$$

where ω_G^2 is the score variance of the prototype. As an observation supporting Definition 6, ω_F^2 of light-tailed distributions is proportional to their variance.

3.5 Influence and uncertainty functions

Suggestion 3. *The normalized sfd,*

$$I_F(x) = \frac{S_F(x)}{ES_F^2}, \tag{3.5}$$

is an influence function of distribution F.

Indeed, in the class of location (t-location) distributions, $I_F(x)$ is the influence function of the ML estimator of the location (t-location) parameter.

Suggestion 4. *Function $U_F(x) = I_F^2(x)$ is an uncertainty function of F.*

Proposition 3. *The mean uncertainty associated with a random variable is its score variance.*

Proof: By (3.5) and (3.3), $EU_F = \omega_F^2$.

It has been shown by Fabián (2009) that for many distributions the square root of mean uncertainty $\omega_F \sim e^{h(f)}$, where $h(f)$ is the differential entropy (3.2) of F.

Fig. 3 shows densities, influence and uncertainty functions of the light-tailed Weibull and heavy-tailed beta-prime distributions with $x^* = 3$ and some values of ω (for Weibull distribution $\omega = 3/c$, see Table 1). Influence and uncertainty function of heavy-tailed distributions are bounded on the side of the heavy tail.

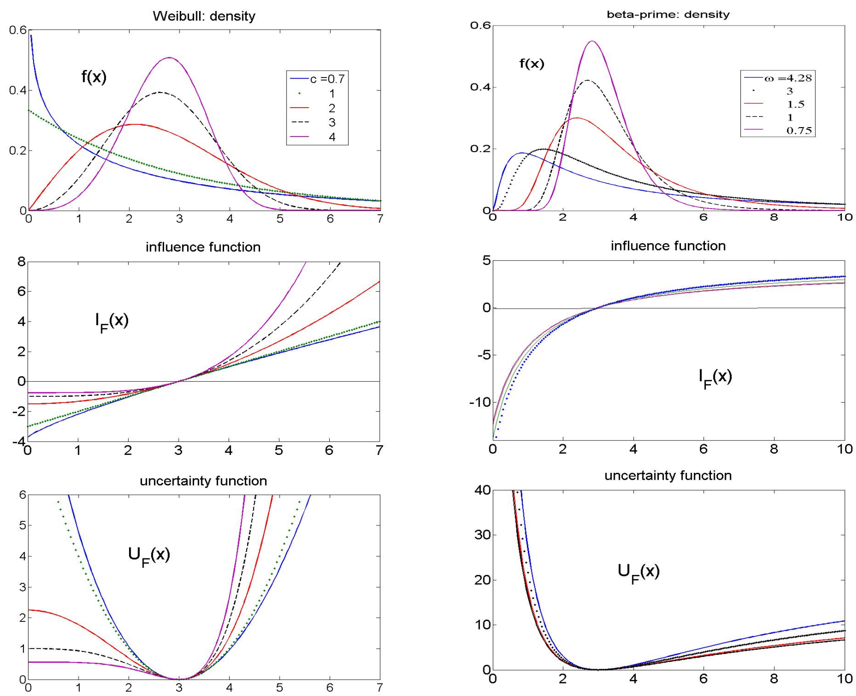

Fig. 3. Influence and uncertainty functions of Weibull and beta-prime distributions.

3.6 Weight function

As a weight of observations $\mathbf{X}_n$ iid according to some $F \in \Pi_{\mathcal{X}}$ are often taken values $w_F(x_i) = 1/f(x_i)$. We suggest a weights based on the sfd.

Suggestion 5. *Function*

$$w_F(x) = \frac{dS_F(x)}{dx}$$

is the weight function of F, describing the relative weight of a data item $x \in \mathcal{X}$.

In contrast with the sfd, the normalizing constant is not interesting from the point of view of a possible use of the weight function, so that in Suggestion 5 might be used the t-score instead of the sfd. Weight function of distributions with linear sfd (normal, gamma, beta) is $w_F(x) = 1$. Weight function of the Weibull distribution is $w_F(x) = (x/\tau)^{c-1}, c > 0$ and of the Fréchet distribution $w_F(x) = (x/\tau)^{c-1}, c < 0$. Fig. 4 shows weight functions of the concatenation of both distributions,

$$f(x) = \frac{c}{\tau}\left(\frac{x}{\tau}\right)^{c-1} e^{-(\frac{x}{\tau})^c}, \quad c \in \mathcal{R} \tag{3.6}$$

for $x^* = \tau = 1$ and $x^* = 3$ and some values of c. Value $c = 1$ corresponds to the exponential distribution.

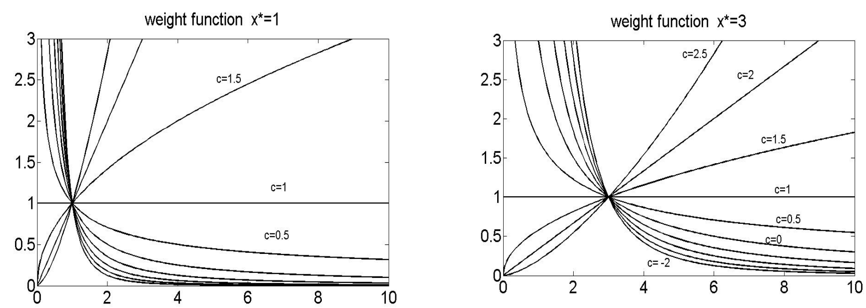

Fig. 4. Weight functions of members of family (3.6).

Function $w_F(x)$ can be used as a metric function, generating in the sample space $\mathcal{X}$ a Riemannian distance

$$d_F(x_2, x_1) = \int_{x_1}^{x_2} w_F(x)\, dx = |S_F(x_2) - S_F(x_1)|, \tag{3.7}$$

providing a one-to-one relation between probability measure F and distance (pseudo-metric) in the sample space $\mathcal{X}$.

4 Applications

4.1 Classification of models

The first attempt to systemize continuous distributions was undertaken by K. Pearson. His system of distributions is actually based on the behavior of the sfd (1.2). Unfortunately, the system does not have a clear systematic basis (Johnson et al., 1994). The reason for it follows from the previous account: formula (1.2) can not be used for transformed distributions.

From the point of view of the behavior of sfds at boundaries of the support, distributions can be classified into four basic classes, UU, BU, UB and BB, where U means unbounded and B bounded sfd. For instance, UB means unbounded on the left and bounded on the the right boundary of the support.

To give examples, we assign distributions mentioned in the paper into the respective classes:

UU: normal, skew-normal, lognormal, Johnson's U_B, GIG, (2.11)

BU: Gumbel, Weibull, gamma, loggamma

UB: extreme value, Fréchet, inverse gamma

BB: logistic, loglogistic, beta-prime, Pareto, beta, (2.3), Kumaraswamy, Topp-Leone.

4.2 Score moment estimates

Score moments of simple distributions are often simple functions of θ so that the estimate $\hat{\theta}_n$ of the 'true' θ_0 based on $\mathbf{X}_n$ assuming a model $\mathcal{F}_{\mathcal{X},\theta}$ can be obtained by means of the generalized moment estimator, the *score moment estimator* (Fabián 2001, 2010), given by the sample version of (3.1),

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

Score moment estimates, proven consistent and asymptotically normal with variance-covariance matrix given in Fabián (2001), are in general not efficient, but for distributions from the *BB* class are all the components of $\hat{\theta}_n$ robust with respect to outliers (either too large/small regular values of heavy-tailed distributions or consequences of a contamination).

Consider here, for instance, the Pareto distribution with density (2.12) and tail index $\gamma = 1/c$. By (2.14), $\gamma = x^* - 1$. Using Table 2, $\omega^2 = \gamma^2(1 + 2\gamma)$. In simulation experiments were data generated from $F_{Pareto}(\omega)$ and γ was estimated from samples of length 75 points both by the ML and score moment method. The dependence of average estimated γ on increasing ω is shown in upper panels of Fig. 45. The ML estimates have lower mean squares error (MSE). Lower four panels show the average estimated γ as a function Ω under increasing contamination represented by data generated from

$$F_c = 0.9F_{Pareto}(\omega) + 0.1F_{Pareto}(\Omega)$$

for two different values ω. It appeared that the score moment estimates of γ becomes less biased than the ML ones when approximately $\Omega > 2\omega$, and the MSE of score estimates become lower than those of ML estimates when approximately $\Omega > 3\omega$.

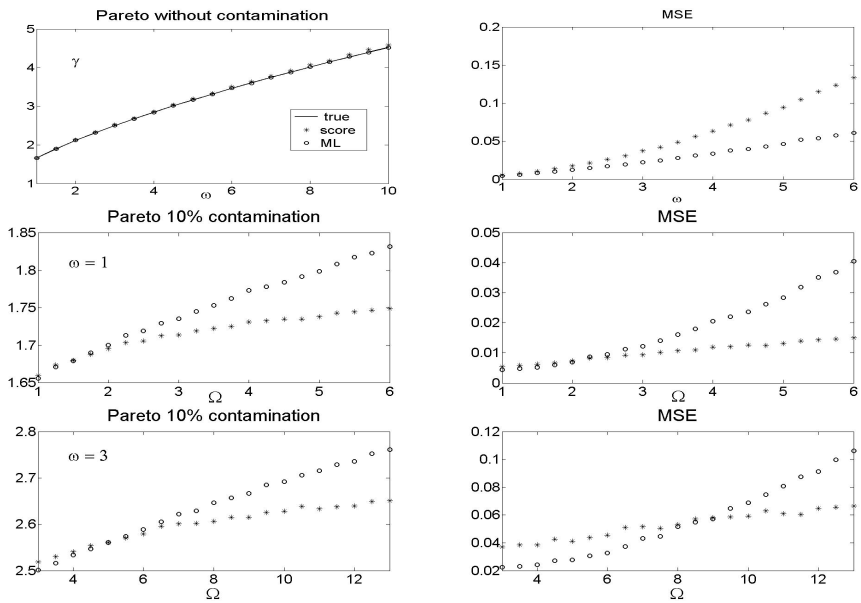

Figure 5. ML and score estimates of the tail index of the Pareto distribution.

Since sfds are scalar-valued functions, sfds of distributions from classes *BU, UB* and *UU* can be modified by robust approaches as trimming or huberising (Beran and Schell, 2012), as in Fabián (2013) in cases of Weibull and gamma distributions.

Finally, let $\hat{\theta}_n$ be a sufficient estimate of θ_0 based on $\mathbf{X}_n$. No matter by which method θ_0 was estimated, the *sample score mean* can be in general constructed as $\hat{x}_n^* = x^*(\hat{\theta}_n)$, and the *sample score variance* as $\hat{\omega}_n^2 = \omega^2(\hat{\theta}_n)$. It makes possible to describe data samples by a central point and dispersion under an arbitrary assumed parametric model and to compare results of the estimation for various models with different types and different number of parameters.

4.3 Score average

Let us take $S_F(X)$ as a random variable associated to X with distribution F. Let $X_1, ..., X_n$ be iid according to F. Let us denote by $\bar{S}_F$ the arithmetic mean

$$\bar{S}_F = \frac{1}{n}\sum_{i=1}^{n} S_F(X_i).$$

Theorem 3.

$$\lim_{n\to\infty} P(|\bar{S}_F| \geq \varepsilon) = 0.$$

As $n \to \infty$,

$$n^{1/2}\bar{S}_F \to Z \sim \mathcal{N}(0, ES_F^2).$$

Proof. The first assertion follows from the law of large numbers. Since $ES_F = 0$ by Proposition 2 and $ES_F^2 < \infty$, the second assertion immediately follows from the Central limit theorem. □

The first equation of the system (4.1) can be in some cases written in the form

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

For distributions of the first kind with one (t-location) parameter, (4.2) is identical with the ML equation. The solution $\hat{x}_n^*$ of (4.2) is $AN(x^*, \sigma_*^2)$, where $\sigma_*^2 = ES_F^2/[ES_*']^2$ and where $S_*'(x) = \frac{\partial S_F(x;x^*)}{\partial x^*}$, by the well-known result from the theory of M-estimates (e.g. Marrona et al. 2006, Theorem 10.7). The confidence intervals for x_n^* can be constructed in accordance with the Rao scores test (Fabián 2009). By Theorem 3, the asymptotic rejected region for x_n^* is given by $|\bar{S}_F(x_n^*)| \geq u_{\alpha/2}\sqrt{nES_F^2(x_n^*)}$, where $u_{\alpha/2}$ is the $\alpha/2$ quantile of the normal distribution.

In general, (4.2) is to be solved by an iterative way. However, the form of sfds of distributions in Table 1 (except the loglogistic one) is so simple that the sample score mean is given by an explicit formula $\hat{x}_n^* = S_F^{-1}(\bar{S}_{F_n})$. In such cases we speak on *score average*, in Table 1 denoted by $\bar{x}_S$. In cases of the Weibull and Fréchet distributions formula for $\bar{x}_S$ holds for a fixed value of c. It is apparent that the score average of distributions the center point with linear sfd is the arithmetic mean, whereas a typical value of heavy-tailed distributions considered here (inverse gamma, Pareto, Fréchet with c =1) is the harmonic mean.

The last observation led to a 'harmonic mean based' Hill estimator of the tail index of Pareto distribution, c.f. Stehlík et al. (2012) and Beran et al. (2014).

4.4 Observed sfd

The sfd evaluated at $\hat{\theta}_n$ is the observed sfd, as well as $ES_F^2(\hat{\theta}_n)$ is the observed Fisher information. Observed sfds of two random variables with arbitrary interval support have been used for estimation of a measure of their association. The distribution-dependent *score correlation coefficient* of random variables X and Y with marginal distributions F_X, F_Y and sfds S_X, S_Y, respectively, was defined in Fabián (2009) by

$$\rho_F(X,Y) = \rho_P(S_X(X), S_Y(Y)) \tag{4.3}$$

where $\rho_P(X,Y)$ is the Pearson's correlation coefficient. Behavior of the sample version of (4.3) were compared by means of simulation experiments (Fabián, 2013b) with behavior of correlation coefficients in current use. For distributions with 'mild' non-symmetry (with $\mathcal{X} = \Pi_{\mathcal{R}}$, gamma, Weibull) the average values of all correlation coefficients were roughly equal to the theoretical value: correlation properties have overcome the structure of distributions. However, in cases of highly non-symmetric distributions from $\Pi_{\mathcal{R}^+}$ results were strongly dependent on the variability of the distribution. Fig 5. shows the dependence of the average score, Spearman and robust (with Huber score function) correlation coefficients between X and $Y = \alpha X + (1 - |\alpha|)Z$, where X and Z were iid according to the beta-prime distribution with true ρ under increasing variability described by ω with a conclusion that only the score and Spearman correlation coefficients are capable to indicate association of random variables with heavy-tailed distributions.

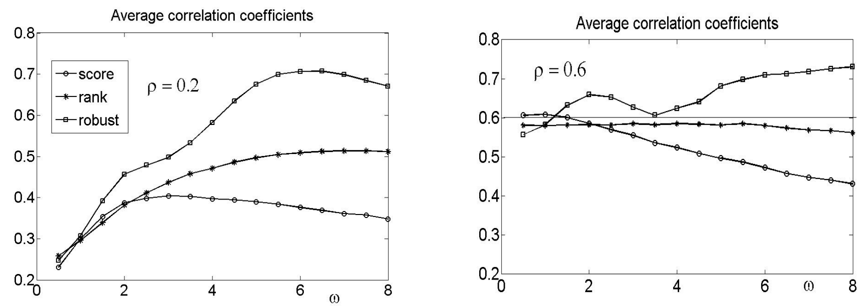

Fig. 6. Average sample correlation coefficients under increasing variability of samples.

Results could be used for distribution-dependent (and robust) spectral analysis of time series, as indicated in Fabián (2011b).

5 Conclusions

To any univariate absolute continuous distribution with arbitrary support $\mathcal{X}$ can be assigned a scalar function, the score function of distribution, a counterpart of the density, describing (for given values of parameters) the relative influence of $x \in \mathcal{X}$ with respect to the central point of the distribution, the score mean. This assignment seems to be unique: in unclear cases it is preferred the choice of the simplest innate mapping (to obtain simplest expressions for score moments) with a comparison of the resulting t-score with t-scores of similar distributions. However, a quantitative comparison of sfds of different distributions on the given support is meaningful only if distributions have the same innate mapping.

The score function of distribution (sfd) appears to be a new alternative description of continuous distributions. The score mean, the zero of the sfd, a finite value without the defect of the mean, can be taken as a center point (typical value) of a distribution. The sfd, a scalar-valued function even if the model distribution has a vector parameter, can be explained as an extended Fisher (maximum likelihood) score for this center point, either it is a value of some parameter (in this case is sfd actually the Fisher score function for it) or a function of parameters.

The score moment estimates of parameters of parametric distributions are in general not efficient, but for distributions with bounded sfds robust with respect to outliers. Unbounded sfds can be in principle easily modified by some of robust approaches.

We think that concept of the sfd provides a new direction in parametric estimation: the estimated parameters are not the final products of estimation procedures. The sample score mean and sample score variance, constructed from them, are comparable among various models with different types of parameters. Methods based on the sfd could be particularly useful in the study of skewed and heavy-tailed distributions. Moreover, we hope that the observed sfd could prove to be useful in other statistical tasks, such as, for instance, correlation and regression.

Appendix. Proof of Theorem 1. Let a location distribution $G_\mu \in \Pi_{\mathcal{R}}$ has density $g(y-\mu)$ and sfd $S_G(y-\mu)$. Consider transformed distribution $F_\tau \in \Pi_{\mathcal{X}}$ with density $f(x;\tau) = g(\eta(x)-\eta(\tau))\eta'(x)$, where τ is given by (2.8), and with score function S_F. Set $u = \eta(x) - \eta(\tau)$. Using (2.2) and the chain rule for integration, we obtain

$$\begin{aligned}\frac{\partial}{\partial\tau}\log f(x;\tau) &= \frac{1}{g(u)\eta'(x)}\frac{\partial}{\partial\tau}[g(u)\eta'(x)]\\ =\frac{1}{g(u)}\frac{dg(u)}{du}\frac{\partial u}{\partial\tau} &= S_G(u)\eta'(\tau),\end{aligned}$$

where $S_G(u) = -g'(u)/g(u))$. By Proposition 1 and (2.2) $T_F(x;\tau) = S_G(u)$ so that from $T_F(x;\tau) = 0$ it follows that $x^* = \tau$, and

$$\frac{\partial}{\partial\tau}\log f(x;\tau) = \eta'(x^*)T_F(x;\tau) = S_F(x;\tau).$$

□

Acknowledgemenst. The work was done by institutional support RVO:67985807.

References

Azzalini, A. (1985). A class of distributions which includes the normal ones. *Scand. J. Statist.*, **12**, 171-178.

Beran, J., Schell, D. (2012). On robust tail index estimation. *Comp. Statist. and Data Analysis,* **56**, 3430-3443.

Beran, J., Schell, D. and Stehlík, M. (2014). The harmonic moment tail index estimator: asymptotic distribution and robustness. *Ann. Inst. Stat. Math. Tokyo*, **66**, 193-220.

Cover, T.M. and Thomas, J.A. (1991). *Elements of Information Theory.* New York: 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 distributions. *Recent Advances in Stochastic Modeling and Data Analysis*, Ed. C. Skiadas, 43-50. 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. (2009). Confidence intervals for a new characteristic of central tendency of distributions. *Comm. Statist. Theory Methods,* **38**, 1804 - 1814.

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

Fabián, Z. (2010). Score moment estimators. Proceedings of the conference COMPSTAT 2010. Physica-Verlag Springer.

Fabián, Z. (2011). Uncertainty of random variables. In *Proceedings Stochastic Modeling Techniques and Data Analysis.* SMTDA Chania, 251-258,
link: http://www.smtda.net/images/SMTDA 2010 Proceedings.

Fabián, Z. (2011b). A new statistical tool: Scalar score function. *Comp. Technology and Application* **2**, 109-116.
Fabián, Z. (2013). Score function of distribution and revival of the moment method. *Technical Report ICS AS CR V-1189*,
link: http://hdl.handle.net/11104/0225830 (to appear in Comm. Statist. Theory Methods).
Fabián, Z. (2013b). Score function of distribution and association of random variables. *Forum Statisticum Slovakum*, **5**, 10-18.
Hampel, F. R., Rousseeuw, P. J., Ronchetti, E. M. and Stahel, W. A. (1986). *Robust Statistic. The Approach Based on Influence Functions*. New York: Wiley.
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). *Continuous univariate distributions 1.* New York: Wiley.
Jones, M. C. and Pewsey, A. (2009). Sinh-arcsinh distributions. *Biometrika*, **96**, 761-780.
Jones, M. C. (2014). On families of distributions with shape parameters, International Statistical Review. DOI: http://dx.doi.org/10.1002/insr.12055
Jurečková, J. (2012). Score functions of distributions and their role. Abstract of lecture hold on *Int. workshop on recent advances in math. stat.* in honor of prof. M. Hušková, Prague.
Kotz, S. and Van Dorp, J. R. (2004). *Beyond beta: Other continuous families of distributions with bounded support and applications.* Singapore, World Scientific.
Lawless, J. F. (2003). *Statistical Models and Methods for Lifetime Data.* 2nd ed., Hoboken, New York: Wiley.
Marrona, R. A., Martin, R. D. and Yohai, V. J. (2006). *Robust statistics. Theory and methods.* Chicester: Wiley.
de Pascoa, M. A. R., Ortega, E. M. M. and Cordeiro, G. M. (2011). The Kumaraswamy generalized gamma distribution with application in survival analysis. *Statistical methodology* **8**, 411-433.
Sen P. K., Singer J. M. and Pedroso de Lima A. C. (2009). *From Finite Sample to Asymptotic Methods in Statistics.* Cambridge Univ. Press.
Stehlík, M., Potocký, R., Waldl, H. and Fabián, Z. (2010). On the favorable estimation of fitting heavy tailed data. *Computational Statistics* **25**, 485-503.
Stehlík, M., Fabián, Z. and Střelec, L. (2012). Small sample robust testing for normality against Pareto tails. *Comm. in Statist., Simulation Computation,* **41**, 1167-1194.