Score Function of Distribution and Revival of the Moment Method. Final Version

Fabián, Zdeněk
2013

Institute of Computer Science
Academy of Sciences of the Czech Republic

SCORE FUNCTION OF DISTRIBUTION AND REVIVAL OF THE MOMENT METHOD
final version

Zdeněk Fabián

Technical report No. 1189

November 2013

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

SCORE FUNCTION OF DISTRIBUTION AND REVIVAL OF THE MOMENT METHOD
final version [1]

Zdeněk Fabián

Technical report No. 1189

November 2013

Abstract:

The paper deals with the scalar-valued score function S_F defined for a regular unimodal and continuous probability distribution F with arbitrary interval support $\mathcal{X} \in \mathbb{R}$, recently introduced by the author. The concept of the central characteristic that describes a relative influence of $x \in \mathcal{X}$ is described in a much more general way in order to improve its usefulness in parametric estimation by means of the general moment method. In particular, the whole approach is elucidated by describing it in a more suitable framework than before. Further we show that the inference function to be used in the moment estimating equations is either the score function of distribution (bounded for heavy-tailed parametric distribution models) or its modification based on the Huber's approach.

Keywords:
score function; characteristics of distributions; parametric methods; general moment method; data characteristics; Huber moment estimator;

[1] Author should like to express thanks to the anonymous referee and to prof. M.Tůma for suggestions which led to considerable improvements of the paper. The work was done with institutional support RVO:67985807 and grant of the Czech Ministry of Education, Youth and Sports under contract No. LG12020.

SCORE FUNCTION OF DISTRIBUTION AND REVIVAL OF THE MOMENT METHOD

Zdeněk Fabián
Institute of Computer Science
Academy of Sciences of the Czech Republic
Pod vodárenskou věží 2
18200 Prague
zdenek@cs.cas.cz

Key Words: score function; characteristics of distributions; parametric methods; general moment method; data characteristics; Huber moment estimator;

ABSTRACT.

The paper deals with the scalar-valued score function S_F defined for a regular unimodal and continuous probability distribution F with arbitrary interval support $\mathcal{X} \in \mathbb{R}$, recently introduced by the author. The concept of the central characteristic that describes a relative influence of $x \in \mathcal{X}$ is described in a much more general way in order to improve its usefulness in parametric estimation by means of the general moment method. In particular, the whole approach is elucidated by describing it in a more suitable framework than before. Further we show that the inference function to be used in the moment estimating equations is either the score function of distribution (bounded for heavy-tailed parametric distribution models) or its modification based on the Huber's approach.

1. INTRODUCTION

The basic objective of parametric estimation of classical statistics is to get reliable estimates of parameters involved in probabilistic description of a given set of random variables. The models of classical statistics are parametric families of distributions $\{F_\theta : \theta \in \Theta \subseteq \mathbb{R}^m\}$ with parent F with density $f(x) > 0$ on an open interval $\mathcal{X} \subseteq \mathbb{R}$ and $f(x) = 0$ on $\mathbb{R} - \mathcal{X}$. $\mathcal{X}$ is called the support of F. The estimation is typically based on averaging functions of the variables generally called *score functions* that measure sensitivity of the corresponding likelihood function with respect to its parameters. More in detail, a value $\psi(x_0; \theta)$ of a score function ψ at an observed point x_0 and for a parameter θ describes the relative influence of x_0 on the estimated characteristic under the chosen model.

The basic and most well-known score functions are vector functions with components

$$U_{\theta_j}(x;\theta) = \frac{\partial}{\partial\theta_j} \log f(x;\theta), \quad j = 1, ..., m,$$

sometimes called *Fisher scores*. These scores naturally arise in the maximum likelihood method, that provides estimates of θ of minimum variance. A disadvantage of the method is that the Fisher scores are not stable under deviations from the assumed model.

Robust statistics represent an important approach to solve the problem of reliable parametric estimation in case of the instabilities caused by outlying data deviations. This approach is based on constructing bounded score functions with the intention to suppress the influence of data outliers. For example, in case of data following a contaminated normal model $\mathcal{N}_c = (1-\varepsilon)\mathcal{N}(\mu,\sigma) + \varepsilon\mathcal{N}(\mu,\sigma_c)$ with $\epsilon << 1$ and $\sigma_c >> \sigma$, the Huber function

$$\psi(x;k) = \max[-k, \min(x,k)]$$

where k is a tuning constant, often play a role of the score function.

This paper develops and further generalizes the approach recently introduced by Fabián (2007) based on generalizing the Fisher score functions, called *score functions of distribution*. These functions are denoted here by S_F and abbreviated by sfd. Here a value $S_F(x_0)$ at $x_0 \in \mathcal{X}$ characterizes relative influence of $x_0 \in \mathcal{X}$ on a certain central characteristic of distribution F. If a parametric distribution has a parameter expressing this central characteristic, the sfd is identical with Fisher score for this parameter. It was shown that the sfd-based approach provides in some important cases (as the heavy-tailed distributions) score functions that are bounded. Further, a unique scalar-valued sfd can be used for tackling point estimation problems even if θ is a vector by means of the generalized moment method. More in detail, the moments $ES_F^k(\theta)$, where E means the expectation, exist and are often simple functions of parameters. More complex score moment estimates of θ may need more computational effort in the estimation, but, at the same time, they provide unique features such as they are robust for all components of θ for the heavy-tailed distributions with bounded sfd's.

The development of the sfd may be characterized as a generalization of the maximum likelihood method such that it captures as many of the important features of the considered probabilistic model as possible. It may be instructive to describe briefly the research that led to the development of the sfd functions since this may also explain its potential for solving new problems.

i) The starting point is the following identity valid for location distributions with support $\mathbb{R}$ and density with location parameter $\mu \in \mathbb{R}$ in the form $f(x-\mu)$:

$$\frac{\partial}{\partial \mu} \log f(x-\mu) = -\frac{1}{f(x-\mu)} \frac{d}{dx} f(x-\mu).$$

The identity says that the Fisher score for location can be obtained by differentiating of $-\log f(x-\mu)$ with respect to the variable. By setting $\mu = 0$, Hampel et al. (1986, pp.104) concluded that the relative rate of the change of f

$$S_F(x) = -f'(x)/f(x) \tag{1}$$

describes, analogically to the Fisher score, the relative influence of $x \in \mathbb{R}$ with respect to the "center" (mode) of the distribution (which is the solution of equation $S_F(x) = 0$). Consistently with this observation, Cover and Thomas (1991, pp.494) consider ES_F^2 as the Fisher information of F.

ii) Function (1), mentioned sometimes as a *generalized score function*, [cf. Sen et al. (2009)], is called the *score function of distribution* by Jurečková (2012). This concept encompasses all parametric distributions with the "full" support $\mathbb{R}$, even those without location parameter, like a distribution with the density

$$f(x;p,q) = \frac{1}{B(p,q)} \frac{e^{px}}{(e^x+1)^{p+q}}$$

where B is the beta function. The score function of distribution is in this case

$$S_F(x;p,q) = -\frac{1}{f(x;p,q)} \frac{d}{dx} f(x;p,q) = \frac{qe^x - p}{e^x + 1},$$

a function different from both Fisher scores for p and q and describing relative influence of x to the "center" of F, the solution of equation $S_F(x;p,q) = 0$.

iii) The generalized score functions (1) fail to describe distributions with support $\mathcal{X} \neq \mathbb{R}$. Examples are the uniform distribution with support $\mathcal{X} = (0,1)$ and $-f'(x)/f(x) = 0$, and the exponential one with support $\mathcal{X} = (0,\infty)$ and $-f'(x)/f(x) = 1$. However, it does not mean that a distribution with "partial" support cannot be described by a function characterizing influence of $x \in \mathcal{X}$ with respect to some its "center point".

Fabián (2001) noticed that if a distribution with support $\mathcal{X} = (0, \infty)$ has a density in the form $\frac{1}{\tau} f(x/\tau)$, it holds that

$$\frac{\partial}{\partial \tau} \log \left(\frac{1}{\tau} f(x/\tau) \right) = \frac{1}{\tau} T_F(x; \tau) \tag{2}$$

where

$$T_F(x; \tau) = -\frac{\tau}{f(x/\tau)} \frac{d}{dx} \left[x \frac{1}{\tau} f(x/\tau) \right]. \tag{3}$$

By setting $\tau = 1$ in (3), an analogue of (1) for distributions with support $(0, \infty)$ was obtained in the form

$$T_F(x) = -\frac{1}{f(x)} \frac{d}{dx} [x f(x)]. \tag{4}$$

Formula (4) was interpreted in this way: if F is taken as transformed distribution $F = G \circ \eta$ with "prototype" $G \in \mathcal{P}_{\mathbb{R}}$ and density

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

where g is the density of G, and if $\eta : (0, \infty) \to \mathbb{R}$ is $\eta(x) = \log x$, the term in square brackets of (4) is the density multiplied by the reciprocal Jacobian of the transformation. According to (2), by setting

$$S_F(x; \tau) = \frac{1}{\tau} T_F(x; \tau), \tag{6}$$

we obtained sfd's of a class of distributions with support $(0, \infty)$, which are identical with Fisher scores for certain parameter.

iv) The above observation was generalized in Fabián (2001) for distributions with arbitrary interval support and arbitrary θ by

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

with a support-dependent function $\eta(x)$ inspired by Johnson (1949)

$$\eta(x) = \begin{cases} x & \text{if} \quad \mathcal{X} = \mathbb{R} \\ \log(x - a) & \text{if} \quad \mathcal{X} = (a, \infty) \\ \log \dfrac{(x - a)}{(b - x)} & \text{if} \quad \mathcal{X} = (a, b). \end{cases} \tag{7}$$

v) To obtain score function affording the correct value of the Fisher information of parametric distributions without parameter τ, Fabián (2007) replaces the term $1/\tau = \eta'(\tau)$ in (6) by a general expression (see the next section).

Nevertheless, (7) was still not fully general.

The paper is organized as follows. In Section 2 a general definition of sfd is provided. Its first four moments are discussed in Section 3. Section 4 is devoted to the score moment estimation of the sample score mean and sample score variance, interesting namely in cases of distributions, the mean and variance of which do not exist. The new general strategy is used to introduce a robust version of the score moment method for the light-tailed distributions with unbounded sfd's. Section 6 outlines the main conclusions.

2. SCORE FUNCTION OF DISTRIBUTION

For every open interval $\mathcal{X} \subseteq \mathbb{R}$, $\mathcal{P}_{\mathcal{X}}$ be the class of probability distributions satisfying the usual regularity conditions. An increasing smooth mapping with derivative $\varphi'(x) = d\varphi(x)/dx$ is called *admissible.*

DEFINITION 1. *Let $F \in \mathcal{P}_{\mathcal{X}}$ has density $f(x)$ and $\varphi : \mathcal{X} \to \mathbb{R}$ be an admissible mapping. The most favorable mapping $\eta : \mathcal{X} \to \mathbb{R}$ for F is defined as follows: If some $\varphi'(x)$ is a proper part of the density formula (5), $\eta(x) = \varphi(x)$. In other cases $\eta(x)$ is given by (7). Set*

$$T_F(x) = -\frac{1}{f(x)}\frac{d}{dx}\left[\frac{1}{\eta'(x)}f(x)\right]. \tag{8}$$

Let the solution x^* to the equation

$$T_F(x) = 0 \tag{9}$$

exists and be unique. Function

$$S_F(x) = \eta'(x^*)T_F(x) \tag{10}$$

is called a score function of distribution (sfd of F).

The concept of the most favorable mapping secures the simplest S_F, since it is given either by

$$T_F(x) = -\frac{1}{f(x)}\frac{d}{dx}(g(\eta(x)),$$

or $f(x)$ in square brackets of (8) is multiplied by the derivative of the mapping most frequently occurring in density formulas of distributions from $\Pi_{\mathcal{X}}$.

According the following theorem, the construction described in Definition 1 leads in particular cases to Fisher scores for the most important parameter of the distribution.

THEOREM 1. Let $G \in \mathcal{P}_{\mathbb{R}}$ be a distribution with location parameter μ. For any interval support $\mathcal{X}$, the sfd of transformed distribution $F = G \circ \eta$ equals the Fisher score for parameter $\tau = \eta^{-1}(\mu)$.

Proof. Let a location distribution $G_\mu \in \mathcal{P}_{\mathbb{R}}$ has density $g(y-\mu)$ and score function $S_G(y-\mu) = U_\mu(y-\mu)$. Let us consider the transformed distribution $F_\tau \in \mathcal{P}_{\mathcal{X}}$ with density $f(x;\tau) = g(\eta(x) - \eta(\tau))\eta'(x)$, where

$$\tau = \eta^{-1}(\mu), \tag{11}$$

and with score function S_F. Set $u = \eta(x) - \eta(\tau)$. Using (5) and the chain rule for integration, we obtain

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

(where, by (1), $S_G(u) = -g'(u)/g(u)$). Since by (8)

$$T_F(x;\tau) = -\frac{1}{g(u)\eta'(x)} \frac{d}{dx} g(u) = -\frac{1}{g(u)\eta'(x)} \frac{dg(u)}{du} \frac{\partial u}{\partial x} = S_G(u) \tag{12}$$

and taking into account that the solution of equation $T_F(x;\tau) = S_G(u) = 0$ is $x^* = \tau$, it holds true that

$$U_\tau(x;\tau) = \eta'(x^*) T_F(x;\tau) = S_F(x;\tau). \tag{13}$$

□

The *transformed location* parameter (11) is taken as a central characteristic of transformed distributions of this type.

The most favorable mapping for distributions from $\mathcal{P}_{\mathbb{R}}$ is usually the identical mapping $\eta(x) = x$ and the corresponding sfd's are given by (1).

However, the most favorable mapping for distribution with support $\mathbb{R}$ and density

$$f(x) = \frac{1}{\sqrt{(1+x^2)}} \frac{e^{\sinh^{-1} x}}{(1+e^{\sinh^{-1}} x)^2}. \tag{14}$$

is $\eta(x) = \sinh^{-1} x$, since $\eta'(x) = \frac{1}{\sqrt{(1+x^2)}}$. From (8) one obtains

$$T_F(x) = \frac{e^{\sinh^{-1} x} - 1}{e^{\sinh^{-1} x} + 1}.$$

As $\eta'(0) = 1$, sfd of (14) is $S_F(x) = T_F(x)$. Obviously, (14) is the density of the transformed logistic prototype. Another examples with non-trivial Jacobians are densities of forms $f(x) = g(\arctan x)/(1+x^2)$ or perhaps $f(x) = g(\tan x)/\cosh^2 x$, where g is the density of a certain prototype.

The most favorable mapping for distributions from $\mathcal{P}_{(0,\infty)}$ is $\eta(x) = \log x$. Some parametric distributions from $\mathcal{P}_{(0,\infty)}$ and their sfd's are listed in Table I.

TABLE I. Sfd's of some parametric distributions from $\mathcal{P}_{(0,\infty)}$.

Distribution	$f(x)$	$T_F(x)$	x^*	$S_F(x)$	ES_F^2
lognormal	$\frac{c}{\sqrt{2\pi}x} e^{-\frac{1}{2}\log^2(\frac{x}{\tau})^c}$	$c\log(\frac{x}{\tau})^c$	τ	$\frac{c}{\tau}\log(\frac{x}{\tau})^c$	$\frac{c^2}{\tau^2}$
Weibull	$\frac{c}{x}(\frac{x}{\tau})^c e^{-(\frac{x}{\tau})^c}$	$c((\frac{x}{\tau})^c - 1)$	τ	$\frac{c}{\tau}((\frac{x}{\tau})^c - 1)$	$\frac{c^2}{\tau^2}$
Fréchet	$\frac{c}{x}(\frac{x}{\tau})^{-c} e^{-(\frac{x}{\tau})^{-c}}$	$c(1 - (\frac{x}{\tau})^{-c})$	τ	$\frac{c}{\tau}(1 - (\frac{x}{\tau})^{-c})$	$\frac{c^2}{\tau^2}$
log-logistic	$\frac{c}{x}\frac{(x/\tau)^c}{((x/\tau)^c+1)^2}$	$c\frac{(x/\tau)^c-1}{(x/\tau)^c+1}$	τ	$\frac{c}{\tau}\frac{(x/\tau)^c-1}{(x/\tau)^c+1}$	$\frac{3c^2}{\tau^2}$
gamma	$\frac{\gamma^\alpha}{x\Gamma(\alpha)} x^\alpha e^{-\gamma x}$	$\gamma x - \alpha$	$\frac{\alpha}{\gamma}$	$\frac{\gamma^2}{\alpha}(x - x^*)$	$\frac{\gamma^2}{\alpha}$
inv. gamma	$\frac{\gamma^\alpha}{x\Gamma(\alpha)} x^{-\alpha} e^{-\gamma/x}$	$\alpha - \gamma/x$	$\frac{\gamma}{\alpha}$	$\frac{\alpha^2}{\gamma}(\frac{1}{x^*} - \frac{1}{x})$	$\frac{\alpha^2}{\gamma}$
beta-prime	$\frac{1}{B(p,q)}\frac{x^{p-1}}{(1+x)^{p+q}}$	$\frac{qx-p}{x+1}$	$\frac{p}{q}$	$\frac{q^2}{p}\frac{x-x^*}{x+1}$	$\frac{q^3}{p(p+q+1)}$
Burr XII	$\frac{kcx^{c-1}}{(x^c+1)^{k+1}}$	$c\frac{kx^c-1}{x^c+1}$	$\frac{1}{k^{1/c}}$	$\frac{ck(x^c-(x^*)^c)}{k^{1/c}(x^c+1)}$	$\frac{c^2k^{1+2/c}}{k+2}$

Distributions in the upper half of Table I are those with transformed location parameter and sfd's equal to Fisher scores for it. The pivotal quantity $\frac{y-\mu}{\sigma}$ of a prototype distribution from $\mathcal{P}_{\mathbb{R}}$ transforms into

$$\frac{y-\mu}{\sigma} = \frac{\log x - \log\tau}{\sigma} = \log\left(\frac{x}{\tau}\right)^{1/\sigma}, \tag{15}$$

the pivotal quantity of distributions from $\mathcal{P}_{(0,\infty)}$. Let us point out here that parameter in denominator of a ratio with variable in numerator is frequently referred to as scale parameter. From our point of view it represents a typical value of a distribution. By (15), parameter c of distributions in Table I can be explained not as expressing the shape, but reciprocal scale.

Distributions in the lower part of Table I have sfd's which are yet unknown functions different from any of Fisher scores and the solution of (9) is $x^* = \exp y^*$ where y^* is the mode of the prototype distribution.

The mapping $\eta(x) = \log\log x$ is the most favorable mapping for the log-gamma distribution with support $\mathcal{X} = (1, \infty)$ and density

$$f(x; \alpha, \gamma) = \frac{\gamma^\alpha}{\Gamma(\alpha)}(\log x)^{\alpha-1}x^{-(\gamma+1)}, \tag{16}$$

since $\eta'(x) = 1/(x\log x)$. Then $T_F(x; \alpha, \gamma) = \frac{1}{f(x)}\frac{d}{dx}[x\log x f(x)] = \gamma\log x - \alpha$, $x^* = e^{\alpha/\gamma}$ and the sfd of log-gamma distribution is $S_F(x; \alpha, \gamma) = \frac{\gamma}{\alpha}e^{-\alpha/\gamma}(\gamma\log x - \alpha)$.

The Pareto distribution has support $(1, \infty)$ and density

$$f(x; c) = cx^{-(c+1)}. \tag{17}$$

According (7), $\eta(x) = \log(x - 1)$ and

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

so that $x^* = \frac{c+1}{c}$ and

$$S_F(x; c) = S_1(x; c) = c^2(1 - x^*/x) \tag{18}$$

If one try to use mapping $\varphi(x) = \log\log x$, one obtains

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

and $S_2(x; c) = ce^{-1/c}(c\log x - 1)$ which is proportional to the Fisher score for c. Fig. 1 shows both S_1 and S_2. Since a distribution with support $(0, \infty)$ and density $f(x, c) = c(x + 1)^{-(c+1)}$ (a "shifted" Pareto) is a member of the beta-prime distribution (Table I) with a bounded sfd, the sfd of the Pareto distribution is $S1$.

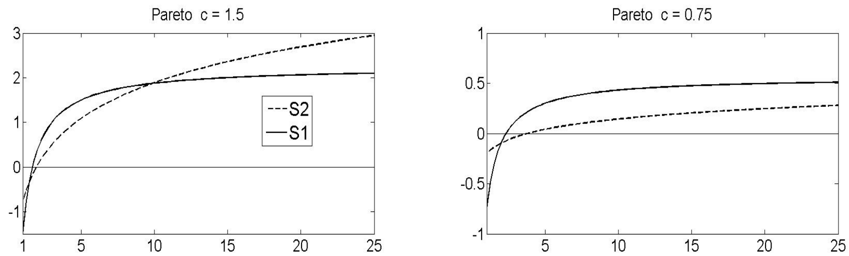

Fig 1. Sfd of Pareto distribution.

Let us consider distributions with finite support $\mathcal{X} = (a,b)$. The most favorable mapping for the beta distribution with support $(0,1)$ and density

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

is apparently (7), as $\eta'(x) = \frac{1}{x(1-x)}$. Then,

$$T_F(x) = \frac{1}{x^{p-1}(1-x)^{q-1}} \frac{d}{dx}[x^p(1-x)^q] = (p+q)x - p, \qquad (19)$$

$x^* = \frac{p}{p+q}$ and $S_F(x) = (p+q)(x/x^* - 1)$. The most favorable mapping for distribution with density

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

is $\eta(x) = -\log(-\log x)$ as $\eta'(x) = -1/x\log x$. The sfd is then

$$S_F(x) = T_F(x) = -\frac{1}{f(x)}\frac{d}{dx}[-x\log x f(x)] = \eta(x).$$

An alternative to (7) for distributions from $\mathcal{P}_{(-1,1)}$ is mapping $\eta(x) = \tanh^{-1}(x)$, $\eta'(x) = 1/\cosh^2(x)$. The most favorable mapping of distributions from $\mathcal{P}_{(-\pi/2,\pi/2)}$ described by means of goniometric functions is often $\eta(x) = \tan x$, $\eta'(x) = 1/\cos^2 x$. For instance, the sfd of a distribution with density $f(x) = e^{-x}/\kappa = \frac{1}{\cos^2 x} e^{-x}\cos^2 x/\kappa$ is

$$S_F(x) = e^x \frac{d}{dx}[e^{-x}\cos^2 x] = \sin 2x - \cos^2 x.$$

Densities and sfd's of distributions from $\mathcal{P}_{(-\pi/2,\pi/2)}$ with densities

$$1\ f(x) = e^{-x}/\kappa \qquad 3\ f(x) = \frac{1}{\sqrt{2\pi}\cos^2 x} e^{-\frac{1}{2}\tan^2 x}$$

$$2\ f(x) = e^{x}/\kappa \qquad 4\ f(x) = \sqrt{\frac{\pi}{2}} \frac{1}{(x+\pi/2)(\pi/2-x)} e^{-\frac{1}{2}\log^2 \frac{\pi/2-x}{x+\pi/2}}$$

are plotted in Fig. 2. The last two distributions have a normal prototype, the latter one is the Johnson's U_B distribution transformed into $(-\pi/2, \pi/2)$.

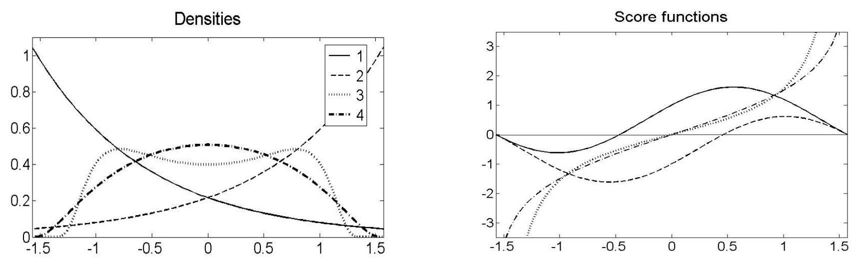

Fig 2. Densities and sfd's of distributions from $\mathcal{P}_{(-\pi/2,\pi/2)}$.

Notice that to find the sfd of $F \in \Pi_{\mathcal{X}}$ one does not usually need to determine the prototype explicitly. It suffices to identify $\eta'(x)$ at (5).

3. SCORE MOMENTS

The moments of the sfd,

$$M_k = ES_F^k(X) = \int_{\mathcal{X}} S_F^k(x) f(x)\, dx, \tag{20}$$

are called *score moments*.

Although S_F can be determined from pure knowledge of the density f, in study of score moments it is useful the concept of prototype distribution. Recall that $G \in \mathcal{P}_{\mathbb{R}}$ is a prototype of $F \in \mathcal{P}_{\mathcal{X}}$ if $F(x) = G(\eta(x))$, where η is the most favorable mapping for F.

THEOREM 2. *Let $G \in \mathcal{P}_{\mathbb{R}}$ with score function S_G be a prototype of $F \in \mathcal{P}_{\mathcal{X}}$ with score function S_F. Let $k \in \mathcal{N}$ and $|ES_G^k| < \infty$. Then,*

$$ES_F^k = [\eta'(x^*)]^k ES_G^k.$$

Proof. According the chain rule in (12),

$$T_F(x) = S_G(\eta(x)). \tag{21}$$

By (10) and (5),

$$ES_F^k = [\eta'(x^*)]^k \int_{\mathcal{X}} T_F^k(x) f(x)\, dx \;=\; [\eta'(x^*)]^k \int_{\mathcal{X}} S_G^k(\eta(x)) g(\eta(x)) \eta'(x)\, dx$$
$$= [\eta'(x^*)]^k \int_{-\infty}^{\infty} S_G^k(y) g(y)\, dy.$$

□

Let the density g of $G \in \mathcal{P}_{\mathbb{R}}$ be unimodal. If $g(y) = O(e^{-y})$ when $y \to \infty$, $S_G(y) = O(1)$. By (5), the transformed distribution $F \in \mathcal{P}_{(0,\infty)}$ has density $f(x) = g(\log x)\frac{1}{x}$ so that $f(x) = O(1/x^2)$ and $S_F(x) = O(1)$ as well. Then, ES_F^k is finite for any $k \geq 1$. Contrary to usual moments, the score moments of heavy-tailed distributions exist.

Let us clarify the meaning of score moments.

i) For any $F \in \mathcal{P}_{\mathcal{X}}$, $ES_F = 0$ due to the fact that $ES_G = 0$ and Theorem 2. By (1), the solution y^* of equation $S_G(y) = 0$ is the mode of G. By (21), $T_F(x^*) = S_G(\eta(x^*))$ so that $x^* = \eta^{-1}(y^*)$ is the transformed mode of the prototype of F. This value, which we call here a *score mean*, is taken as a *typical value* of distribution F. It exists and is unique for distributions with unimodal prototypes (distributions with multimodal prototypes could be perhaps viewed as mixtures). Let us note that there are three distributions with linear sfd's, the normal (Table II), gamma (Table I) and beta (19), score mean (typical value) x^* of which is the mean.

ii) By (13), ES_F^2 of transformed location distributions is the Fisher information for τ. Analogously, we interpret ES_F^2 of any continuous distribution as Fisher information for x^* (or simply the mean information of distribution F). Function

$$I_F(x) = S_F^2(x), \tag{22}$$

increases from the least informative point x^* in both directions to the end-points of the support interval with the average amount of information of distribution F. We interpret (22) as an *information function*, expressing relative information about x^* contained in x (Fabián, 2012).

Fisher information for x^* of some parent distributions from $\mathcal{P}_{\mathbb{R}}$ is given in Table II. Fig. 3 shows densities, score functions and information functions of two distributions from Table I with $x^* = 5$. Score functions and information functions of the Weibull distribution are unbounded when $x \to \infty$, whereas those of beta-prime distribution are bounded. In the latter case, information contained in observations near zero (with a low probability of their occurrence) is high, but finite.

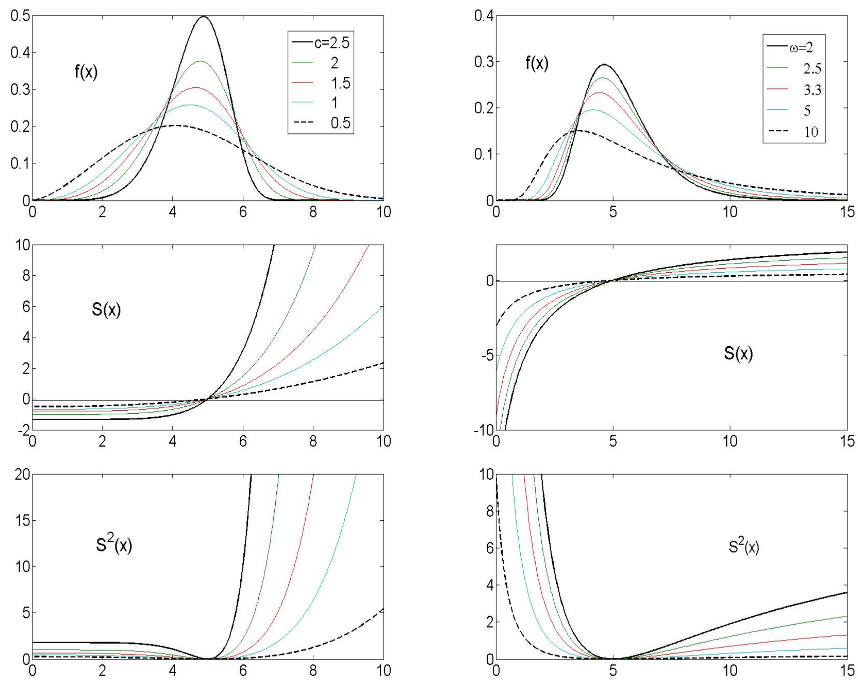

Fig. 3. Functions characterizing Weibull (left) and beta-prime distributions (right) with $x^* = 5$.

Based on analogy with the Cramér-Rao theorem for variance of efficient estimators, and by Example 1 in Hampel et al. (1986), pp.104, the reciprocal value of Fisher information called here *score variance*

$$\omega^2 = \frac{1}{ES_F^2} \tag{23}$$

was suggested by Fabián (2007) as a measure of variability of distribution F. Its square root $\omega = \sqrt{(\omega^2)}$, a *score deviation*, represents a characteristic radius of the distribution. By Theorem 2, the score variance of $F \in \mathcal{P}_{(0,\infty)}$ with prototype G is $\omega^2 = (x^*)^2/ES_G^2$. In Fig. 4 are plotted densities of Weibull and beta-prime distributions, all with $\omega^2 = 1$. Note that they look as if having a similar variability. We add that the densities in Fig. 3 differ just due to various $\omega^2 = \tau^2/c^2$ (Weibull) and $\omega^2 = p(p+q+1)/q^3$ (beta prime) of distributions.

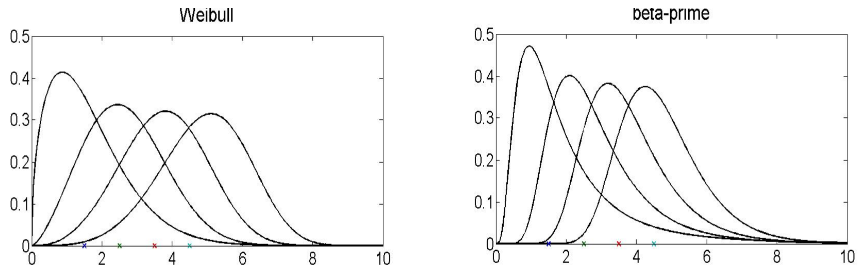

Fig. 4. Densities of distributions with equal score deviance $\omega = 1$. Typical values are marked by crosses on the x-axis.

iii) M_3 characterizes skewness. The negative/positive value of M_3 indicates a negative/positive skewness. If $M_3 = 0$, distribution can be taken as "symmetric on $\mathcal{X}$". Particularly, $M_3 = 0$ if $f(-x) = f(x)$ when $\mathcal{X} = \mathbb{R}$, $f(1/x) = x^2 f(x)$ when $\mathcal{X} = (0, \infty)$ and $f(1-x) = f(x)$ when $\mathcal{X} = (0,1)$. Note that $M_3 \neq 0$ of $F \in \mathcal{P}_{(0,\infty)}$ means a departure from the "symmetric form on $(0,\infty)$", which is itself skewed.

iv) M_4 characterizes flatness of the distribution, described by an analog of Pearson's measure of kurtosis γ_2, coefficient $\tilde{\gamma}_2 = M_4/M_2^2$. The values γ_2 and $\tilde{\gamma}_2$ of some symmetric distributions from $\mathcal{P}_\mathbb{R}$ with various behavior of sfd's are shown in Table II. The values of $\tilde{\gamma}_2$ forms a logical structure reversed to kurtosis. To obtain a clearer picture we omitted in the table $\tilde{\gamma}_2$ of non-symmetric distributions. γ_2 of the Cauchy distribution does not exist.

TABLE II. Score moments of some prototype distributions.

distribution	$f(x)$	$S_F(x)$	M_2	M_3	M_4	$\tilde{\gamma}_2$	γ_2
no name	$\frac{1}{c}e^{-x^4/4}$	x^3	2.028	0	45	10.94	1.707
normal	$\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}x^2}$	x	1	0	3	3	3
extr. value	$e^{-x}e^{-e^{-x}}$	$1-e^{-x}$	1	-2	9		
Gumbel	$e^{x}e^{-e^{x}}$	$e^x - 1$	1	2	9		
logistic	$\frac{e^x}{(1+e^x)^2}$	$\frac{e^x-1}{e^x+1}$	1/3	0	1/5	1.8	4.2
Cauchy	$\frac{1}{\pi(1+x^2)}$	$\frac{2x}{1+x^2}$	1/2	0	3/8	1.5	-
Laplace	$\frac{1}{2}e^{-\lvert x\rvert}$	sgn x	1	0	1	1	6

4. THE SCORE MOMENT METHOD AND CHARACTERISTICS OF DATA SAMPLES

The *score moment (SM) estimator* $(\hat{\theta}_{SM})_n$ is athe solution to implicit estimating equations, the finite parametric versions of (20)

$$\frac{1}{n}\sum_{i=1}^{n} S_F^k(x_i;\theta) = ES_F^k(\theta), \qquad k = 1, ..., m. \tag{24}$$

(24) is a special form of an M-estimator with inference function

$$\mathbf{\Psi}(x;\theta) = [S_F(x;\theta), S_F^2(x,\theta) - ES_F^2, ..., S_F^m(x;\theta) - ES_F^m].$$

The conditions for existence of "well-behaved" (unique, consistent and asymptotically normal) M-estimators of several parameters are well-known, see for instance Serfling (1980), Hampel et al. (1986), Marrona et al. (2006), Huber and Ronchetti (2009) etc. Since $ES_F = 0$ and higher score moments are finite, the sufficient conditions are that i) moments $ES^k(\theta)$ are differentiable with respect to any θ_k, ii) matrix $\mathbf{B}$ with elements

$$B_{jk} = E[kS_F^{k-1}(x;\theta)\frac{\partial S_F(x;\theta)}{\partial \theta_j}]_{\theta=\theta_0} - ES_F^k(\theta_0)$$

is non-singular. The last condition must be dealt with separately in each situation; in simple setups with two-parameter distributions we did not encounter any violation.

From the above considerations it follows

THEOREM 3. Let $(X_1, ..., X_n)$ be a random sample from distribution F_{θ_0}, θ_0 unknown, and let the score function of distribution F_θ, $S_F(x;\theta)$, satisfy the above conditions. The solution of equations (24) is consistent and asymptotically $\mathcal{N}(\theta_0, \mathbf{B}^{-1}\mathbf{A}(\mathbf{B}^{-1})')$, where $A = E\mathbf{\Psi}(x;\theta_0)\mathbf{\Psi}(x,\theta_0)'$.

The SM estimation equations are often quite simple and score moments are often simple functions of parameters. In what follows the SM estimators are written without suffixes.

EXAMPLE 4.1. Estimating equations (24) for Weibull distribution with semi-bounded score function (Table I) are

$$\sum_{i=1}^{n}[(x_i/\tau)^c - 1] = 0 \qquad \frac{1}{n}\sum_{i=1}^{n}[(x_i/\tau)^c - 1]^2 = 1.$$

$\hat{c}$ is thus a solution of equation $n\sum_{i=1}^{n} x_i^{2c} = 2\left(\sum_{i=1}^{n} x_i^{c}\right)^2$. From the first equation $\hat{\tau} = \frac{1}{n}\left(\sum_{i=1}^{n} x_i^{\hat{c}}\right)^{1/\hat{c}}$.

EXAMPLE 4.2. A particular case of the Pearson VI distribution, the beta-prime distribution, called also the beta of the II kind [Johnson, Kotz and Balakrishnan (1995)], is heavy-tailed if $0 < q < 2$. However, it has a bounded sfd (Table I) even when $q \geq 2$, so that SM estimate can be robust even in some cases of light-tailed distributions. Since

$$ET^2 = \int_0^\infty \left(\frac{qx-p}{x+1}\right)^2 \frac{1}{B(p,q)} \frac{x^{p-1}}{(x+1)^{p+q}}\, dx = \frac{pq}{(p+q+1)},$$

the estimating equations (24) are

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

from which

$$\hat{x}^* = \frac{\sum_{i=1}^{n} \frac{x_i}{1+x_i}}{\sum_{i=1}^{n} \frac{1}{1+x_i}}, \tag{25}$$

$\hat{p} = \hat{x}^*\hat{q}$ and $\hat{q} = (\hat{x}^*/\xi(\hat{x}^*) - 1)/(\hat{x}^* - 1)$.

Some results of Monte Carlo simulation experiments are presented in Table III. Random samples of length $n = 50$ were generated from the Weibull and beta-prime distributions. Average relative efficiencies of x^* and ω^2, given by $\mathrm{e}(\hat{x}^*) = \mathrm{var}(\hat{x}^*_{ML})/\mathrm{var}(\hat{x}^*)$ and $\mathrm{e}(\hat{\omega}^2) = \mathrm{var}(\hat{\omega}^2_{ML})/\mathrm{var}(\hat{\omega}^2)$ were determined over 10 000 samples. Their values indicate that the SM estimates of typical value have often an admissible accuracy. A general observation is that the accuracy of estimates decreases if densities have the mass concentrated near zero and very long tail. Further, in the case of Weibull distribution the accuracy of estimates of ω is low if the density has a sharp narrow peak distant from zero.

TABLE III. Efficiencies of SM estimates. Left: Weibull, right: beta-prime.

x^*	ω	0.5	1	2	3	4	0.5	1	2	3	4
1		0.96	0.93	0.93	0.88	0.83	1.0	0.99	0.87	0.82	0.73
3	e(x^*)	0.96	0.95	0.94	0.93	0.92	1.0	1.0	0.99	0.99	0.98
5		0.95	0.95	0.94	0.94	0.94	1.0	1.0	1.0	1.0	1.0
1		0.78	0.94	0.95	0.99	1.0	0.99	0.98	0.78	0.67	0.60
3	e(ω)	0.70	0.72	0.84	0.93	0.99	0.99	0.97	0.91	0.85	0.79
5		0.66	0.69	0.77	0.81	0.86	0.99	0.98	0.95	0.92	0.85

Some other examples can be found in Fabián (2010) and Stehlík et al. (2011).

The *sample score mean* $\hat{x}^* = x^*(\hat{\theta}_n)$, where $\hat{\theta}_n$ is some consistent estimate of θ, can be considered as a "center" of a random sample from F_θ.

In cases of one-parameter distributions, the first equation of system (24) can be often written as

$$\sum_{i=1}^{n} S_F(x_i; x^*) = 0 \qquad (26)$$

with x^* taken as parameter. Let us call the solution $\hat{x}^*$ to the equation (26) expressed as an explicit function of sample observations a *score average*. The asymptotic variance of score averages is clearly (c.f. Fabián, 2009)

$$\sigma_a^2 = ES^2/[\frac{\partial}{\partial x^*} S_F(x; x^*)]^2.$$

Let us give some examples. The score average of a sample from the normal distribution is the arithmetic mean, the score average of a sample from the Gumbel distribution with $f(x) = e^{x-\mu}e^{-e^{x-\mu}}$ and $S_F(x) = e^{x-\mu} - 1$ is

$$\hat{x}^* = \hat{\mu} = \log\left(\frac{1}{n}\sum_{i=1}^{n} e^{x_i}\right), \qquad (27)$$

which equals to the ML estimate of the location parameter. To give an example of a distribution without location parameter, a prototype of the gamma distribution has density $f(x; \alpha, \gamma) = \frac{\gamma^\alpha}{\Gamma(\alpha)} e^{\alpha x} e^{-\gamma e^x}$ and sfd $S_F(x; \alpha, \gamma) = \gamma e^x - \alpha$ so that $x^* = \log(\alpha/\gamma)$ and the score average of the data from it is, incidentally, given by (27) as well. Score average of a sample from the Laplace distribution is median.

Score averages of samples from members of $\mathcal{P}_{(0,\infty)}$ listed in Table I are:

i) If c is a constant, lognormal $\hat{x}^* = (\frac{1}{n}\prod x_i^c)^{1/c}$, Weibull $\hat{x}^* = (\frac{1}{n}\sum x_i^c)^{1/c}$, Fréchet $\hat{x}^* = 1/(\frac{1}{n}\sum \frac{1}{x_i^c})^{1/c}$. If $c = 1$, the score average of the lognormal distribution is the geometric mean, of Weibull the mean and of the Fréchet distribution the harmonic mean.

ii) The score average of a sample from the gamma distribution is the mean, from inverted gamma the harmonic mean, from beta-prime distribution is given by formula (25), from Pareto distribution (17) with sfd (18) the harmonic mean and score average of the sample from the log-gamma distribution (16) is $\hat{x}^* = \frac{1}{n}\sum \log x_i$.

Further, the estimate of the score variance, the *sample score variance*, is given by $\hat{\omega}^2 = \omega^2(\hat{\theta}_n)$ or as a finite version of (23), that is,

$$\hat{\omega}^2 = \frac{n}{\sum_{i=1}^n S_F^2(x_i;\hat{\theta}_n)}.$$

5. ESTIMATION IN THE PRESENCE OF OUTLYING VALUES

If the data iid according distributions with unbounded or semi-bounded sfd's are contaminated, it is necessary to modify inference function by the use of some of procedures of robust statistics. Since S_F is a scalar function, such modification is in principle easy to apply.

To obtain robust score moment estimators for distributions with unbounded or semi-bounded score functions, we chosen Huber's suggestion (1964) modified by Huber and Ronchetti (2005) to use as an inference function of distributions from $\mathcal{P}_{\mathbb{R}}$

$$\psi(x) = \begin{cases} S_F(x-\mu) & \text{if} \quad |x-\mu| \le v \\ r\,\text{sgn}(x-\mu) & \text{if} \quad |x-\mu| > v, \end{cases}$$

where v is some bound and $r = S_F(v)$. According to Beran and Schell (2010), the procedure is called "huberizing".

DEFINITION 2. *Let $S_F(x;\theta)$ be sfd of $F_\theta \in \mathcal{P}_{\mathcal{X}}$ where $\mathcal{X} = (a,b)$ and let $a \le u < v \le b$. Set*

$$\psi_k(x;\theta) = [S_F^k(x;\theta)]_u^v - E_\theta\{[S_F^k(x;\theta)]_u^v\}, \tag{28}$$

where $[y]_u^v = \min(\max(y, v), u)$. *The M-estimator* $(\hat{\theta}_H)_n$ *defined as the solution of equations*

$$\sum_{i=1}^{n} \psi_k(x_i; \theta) = 0, \qquad k = 1, ..., m \tag{29}$$

will be called a huberized score moment estimator.

THEOREM 4. Let $(\hat{\theta}_H)_n \to_p \theta_0$, $E\psi_k(x;\theta)$ be differentiable at θ_0, and ψ_k be continuously differentiable according θ_k. Let matrix $\mathbf{B}$ of derivatives with elements $\dot{\psi}_{jk} = \partial\psi_k/\partial\theta_j|_{\theta=\theta_0}$ be nonsingular, $|\dot{\psi}_{jk}(x;\theta)| \leq K(x)$ for $j, k = 1, ..., m$ where $EK(x) < \infty$, and $E|\psi_k(x;\theta_0)|^2$ be finite. Then,

$$\sqrt(n)(T_n - \theta_0) \to_d \mathcal{N}_p(0, \mathbf{B}^{-1}\mathbf{A}(\mathbf{B}^{-1})')$$

where $\mathbf{A} = E\psi_{\mathbf{k}}(x;\theta_0)\psi_{\mathbf{k}}(x;\theta_0)'$.

Proof. Assumptions of the theorem agree with assumptions of the well-known result (cf. Theorem 10.11, Maronna et al., 2006). □

Set

$$I_{k|cd}(\theta) = \int_c^d S_F^k(x;\theta)\, dF_\theta(x),$$

$I_{ku}(\theta) = S_F^k(u;\theta)F_\theta(u)$ and $I_k^v = S_F^k(v;\theta)(1 - F_\theta(v))$. Equations (29) can be then written in the form

$$\frac{1}{n}\sum_{i=1}^{n} S_F^k(\tilde{x}_i;\theta) - ES^k(\theta) = -\{I_{k|au}(\theta) + I_{k|vb}(\theta)\} + \{I_{ku}(\theta) + I_k^v(\theta)\}, \tag{30}$$

where

$$\tilde{x}_i = \begin{cases} r_1 & \text{if} \quad x_i < u \\ x_i & \text{if} \quad u \leq x_i \leq v \\ r_2 & \text{if} \quad x_i > v, \end{cases} \tag{31}$$

where $r_1 = S_F(u;\theta_{in}), r_2 = S_F(v;\theta_{in})$ and where θ_{in} is some initial value of θ. As initial robust estimates of x^* and ω can be used $\hat{x}_0^* = \text{median}(\mathbf{x})$ and $\hat{\omega}_0 = q\text{MAD}(\mathbf{x})$, where $\text{MAD} = \text{median}(|x - \text{median}(x)|)$ and q is a constant. Initial estimates of the parameter vector $\theta = (\theta_1, \theta_2)$ were determined as

$$\theta_{in} = \theta(\hat{x}_0^*, \hat{\omega}_0). \tag{32}$$

For two-parameter distributions, relation (32) is usually one-to-one.

This general scheme will be now used for a study of properties of huberized score moment estimators of simple distributions with unbounded or semi-bounded score functions.

EXAMPLE 5.1 Normal distribution $\mathcal{N}(\mu,\sigma)$, $x^* = \mu$, $\omega = \sigma$. Set $u = \mu_0 - r\sigma_0, v = \mu_0 + r\sigma_0$. The huberized sfd is

$$\psi(x) = \begin{cases} -r & \text{if} \quad x < u \\ \frac{x-\mu}{\sigma} & \text{if} \quad u \le x \le v \\ r & \text{if} \quad x > v. \end{cases} \tag{33}$$

Since $E\psi = 0$ and, by (30),

$$E\psi^2 = 1 - \frac{2}{\sqrt{2\pi}} \int_r^\infty \xi^2 e^{-\frac{1}{2}\xi^2}\, d\xi + r^2 \frac{2}{\sqrt{2\pi}} \int_r^\infty e^{-\frac{1}{2}\xi^2}\, d\xi,$$

it follows from (29) that $\hat{\mu}_H = \frac{1}{n}\sum_{i=1}^n \tilde{x}_i$ and

$$\hat{\sigma}_H^2 = \frac{\frac{1}{n}\sum_{i=1}^n (\tilde{x}_i - \hat{\mu}_H)^2}{1 - \sqrt{\frac{2}{\pi}} r e^{-\frac{1}{2}r^2} + (r^2-1)(1-\mathrm{erf}(r/\sqrt{2}))}. \tag{34}$$

In simulation experiments, 2 000 samples of length $n = 50$ were taken from a contaminated distribution

$$F_{cont}(\mu,\sigma) = (1-\epsilon)\Phi(0,1) + \epsilon\Phi(0,1+k)$$

with $\epsilon = 0.1$. Average ML and huberized score moment (H) estimates of σ are plotted together with their standard deviations against increasing k for different r in Fig. 5. The ML estimates with increasing k are increasing linearly, the huberized estimates are much useful, but higher than the true value, indicating thus a contamination.

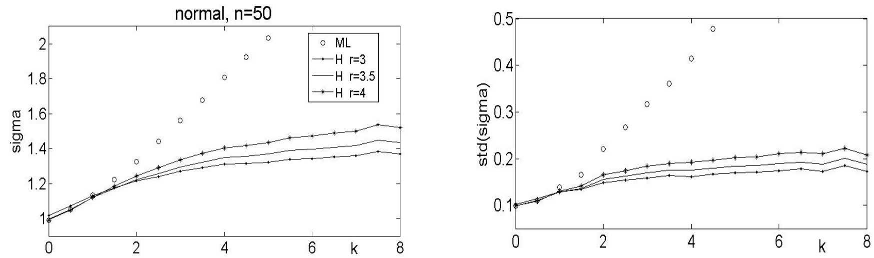

Fig. 5. Estimates of σ of contaminated $\mathcal{N}(0,1)$ under increasing contamination, left: $\hat{\theta}_H$, right: $\mathrm{std}(\hat{\theta}_H)$.

EXAMPLE 5.2. Weibull distribution (Table 2) has the sfd when $x \to \infty$. Let us take as an inference function the tapered simplified sfd,

$$\psi(x;\tau,c) = \begin{cases} (x/\tau)^c - 1 & \text{if} \quad x \le v \\ r & \text{if} \quad x > v, \end{cases} \tag{35}$$

where $r = (v/\tau)^c - 1$. The first and third members of r.h.s. of (30) are zero. Denote by $\lambda(d)$ function

$$\lambda(d) = \int_v^\infty [(x/\tau)^c - 1]^d \frac{c}{\tau}(x/\tau)^{c-1} e^{-(x/\tau)^c}\, dx = \int_w^\infty (\xi - 1)^d e^{-\xi} d\xi$$

where $w = (v/\tau)^c$. Since $I_k^v(\theta) = r^k \int_w^\infty [1 - (1 - e^{\xi})]\ d\xi$, the estimation equations (30) are

$$\begin{aligned} \frac{1}{n}\sum_{i=1}^n ((\tilde{x}_i/\tau)^c - 1) &= -\lambda(1) + r\lambda(0) \\ \frac{1}{n}\sum_{i=1}^n ((\tilde{x}_i/\tau)^c - 1)^2 - 1 &= -\lambda(2) + r^2\lambda(0). \end{aligned} \tag{36}$$

Set now

$$v = \tau_0 + k\omega_0 = \tau_0(1 + k/c_0)$$

where $\tau_0 = \mathrm{median}(x)$ and $\omega_0 = \mathrm{MADN}(x) = \mathrm{MAD}(x)/0.675$. Then $w = (1 + k/c_0)^{c_0}$, $r = w - 1$, $\lambda(0) = e^{-w}$, $\lambda(1) = we^{-w}$ and $\lambda(2) = (1 + w^2)e^{-w}$ so that the equations turn into

$$\begin{aligned} \tau^c &= \frac{\frac{1}{n}\sum_{i=1}^n \tilde{x}_i^c}{1 - e^{-w}} \\ \tau^{2c} &= \frac{\frac{1}{n}\sum_{i=1}^n \tilde{x}_i^{2c}}{2[1 - (w+1)e^{-w}]}. \end{aligned}$$

By subtracting the second equation from the square of the first, we obtain by an iterative way $\hat{c}_H$, and then $\hat{\tau}_H = \left(\frac{1}{n}\sum_{i=1}^{n} \tilde{x}_i^{\hat{c}_H}/(1-e^w)\right)^{1/\hat{c}_H}$. As a result we obtain the huberized score moment estimates $\hat{\tau}_H$ of typical value and $\hat{\omega} = \hat{\tau}_H/\hat{c}_H$ of score deviance as functions of k.

Let us give some results of simulation experiments. We refer the density of any two-parameter distribution as a function of x^* (9) and ω (23). In cases of distributions with support $(0, \infty)$, the contaminated distribution was in the form

$$f_c(x^*, \omega) = (1-\epsilon)f(x^*, \omega) + \epsilon f(x^* + k, \omega)$$

with fixed $\epsilon = 0.1$. Average ML and H estimates are plotted together with their standard deviations against increasing k for some tuning values r in Fig. 6. Similarly as in the previous case, ML estimates of a positive random variable are increasing linearly with increasing k, the huberized estimates stabilize at certain level, which is, however, higher than the true value, indicating thus contamination.

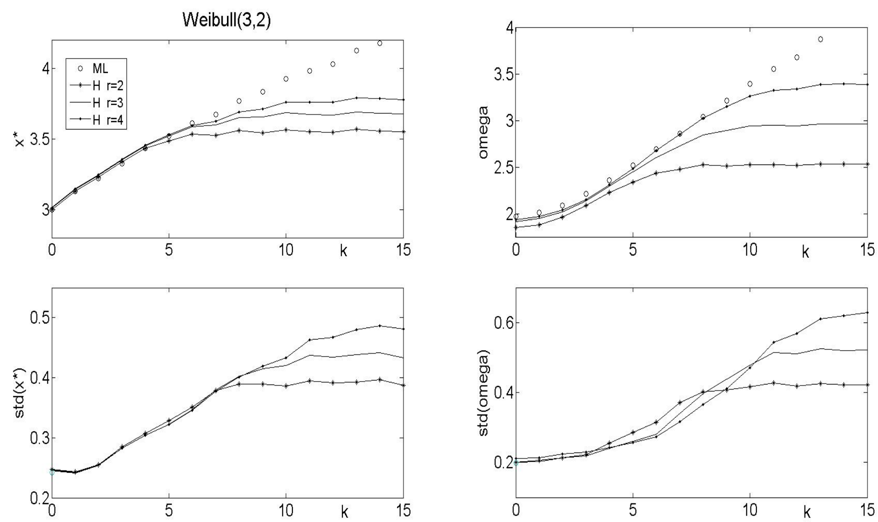

Fig. 6. Average estimate of typical value and score deviation of contaminated Weibull distribution and their standard deviations.

Average efficiencies of huberized moment estimates for various combinations of τ and ω are presented in Table IV. The main technical problem appeared to be the choice of initial values x_0^* and ω_0. Estimates successfully used for contaminated normal (median and MAD) can be used in cases of data from skewed distributions from $\mathcal{P}_{(0,\infty)}$ only in cases that ω is not too large with respect to x^*, that is, in cases of densities with a relatively sharp peaks or densities quickly decreasing to zero. In cases where $x^* < \omega$ (in Table IV marked by "$-''$), it is to use other input values. The problem needs further investigations.

TABLE IV. Comparison of efficiencies of SM and H estimates for Weibull.

x^*	ω	0.5	1	2	3	0.5	1	2	3
1		0.96	0.93	0.93	0.88	0.94	0.93	-	-
3	e(x^*)	0.96	0.95	0.94	0.93	0.95	0.95	0.94	0.95
5		0.95	0.95	0.94	0.94	0.95	0.94	0.95	0.94
1		0.78	0.94	0.95	0.99	1.07	0.95	-	-
3	e(ω)	0.70	0.72	0.84	0.93	0.74	0.93	1.0	0.96
5		0.66	0.69	0.77	0.81	0.71	0.78	0.99	1.09

EXAMPLE 5.3. The sfd of the Gamma distribution (Table I) is an semi-bounded function $S_F(x) = \frac{x-x^*}{\omega^2}$, $x^* = \alpha/\gamma$ and $\omega = \alpha/\gamma^2$. By setting $u = 0$ and $v = x_0^* + k\omega_0$, the huberized sfd is

$$\psi(x) = \begin{cases} x - x^* & \text{if} \quad x \le v \\ r & \text{if} \quad x > v. \end{cases}$$

By observing that $E(x - x^*)^2 = \omega^2$, we tried to use simplified equations (29)

$$\begin{aligned} \frac{1}{n}\sum(\tilde{x}_i - \hat{x}_H^*) &= 0 \\ \frac{1}{n}\sum(\tilde{x}_i - \hat{x}_H^*)^2 &= \hat{\omega}_H^2, \end{aligned}$$

where $\tilde{x}_i$ are given by (31). Surprisingly, even biased solutions from these simple equations (Fig. 7) are reasonably efficient.

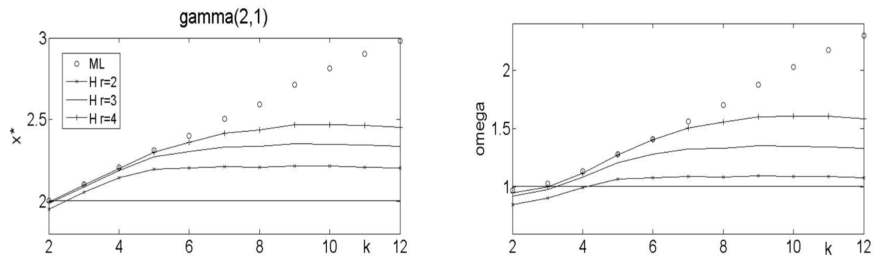

Fig. 7. Average estimate of typical value and score deviation of contaminated gamma distribution.

Fig. 8 shows the ML and H estimates and 10% and 20% trimmed mean of typical value x^* of the gamma distribution contaminated by the same way as in Example 5.2. $\hat{x}^*_{ML}$ is approximately linearly increasing and the trimmed mean depends on the "guessed" percent of contamination. Trimmed mean is a very unstable estimate, which documents the behavior of standard deviations.

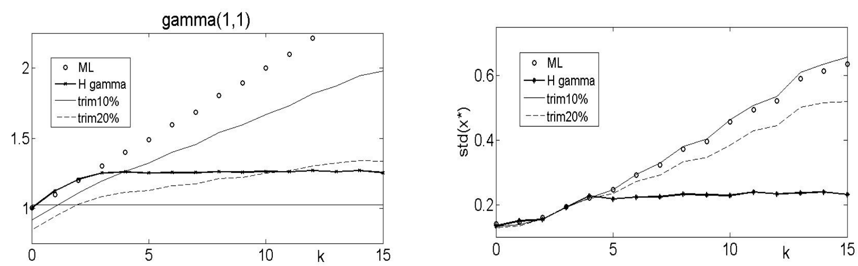

Fig. 8. Robust estimates under increasing contamination.

Fig. 9. shows that the assumption on the underlying distribution is important. Distributions gamma(x^*, ω) and Weibull(x^*, ω), in case $x^* = 1, \omega = 1$ identical, are rather different distributions if $x^* = 3, \omega = 2$. The data generated from both distributions with these values were estimated by both huberized gamma and huberized Weibull estimators. Average values of $\hat{x}^*$ exhibit a large bias when using an improper model.

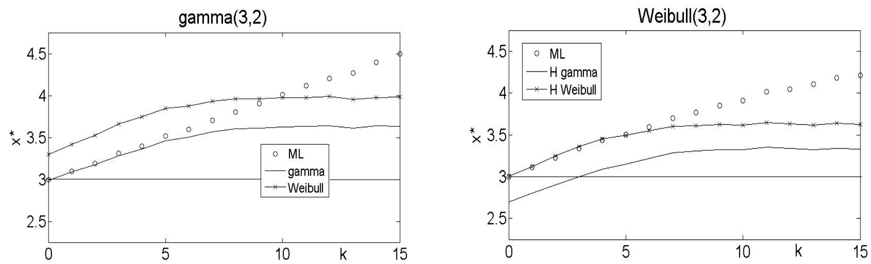

Fig. 9. Using the proper and improper estimator. Data are generated from a distribution stated in the headline.

A comparison of models F and G can be based on the score divergence, a function of θ suggested in a slightly different form as core divergence by Fabián and Vajda (2003),

$$D_{FG} = \int_{\mathcal{X}} \left(S_G(x;\theta) - S_F(x;\theta)\right)^2 f(x;\theta)\, dx,$$

where S_F and S_G are the corresponding sfd's.

EXAMPLE 5.4. K=2 000 samples of length n=50 were generated both from Weibull($x^* = 1, \omega$) and gamma($x^* = 1, \omega$) for increasing ω, and their parameters were estimated under assumption of both F : Weibull and G : gamma. Fig. 10 shows the empirical distance

$$D_{FG}(\omega) = \frac{1}{Kn} \sum_{k=1}^{K} \sum_{i=1}^{n} \left[\frac{\hat{c}}{\hat{\tau}}((x_i/\hat{\tau})^{\hat{c}} - 1) - \frac{x_i - \hat{x}^*}{\hat{\omega}^2} \right]^2$$

as functions of increasing score deviation ω of the generated samples. Estimates were determined by both ML and SM method. For samples from the Weibull, the SM method affords indiscernibly worse efficiencies, but is much more robust if the data originate from the gamma.

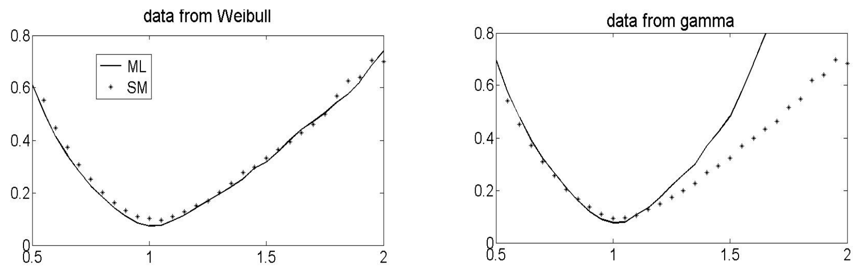

Fig. 10. Observed score divergence of the Weibull and gamma distributions as functions of ω.

6. CONCLUSIONS

Score function of distribution (sfd) is a function describing relative influence of an observation on central characteristic of the given model with support $\mathcal{X}$. The concept is based on a finding that any density formula can be explained as a product of some transformed basic form and Jacobian of the transformation $\eta : \mathcal{X} \to \mathbb{R}$. In the paper, we developed this concept in a general way.

Sfd's can be used for parametric estimation by means of the general moment method. The score moment estimates are often not efficient, but in cases of bounded sfd's (a characteristic property of heavy-tailed distributions) robust for all the components of parametric vector. By using inference function in the form of a huberized sfd, one can obtain a suitable tradeoff between efficiency and robustness of estimators even if light-tailed distributions are highly non-symmetric. As an unsolved problem remains the choice of initial values of parameters of iterative procedures.

Further, the estimates of parameters of the assumed parametric model need not be final results of parametric inference. The more interesting values are the sample score mean (as a typical value of the sample), the sample variance (as a variability of the sample) and, perhaps, the higher score moments, as functions of estimated parameters, which enable comparing of results of estimation under various differently parametrized models.

ACKNOWLEDGEMENT. Author should like to express thanks to the anonymous referee and to prof. M.Tůma for suggestions which led to considerable improvements of the paper. The work was done with institutional support RVO:67985807 and grant of the Czech Ministry of Education, Youth and Sports under contract No. LG12020.

BIBLIOGRAPHY

Beran J. and Schell, D. (2010). On robust tail index estimation. *Computational Statistics and Data Analysis,* doi:10.1016/j.csda.2010.05.028.

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

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

Fabián, Z., Vajda, I. (2003). Core functions and core divergences of regular distributions. *Kybernetika* 39, 29–41.

Fabián, Z. (2007). Estimation of simple characteristics of samples from skewed and heavy-tailed distribution. In Skiadas, C. (ed.) *Recent Advances in Stochastic Modeling and Data Analysis*, Singapore, World Scientific, 43–50.

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. (2010). Score moment estimators. In *Proc. of conference COMPSTAT*, Physica-Verlag, Springer.

Fabián, Z. (2011). A new statistical tool: Scalar score function. *Comp. Technology and Application* 2, 109–116.

Fabián, Z. (2012). Uncertainty of random variables. Proceedings SMTDA 2010, http://www.smtda.net, 251-258.

Hampel, F. R., Rousseeuw, P. J., Ronchetti, E. M. and Stahel, W. A. (1986). *Robust statistic. The approach based on influence functions,* Wiley, N. York.

Huber, P. J. (1964). Robust estimation of a location parameter. *The Annals of Mathematical Statistics* 36, 1753–1758.

Huber, P. J., Ronchetti, E. M. (2009). *Robust statistics.* Wiley, New York.

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

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

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.

Marrona, R. A., Martin, R. D. and Yohai, V. J. (2006). *Robust statistics. Theory and methods.* Wiley, Chicester.

Serfling, R. J. (1980). *Approximation theorems of mathematical statistics*, Wiley.

Sen, P.K., Singer J.M., Pedroso da Lima, A. (2010). *From finite sample to asymptotic methods of statistics.* Cambridge univ. press, New York.

Stehlík, M., Potocký, R., Waldl, H. and Fabián, Z. (2011). On the favorable estimation for fitting heavy tailed data. *Comput. Stat.* 25, 485–503.