Score Function of Distribution and Revival of the Moment Method

Fabián, Zdeněk
2012

Institute of Computer Science
Academy of Sciences of the Czech Republic

SCORE FUNCTION OF DISTRIBUTION AND REVIVAL OF THE MOMENT METHOD

Zdeněk Fabián

Technical report No. 1176

December 2012

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 [1]

Zdeněk Fabián

Technical report No. 1176

December 2012

Abstract:

A score function of distribution with arbitrary interval support is introduced. The function was suggested recently by the author under various names, but its definition was not sufficiently general. In some particular cases it equals the (likelihood) partial score function for the most important parameter of the distribution, in other cases it is yet unknown function. It is shown that it appears to be a suitable function for using in the general moment method. Score functions of distribution of heavy-tailed distributions are bounded and the corresponding moment estimators robust. For distributions with unbounded score function of distribution, robust version of the moment method is introduced.

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

[1] This work is done with institutional support RVO:67985807.

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
18207 Prague
zdenek@cs.cas.cz

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

ABSTRACT.

A score function of distribution with arbitrary interval support is introduced. The function was suggested recently by the author under various names, but its definition was not sufficiently general. In some particular cases it equals the (likelihood) partial score function for the most important parameter of the distribution, in other cases it is yet unknown function. It is shown that it appears to be a suitable function for using in the general moment method. Score functions of distribution of heavy-tailed distributions are bounded and the corresponding moment estimators robust. For distributions with unbounded score function of distribution, robust version of the moment method is introduced.

1. INTRODUCTION

For every open interval $\mathcal{X} \subseteq R$, let $\Pi_{\mathcal{X}}$ be the class of distributions F absolutely continuous with respect to the Lebesgue measure λ, supported by $\mathcal{X}$ and with well defined derivatives $f'(x) = df(x)/dx$ of the respective Lebesgue densities $f = dF/d\lambda$. In parametric setup, an unknown distribution F of iid random variables $X_1, ..., X_n$ is supposed to be a member of a parametric family $\mathcal{F}_{\mathcal{X},\theta} = \{F_\theta : \theta \in \Theta \subseteq \mathbb{R}^m\}$ with densities $f(x;\theta)$. The problem is to find such $\hat{\theta} \in \Theta$ based on a random sample from F, for which $F_{\hat{\theta}}$ is a good approximation of F.

There are two classical ways to solve this problem.

i) Let S be a suitable function defined on $\mathcal{X}$, and X be a random variable with distribution F. For $k \in \mathcal{N}$, moments of random variable $S(X)$ are

$$M_k = ES^k(X) = \int_{\mathcal{X}} S^k(x)f(x)\,dx. \tag{0.1}$$

The moment method consists in estimation of θ from finite parametric versions of (0.1). The approach provides moment estimates $\hat{\theta}_n^S$ as a solution of equations

$$\frac{1}{n}\sum_{i=1}^{n} S^k(x_i;\theta) = ES^k(\theta), \qquad k = 1, ..., m. \tag{0.2}$$

ii) Taking $\log f(x;\theta)$ as function of θ, the (likelihood) score function is the vector $U(x;\theta) = (U_{\theta_1}(x;\theta), ..., U_{\theta_m}(x;\theta))$ of gradients of the log-likelihood, that is

$$U_{\theta_k}(x;\theta) = \frac{\partial}{\partial\theta_k}\log f(x;\theta). \tag{0.3}$$

The maximum $\hat{\theta}_n^{ML}$ of the log-likelihood surface in Θ is a solution of the maximum likelihood (ML) estimating equations

$$\sum_{i=1}^{n} U_{\theta_k}(x_i;\theta) = 0, \qquad k = 1, ..., m. \tag{0.4}$$

Although the ML method yields estimates with the lowest variance and, consequently, the best approximated density $f(x;\hat{\theta}_n^{ML})$, it does not offer numerical characteristics of observed samples. On the other hand, $\hat{\theta}_n^S$ is often not efficient, but the method provides direct sample characteristics, the sample moments. However, since a suitable function in (0.1) has been unknown, the function

$$S(X) = X \tag{0.5}$$

has been used, yielding "Euclidean moments", the mean value EX and central moments $E(X-EX)^k$. The sample values of moments, particularly the sample mean $\bar{x}$ and sample variance, are often used as data characteristics, excepting cases in which function (0.5) does not match the density to such an extent that some of integrals (0.1) do not exist. But how to characterize data from such distributions?

Curiously, in case $m = 1$ there exist distributions for which both the moment and ML solutions coincide. They are the location distributions $F_\mu \in \mathcal{P}_\mathbb{R}$ with densities in the form $f(x-\mu), \mu \in \mathbb{R}$. The score function for μ is

$$U_\mu(x-\mu) = \frac{\partial}{\partial\mu}\log f(x-\mu) = -\frac{f'(x-\mu)}{f(x-\mu)}. \tag{0.6}$$

By using in (0.1) function

$$S(x) = -\frac{f'(x)}{f(x)}, \tag{0.7}$$

the ML equation (0.4) for μ and the first moment equation (0.2) become identical since $ES = 0$. Function (0.7) can be taken as a *score function of distribution* F having its support the entire $\mathbb{R}$, cf. Hampel et al. (2086, pp.104), Jurečková (2012). Accordingly, value ES^2 can be taken as Fisher information of distribution F, cf. Cover and Thomas (1991, pp.494). The generalization of the r.h.s. of (0.6) by Huber (1964) gave rise to methods of robust statistics.

By differentiating parametric density $f(x;\theta)$ according the variable as in (0.7) one obtains a parametric score function $S(x;\theta)$ of distribution $F_\theta \in \mathcal{P}_\mathbb{R}$. If it is bounded, estimates of all components of θ from equations (0.2) are robust, which could be, apart from possible non-efficiency, an important advantage of the score moment estimators with respect to the maximum likelihood ones in cases that data are contaminated. Moreover, the solution $x^*(\theta)$ of equation $S(x;\theta) = 0$ is the mode, which is an excellent central characteristic of distributions from $\mathcal{P}_\mathbb{R}$.

However, this promising approach is not used in mathematical statistics. The reason is, we judge, that function (0.7) cannot be used universally, since for distributions from $\mathcal{P}_\mathcal{X}$ with $\mathcal{X} \neq \mathbb{R}$ it looses any sense (uniform distribution !). A generalization of the concept for distributions with support $\mathcal{X} \neq \mathbb{R}$ has been recently suggested by Fabián (2001) who noticed that for distributions from $\mathcal{P}_{(0,\infty)}$, an analogue of (0.7) is function

$$T(x) = -\frac{1}{f(x)}\frac{d}{dx}[xf(x)]. \tag{0.8}$$

The reason was that for distributions with densities in the form $f(x/\tau) = \frac{1}{\tau}\tilde{f}(\frac{x}{\tau})$ it holds true that

$$S(x;\tau) \equiv \tau T(x;\tau) = U_\tau(x;\tau), \tag{0.9}$$

where U_τ is the (likelihood) score function for τ. By Fabián (2001), the explanation of formula (0.8) is as follows: if one takes F as a transformed distribution $F(x) = G(\eta(x))$ with $G \in \mathcal{P}_\mathbb{R}$ and $\eta : \mathcal{X} \to \mathbb{R}$ is given by $\eta(x) = \log x$, the term in brackets of the formula (0.8) is the density multiplied by the reciprocal Jacobian of the transformation.

Fabián (2001) generalized this observation for distributions with arbitrary interval support by using, instead of $\eta(x) = \log x$, support-dependent mappings inspired by Johnson (1949), which, being Fabián (2001) too special, stabilized in later works into

$$\eta(x) = \begin{cases} x & \text{if} \quad \mathcal{X} = \mathbb{R} \\ \log(x-a) & \text{if} \quad \mathcal{X} = (a,\infty) \\ \log\dfrac{(x-a)}{(b-x)} & \text{if} \quad \mathcal{X} = (a,b). \end{cases} \tag{0.10}$$

It turned out, however, that for purposes of the moment estimation, mappings (0.10) are not sufficiently general.

In this paper we try to show that:

i) To derive the score function of a distribution with arbitrary interval support, there is no need to prescribe any specific mapping $\eta : \mathcal{X} \to \mathbb{R}$. A natural way is to choose such the mapping, which is an inner part of the density formula.

ii) The resulting score functions of distribution can be used in (0.1) and (0.2) with remarkable results, namely in situations when a parametric model of the data is known and data are contaminated.

2. SCORE FUNCTION OF DISTRIBUTION

DEFINITION 1. *Let $\mathcal{X} \subseteq \mathbb{R}$ be an open interval, $F \in \mathcal{P}_{\mathcal{X}}$ and $\eta : \mathcal{X} \to \mathbb{R}$ be a continuous increasing mapping. Set*

$$T(x) = -\frac{1}{f(x)} \frac{d}{dx} \left(\frac{1}{\eta'(x)} f(x) \right) \tag{0.11}$$

where f is the density of F and $\eta'(x) = d\eta(x)/dx$. Let the solution x^* to the equation

$$T(x) = 0 \tag{0.12}$$

be unique. Function

$$S(x) = \eta'(x^*)T(x) \tag{0.13}$$

will be called an $\eta-$score of distribution F.

The idea behind Definition 1 is that any $F \in \mathcal{P}_{\mathcal{X}}$ is taken as a transformed distribution with density

$$f(x) = g(\eta(x))\eta'(x), \tag{0.14}$$

where g is the density of the "prototype distribution" $G(y) = F(\eta^{-1}(y)) \in \mathcal{P}_{\mathbb{R}}$.

To obtain an unambiguous score function, we introduce at the first sight a vague concept.

DEFINITION 2. *The mathematically simplest η-score of F will be called the score function of distribution F. The corresponding $\eta : \mathcal{X} \to \mathbb{R}$ will be called the most favorable mapping.*

To clarify the definition, notice that according to formula (0.14), the most favorable mapping for distribution F is the mapping, the reciprocal derivative of which is *a componenet of the density formula* $f(x)$ so that, by (0.11),

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

As an example, the lognormal distribution with density $f(x) = \frac{1}{\sqrt{2\pi}x} e^{-\frac{1}{2}\log^2 x}$ has apparently the most favorable mapping $\eta(x) = \log x$, $\eta'(x) = 1/x$. Then, $T(x) = -xe^{\frac{1}{2}\log^2 x} \frac{d}{dx} e^{-\frac{1}{2}\log^2 x} = \log x$, and since $x^* = 1$, $S(x) = T(x)$. Some distributions may get such a form after a simple modification. An example is the exponential distribution with density $f(x) = e^{-x} = (\log x)' x e^{-x}$, so that its score function is $S(x) = T(x) = -e^x \frac{d}{dx}(xe^{-x}) = x - 1$.

To find the most favorable mapping of a distribution is often an easy task. Before discussing it, we prove a theorem showing that the η-score reduces in particular cases to the score function for an important parameter.

THEOREM 1. For any interval support $\mathcal{X}$ and any continuous, strictly increasing $\eta : \mathcal{X} \to \mathbb{R}$ there exists a class $\Pi_{\mathcal{X}}$ of distributions with parameter τ, say, such that for any $F_\tau(x) \in \Pi_{\mathcal{X}}$ the $\eta-$score equals to the score function for τ.

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

$$\tau = \eta^{-1}(\mu), \tag{0.15}$$

and with score function S_F. Set $u = \eta(x) - \eta(\tau)$. Using (0.14) 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 (0.7), $S_G(u) = -g'(u)/g(u)$). Since by (0.11)

$$T(x;\tau) = -\frac{1}{g(u)\eta'(x)} \frac{d}{dx} g(u) = -\frac{1}{g(u)\eta'(x)} \frac{dg(u)}{du} \frac{\partial u}{\partial x} = S_G(u) \tag{0.16}$$

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

$$U_\tau(x;\tau) = \eta'(x^*)T(x;\tau) = S_F(x;\tau). \tag{0.17}$$

□

By (0.15), parameter τ is the "image" of the location of the prototype. Let us call it a *transformed location parameter*. If $\mathcal{X} \neq \mathbb{R}$, the class $\Pi_{\mathcal{X}}$ consists of distributions with transformed location parameter.

Let us turn to the problem of finding the most favorable mapping for a given distribution.

i) Distributions from $\mathcal{P}_\mathbb{R}$ often have the most favorable mapping the identical mapping $\eta(x) = x$ and score functions (0.7). The score function of standard normal distribution is $S(x) = x$, that of $\mathcal{N}(\mu, \sigma)$ is $S(x) = \frac{x-\mu}{\sigma^2}$, another examples are given in Table II. However, let us consider a distribution with density

$$f(x) = \frac{1}{\sqrt{(1+x^2)}} \frac{e^{\sinh^{-1} x}}{(1 + e^{\sinh^{-1}} x)^2}. \tag{0.18}$$

It is easy to see that the most favorable mapping $\eta : \mathbb{R} \to \mathbb{R}$ here is $\eta(x) = \sinh^{-1} x$, $\eta'(x) = \frac{1}{\sqrt{(1+x^2)}}$. From (0.11) one obtains

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

and, since $\eta'(0) = 1$, score function of distribution (0.18) is $S(x) = T(x)$. Obviously, (0.18) is the density of the transformed logistic prototype.

ii) The most favorable mapping of distributions from $\mathcal{P}_{(0,\infty)}$ is often $\eta(x) = \log x$, giving so called Johnson scores studied in previous author's papers. Some parametric distributions from $\mathcal{P}_{(0,\infty)}$ and the corresponding score functions are listed in Table I.

Table I. Scalar scores of some parametric distributions from $\mathcal{P}_{(0,\infty)}$.

Distribution	$f(x)$	$T(x)$	x^*	$S(x)$	ES^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	$\frac{kcx^{c-1}}{(x^c+1)^{k+1}}$	$c\frac{kx^c-1}{x^c+1}$	$\frac{1}{k^{1/c}}$	$\frac{ck(x^c-(x^*)^c)}{k^{1/c}(x^c+1)}$	$\frac{c^2k^{1+2/c}}{k+2}$

In the upper half of the table, there are distributions from $\Pi_{(0,\infty)}$. The pivotal quantity $\frac{y-\mu}{\sigma}$ of prototype distributions transforms into

$$\frac{y-\mu}{\sigma} = \frac{\log x - \log \tau}{\sigma} = \log \left(\frac{x}{\tau}\right)^{1/\sigma}. \tag{0.19}$$

Let us point out here that parameters in denominators of ratios with variable in numerator (as in the r.h.s. of (0.19)) are frequently referred to as scale parameters. From the point of view of distributions from $\mathcal{P}_{(0,\infty)}$ as transformed distributions, the transformed location $\tau = \eta^{-1}(\mu) = e^{\mu}$ represents the typical value. By (0.19), parameter c of these distributions can be explained not as a shape, but as a reciprocal scale parameter.

In the lower part of the table there are distributions not being members of $\Pi_{(0,\infty)}$. Scalar scores of such distributions are yet unknown functions and the typical value x^* and ES^2 (see the next section) are their new descriptions.

iii) In the case $\mathcal{X} = (1,\infty)$, there are at least two competitive mappings: the Johnson mapping $\eta_1(x) = \log(x-1)$ and $\eta_2(x) = \log\log x$. There are two possible η-scores of the Pareto distribution with density

$$f(x) = cx^{-(c+1)}. \tag{0.20}$$

By (0.11), the first one is, as $\eta_1'(x) = (x-1)^{-1}$,

$$T_1(x) = -\frac{1}{f(x)}\frac{d}{dx}[(x-1)f(x)] = c - \frac{c+1}{x}$$

so that $x_1^* = \frac{c+1}{c}$ and the $\eta-$score is

$$S_1(x;c) = c(\frac{1}{x^*} - \frac{1}{x}). \tag{0.21}$$

Using the latter mapping with $\eta_2'(x) = (x\log x)^{-1}$,

$$T_2(x) = -\frac{1}{f(x)}\frac{d}{dx}[x^{-c}\log x] = c\log x - 1$$

with $x_2^* = e^{1/c}$. The corresponding η-score $S_2(x;c) = e^{-1/c}(c\log x - 1)$ is proportional to the (likelihood) score for c. Fig. 1 shows both candidates, S_1 should be preferred to be taken as a score function since bounded inference functions secure robust estimates.

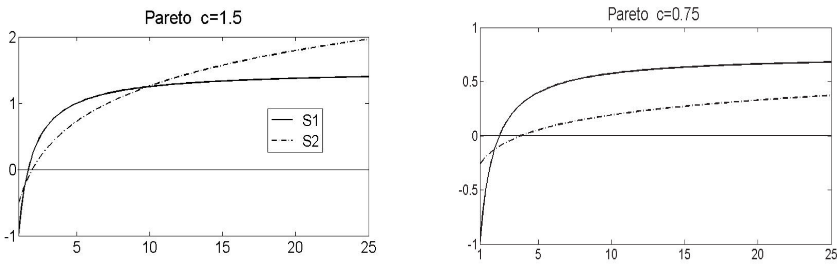

Fig. 1. Two $\eta-$scores of Pareto distribution.

The log-gamma distribution with density

$$f(x) = \frac{c^{\alpha}}{\Gamma(\alpha)}(\log x)^{\alpha-1}x^{-(\gamma+1)} \tag{0.22}$$

has apparently as the most favorable mapping the latter one, since $\eta'(x) = 1/(x\log x)$. Here $T(x) = \frac{1}{f(x)}\frac{d}{dx}[x\log x f(x)] = \gamma\log x - \alpha$, $x^* = e^{\alpha/\gamma}$ and the score function of log-gamma distribution is $S(x) = e^{-\alpha/\gamma}(\gamma\log x - \alpha)$.

iv) In the case of a finite support interval there is a great variety of different mappings $\eta(x)$: $(a,b) \to \mathbb{R}$. For distributions from $\mathcal{P}_{(0,1)}$ one can use in principle any quantile function. However, we find only few of them to be a most favorable mapping of currently used distributions. In the case of the beta distribution with density

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

the most favorable mapping is apparently the Johnson mapping (0.10) with derivative $\eta'(x) = \frac{1}{x(1-x)}$. Then,

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

$x^* = \frac{p}{p+q}$ and $S(x) = (p+q)(x/x^* - 1)$. For a distribution with density

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

the most favorable mapping is $\eta(x) = -\log(-\log x)$ since $\eta'(x) = -1/x \log x$. The score function of the distribution is then

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

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

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

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

1 $f(x) = e^{-x}/\kappa$	3 $f(x) = \frac{1}{\sqrt{2\pi}\cos^2 x} e^{-\frac{1}{2}\tan^2 x}$
2 $f(x) = e^{x}/\kappa$	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. 1. The last two distributions have a normal prototype and unbounded score functions. The latter one is the Johnson's U_B distribution transformed into $(-\pi/2, \pi/2)$.

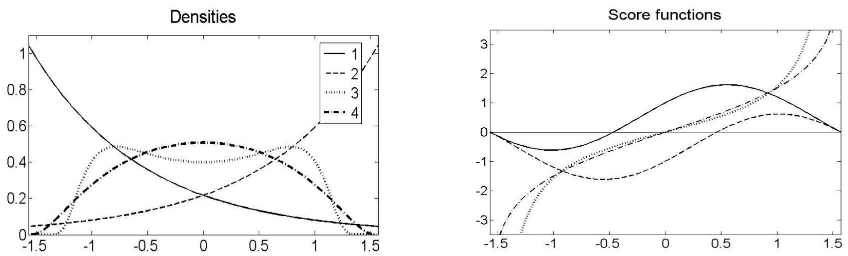

Fig 2. Densities and scalar scores of distributions from $\mathcal{P}_{(-\pi/2,\pi/2)}$.

3. SCORE MOMENTS

Moments (0.1), where S is the score function of distribution F, will be called *score moments*.

Although S can be determined from the knowledge of the density f only, a study of score moments is facilitated by the concept of a 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, and, as shown in (0.16),

$$T_F(x) = S_G(\eta(x)). \tag{0.23}$$

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

$$ES_F^k = [\eta'(x^*)]^k ES_G^k. \tag{0.24}$$

Proof. By (0.13), (0.23) and (0.14),

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

□

Let $G \in \mathcal{P}_{\mathbb{R}}$ have unimodal, twice continuously differentiable density g and score function of distribution S_G. If $g(y) = O(e^{-y})$ when $y \to \infty$, $S_G(y) = O(1)$. 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^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 (0.7), the solution y^* of equation $S_G(y) = 0$ is the mode of G. By (0.23), $T_F(x^*) = S_G(\eta(y^*))$ so that $x^* = \eta^{-1}(y^*)$ is the transformed mode of the prototype of F. This value, not very successfully named the *t-mean* or *transformation-based mean*, Fabián (2010, 2011), we take as a *typical value* of distribution F. The typical value exists and is unique for distributions with unimodal prototypes (distributions with multimodal prototypes could be perhaps viewed as mixtures). Referring to the meaning of score functions in robust statistics, $S(x)$ expresses the relative influence of $x \in \mathcal{X}$ on the typical value x^*. In the preceding section we encountered three distributions with linear score functions (normal, gamma and beta), typical value x^* of which is the mean.

ii) By (0.17), ES^2 of transformed location distributions is Fisher information for τ. Analogously, we interpret ES^2 of any continuous distribution as Fisher information for x^* or even the mean information of distribution F (Fabián, 2012). This point of view corresponds with that of Cover and Thomas (1991, pp.494) for distributions from $\mathcal{P}_{\mathbb{R}}$. Function

$$I(x) = S^2(x), \tag{0.25}$$

increases from the least informative point x^* in both directions to the end-points of the support interval. Its mean value is Fisher information for x^*. We thus interpret (0.25) as an *information function*, expressing relative information about x^* contained in x.

Fisher information for x^* of some distributions is given in Table I. Fig. 2 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.

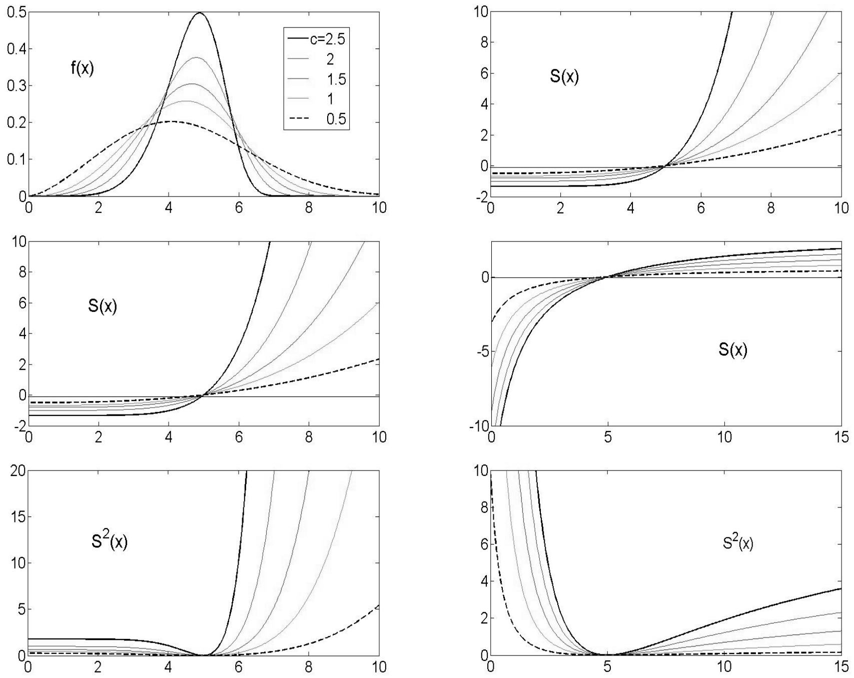

Fig. 3. Functions characterizing Weibull (left) and beta-prime distributions (right) with $x^* = 5$.

The reciprocal value of Fisher information,

$$\omega^2 = \frac{1}{ES^2}, \tag{0.26}$$

was suggested by Fabián (2007) as a measure of variability of distribution F. Let us call it according Fabián (2010) a *score variance*. 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 $\omega^2 = (x^*)^2/ES_G^2$. To indicate that ω^2 is a reasonable concept, we plotted in Fig. 4 densities of Weibull and beta-prime distributions, all with $\omega^2 = 1$. They look like having the same variability. We add that the densities in Fig. 3 differ just due to various $\omega = \tau/c$ (Weibull) and $\omega = p(p+q+1)/q^3$ (beta prime) of distributions.

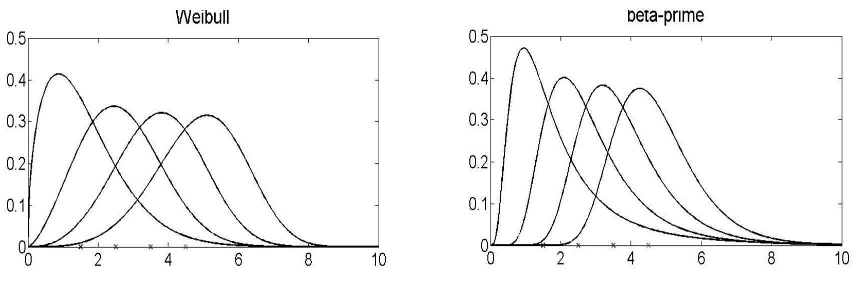

Fig. 4. Densities of distributions with equal score deviance $\omega = 1$. Typical values are marked by crosses on the x-axis.

iii) M_3 characterizes skewness. The negative/positive value of M_3 indicates a negative/positive skewness. If $M_3 = 0$, distribution can be called '*S-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 S-symmetric form, which is itself skewed.

iv) M_4 characterizes flatness of the distribution. Let us introduce an analog of Pearson's measure of kurtosis γ_2, a 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 score functions of distribution are given in Table II. It is apparent that $\tilde{\gamma}_2$ forms a logical structure reversed to kurtosis. To obtain a clearer picture of symmetric distribution, we omitted in the table the values of non-symmetric ones. γ_2 of the Cauchy distribution does not exist.

Table II. Score moments of some prototype distributions.

distribution	$f(x)$	$S(x)$	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. SCORE MOMENT METHOD AND CHARACTERISTICS OF DATA SAMPLES

The *score moment (SM) estimator* $\hat{\theta}_S \equiv \hat{\theta}_n^S$ is a solution to implicit estimating equations (0.2), where S means the score function of the assumed model F_θ. It is a special form of an M-estimator with inference function

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

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 $ES = 0$ and, at least for distributions with unimodal prototypes, $ES^k < \infty$ for $k = 2, ..., m$, the sufficient conditions are that moments $ES^k(\theta)$ are differentiable with respect to any θ_k and that a matrix $\mathbf{B}$ with elements

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

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 us have a random sample $(X_1, ..., X_n)$ from distribution F_{θ_0}, θ_0 unknown, and let the corresponding score function of distribution $S(x;\theta)$ satisfy the above conditions. The solution of equations (0.2) is consistent and asymptotically $\mathcal{N}(\theta_0, \mathbf{B}^{-1}\mathbf{A}(\mathbf{B}^{-1})')$, where $A = E\mathbf{\Psi}(x;\theta_0)\mathbf{\Psi}(x,\theta_0)'$.*

Let us present a few examples indicating usefulness of studying SM estimators. The score moment equations are, as a rule, much simpler than ML equations, and score moments are usually expressed by means of parameters and not by special functions of parameters. Moreover, it follows from our simulation experiments that SM estimates have in many cases acceptable relative efficiencies.

EXAMPLE 4.1. Estimating equations (0.2) for Weibull distribution with semi-bounded score function (Table I) are

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

$\hat{c}_S$ is 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 $\hat{\tau}_S = \frac{1}{n}\left(\sum_{i=1}^n x_i^{\hat{c}_S}\right)^{1/\hat{c}_S}$.

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 score function (Table I) even when $q \geq 2$, so that, in this case, the score moment estimate is robust even in cases of light-tailed distribution. Since

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

where $B(p,q)$ is the beta function, the estimating equations (0.2) 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 the first equation we have

$$\hat{x}_S^* = \hat{x}^* = \sum_{i=1}^n \frac{x_i}{1+x_i} \Big/ \sum_{i=1}^n \frac{1}{1+x_i} \tag{0.27}$$

and $\hat{p}_S = \hat{x}^*\hat{q}_S$, from the second one $\hat{q}_S = (\hat{x}^*/\xi(\hat{x}^*) - 1)/(\hat{x}^* - 1)$.

Relative efficiencies of the score moment (SM) estimates were in both cases tested by means of Monte Carlo simulations. Random samples of length $n = 50$ were generated from the Weibull and beta-prime distributions (in the latter case as transformed values of the beta distribution) and average efficiencies SM estimates of x^* and ω^2, $\mathrm{e}(x_S^*) = \mathrm{var}(\hat{x}_{ML}^*)/\mathrm{var}(\hat{x}_S^*)$ and $\mathrm{e}(\hat{\omega}_S^2) = \mathrm{var}(\hat{\omega}_{ML}^2)/\mathrm{var}(\hat{\omega}_S^2)$ were computed over 10 000 samples.

Values of average efficiencies in Table III indicate that the SM estimates of typical value have often a sufficient accuracy. For Weibull distribution the accuracy of estimates of variability decreases for densities having sharp narrow peak distant from zero, whereas in the case the beta-prime distribution accuracy decreases for densities with the mass concentrated near zero and very long tail.

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

EXAMPLE 4.3. By the score moment method it is even possible to estimate characteristics of distributions having the end point as a parameter. Consider the uniform distribution with support $(0,\delta)$. The maximum likelihood estimator of δ is $\hat{\delta}_{ML} = x_{(n)} = \max(x_1,...,x_n)$. According to (0.10), $\eta(x) = \log\frac{x}{\delta - x}$ and $T(x) = -\frac{d}{dx}[x(\delta - x)]$. The solution of equation $\sum_{i=1}^n T(x_i) = 0$ gives $\hat{\delta}_n = 2\bar{x}$. To exclude cases where $\hat{\delta}_n < x_{(n)}$, we set the 'adjusted' score moment solution as $\hat{\delta}_n^S = \max(x_{(n)}, 2\bar{x})$. After 10 000 experiments we obtained for $n = 5$, 10, 20 and 50 ML estimates $\hat{\delta}_n^{ML} = 0.87$, 0.91, 0.95 and 0.98, whereas $\hat{\delta}_n^S = 1$ for $n \geq 5$ with accuracy to three decimal points.

Other examples of of using score moment estimators can be found in Fabián (2010) and Stehlík et al. (2011).

In Section 3, reasons for taking x^* as a typical value of distribution $F \in \mathcal{P}_{\mathcal{X}}$ were clarified. The *sample typical value* $\hat{x}^* \equiv \hat{x}^*_n = x^*(\hat{\theta}_n)$, where $\hat{\theta}_n$ is a 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 (0.2) can be often written as

$$\sum_{i=1}^{n} S(x_i; x^*) = 0. \tag{0.28}$$

If the solution $\hat{x}^*_S$ of (0.28) is expressed as an explicit function of sample observations, we call it a *score average.* For example, it follows from (0.28) for Gumbel distribution with density $f(x) = e^{x-\mu}e^{-e^{x-\mu}}$ and score function $S(x) = e^{x-\mu} - 1$ that the score average is

$$\hat{x}^*_S = \hat{\mu} = \log\left(\frac{1}{n}\sum_{i=1}^{n} e^{x_i}\right), \tag{0.29}$$

which equals to the ML estimate of the location parameter. The prototype of the gamma distribution (to give an example of a distribution without location parameter) has density $g(x) = \frac{\gamma^\alpha}{\Gamma(\alpha)}e^{\alpha x}e^{-\gamma e^x}$ and score function $S(x) = \gamma e^x - \alpha$ so that $x^*_S = \log(\alpha/\gamma)$ and the score average is, incidentally, given by (0.29) as well. Score average of the Laplace distribution (Tab. II) is the median.

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

i) If c is 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 $c = 1$, score average of the lognormal distribution is the geometric mean, of Weibull the mean and of Fréchet the harmonic mean.

ii) $\hat{x}^*$ of the gamma distribution is the mean, $\hat{x}^*$ of inverted gamma the harmonic mean and score average of the sample from the beta-prime distribution is given by (0.27).

Using as the score function of the Pareto distribution with density (0.20) and bounded $\eta-$ score (0.21), the estimating equation is

$$\sum_{i=1}^{n}(1/x^* - 1/x_i) = 0$$

from which it follows that the typical value of the sample from Pareto distribution is the harmonic mean.

The asymptotic variance of score averages is (Fabián, 2009)

$$\sigma^2_{as} = ES^2/[\frac{\partial}{\partial x^*}S(x; x^*)]^2.$$

Similarly, the estimate of the score variance, the *sample score variance*, is given by $\hat{\omega}^2 \equiv \hat{\omega}^2_n = \omega^2(\hat{\theta}_n)$ or as a finite version of (0.26), that is,

$$\hat{\omega}^2 = \frac{n}{\sum_{i=1}^{n} S^2(x_i; \hat{\theta}_n)}. \tag{0.30}$$

For a few distributions, $\hat{\omega}^2$ is expressed as explicit function of sample observations, too. For example, $\hat{\omega}^2$ of the gamma distribution equals the variance and $\hat{\omega}^2$ of the inverted gamma distribution is $\hat{\omega}^2 = \bar{x}^2_H(\bar{x}^2_H\frac{1}{n}\sum 1/x_i^2 - 1)$, where $\bar{x}_H$ means the harmonic mean.

5. ESTIMATION IN THE PRESENCE OF OUTLYING VALUES

The score moment estimates of parameters of families with bounded score functions have an attractive property: the estimates are insensitive to incidental large values from other contaminating source.

If the data from distributions with unbounded or semi-bounded score functions are contaminated by another source, it is necessary to modify score functions using some of the procedures suggested by robust statistics (trimming the data or tapering the inference function). Since S is a unique function, such modification is in principle easy to apply.

To obtain robust score estimators for distributions with bounded or semi-bounded score functions, we use Huber's famous suggestion (1964). Its modification by Huber and Ronchetti (2005) consists in using as an inference function of distributions from $\mathcal{P}_{\mathbb{R}}$ function

$$\psi(x) = \begin{cases} S(x-\mu) & \text{if} \quad |x-\mu| \leq v \\ b\,\mathrm{sgn}(x-\mu) & \text{if} \quad |x-\mu| > v, \end{cases}$$

where v is some bound and b a tuning value. We suggest in a general case to use the tapered score function of distribution. According to Beran and Schell (2010), let us call this procedure "huberizing".

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

$$\psi_k(x;\theta) = [S^k(x;\theta)]_u^v - E_\theta\{[S^k(x;\theta)]_u^v\}, \tag{0.31}$$

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

$$\sum_{i=1}^{n} \psi_k(x_i;\theta) = 0, \qquad k = 1, ..., m \tag{0.32}$$

will be called a huberized score moment estimator.

THEOREM 5. *Let $T_n \equiv \hat{\theta}_n^H \to_p \theta_0$, $E\psi_k(x;\theta)$ be differentiable at θ_0, and ψ_k be continuously differentiable. Let matrix $\mathbf{B}$ of derivatives with elements $\dot{\psi}_{jk} = \partial\psi_k/\partial\theta_j|_{\theta=\theta_0}$ be nonsingular and $|\dot{\psi}_{jk}(x;\theta)| \leq K(x)$ for $j,k = 1, ..., m$ where $EK(x) < \infty$. Let $E|\psi_k(x;\theta_0)|^2$ 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).

Let us further set

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

$I_{ku}(\theta) = S^k(u;\theta)F_\theta(u)$ and $I_k^v = S^k(v;\theta)(1-F_\theta(v))$. Equations (0.32) can be then written in the form

$$\frac{1}{n}\sum_{i=1}^{n} S^k(\tilde{x}_i;\theta) - ES^k(\theta) = -\{I_{k|au}(\theta) + I_{k|vb}(\theta)\} + \{I_{ku}(\theta) + I_k^v(\theta)\}, \tag{0.33}$$

where

$$\tilde{x}_i = \begin{cases} r_1 & \text{if} \quad x_i < u \\ x_i & \text{if} \quad u \leq x_i \leq v \\ r_2 & \text{if} \quad x_i > v, \end{cases} \tag{0.34}$$

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

$$\theta_{in} = \theta(\hat{x}_0^*, \hat{\omega}_0). \tag{0.35}$$

For two-parameter distributions, relation (0.35) is often one-to-one.

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

EXAMPLE 5.1. Normal distribution $\mathcal{N}(\mu,\sigma)$, $x^* = \mu$ and $\omega = \sigma$. Set $u = \mu_0 - b\sigma_0, v = \mu_0 + b\sigma_0$. The huberized score function a function

$$\psi(x) = \begin{cases} -r & \text{if} \quad x < u \\ \frac{x-\mu}{\sigma} & \text{if} \quad u \leq x \leq v \\ r & \text{if} \quad x > r. \end{cases} \tag{0.36}$$

Since $E\psi = 0$ and, by (0.33),

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

it follows from (0.31) that $\hat{\mu}_H = \frac{1}{n}\sum_{i=1}^n \tilde{x}_i$ and

$$\hat{\sigma}_H^2 = \frac{\frac{1}{n}\sum_{i=1}^n (\tilde{x}_i - \hat{\mu}_H)^2}{1 - \sqrt{\frac{2}{\pi}} b e^{-\frac{1}{2}r^2} + (r^2-1)(1 - \mathrm{erf}(r/\sqrt{2}))}. \tag{0.37}$$

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, thus indicating contamination.

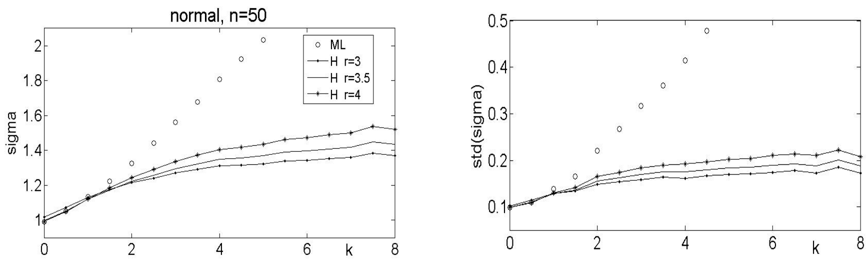

Fig. 5. Estimates of σ of contaminated $\mathcal{N}(0,1)$ under increasing contamination, left: $\hat{\theta}_H$, right: $\mathrm{std}(\hat{\theta}_H)$.

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

$$\psi(x;\tau,c) = \begin{cases} (x/\tau)^c - 1 & \text{if } \; x \le v \\ r & \text{if } \; x > v \end{cases} \tag{0.38}$$

where $r = (v/\tau)^c - 1$. The first and third members of r.h.s. of (0.33) 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 (0.33) are

$$\begin{aligned} \frac{1}{n}\sum_{i=1}^n ((\tilde{x}_i/\tau)^c - 1) &= -\lambda(1) + r\lambda(0) \qquad (0.39) \\ \frac{1}{n}\sum_{i=1}^n ((\tilde{x}_i/\tau)^c - 1)^2 - 1 &= -\lambda(2) + r^2\lambda(0). \end{aligned}$$

Set now

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

where $\tau_0 = \mathrm{med}(x)$ and $\omega_0 = \mathrm{MADN}(x) = \mathrm{MAD}(x)/0.675$ and compute members on the r.h.s. by using them. We obtain $w = (1+k/c_0)^{c_0}$, $r = w - 1$, $\lambda(0) = e^{-w}$, $\lambda(1) = we^{-w}$ and $\lambda(2) = (1+w^2)e^{-w}$ so that

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

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.

We refer to the density of any two-parameter distribution as a function of $x^* = \tau$ and $\omega = \tau/c$. In simulation experiments, the contaminated distribution was

$$f_c(x^*, \omega) = (1-\epsilon)f(x^*, \omega) + \epsilon f(x^* + k, \omega) \tag{0.40}$$

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 with increasing k increasing linearly, the huberized estimates stabilize at certain level, which is, however, higher than the true value, indicating thus contamination.

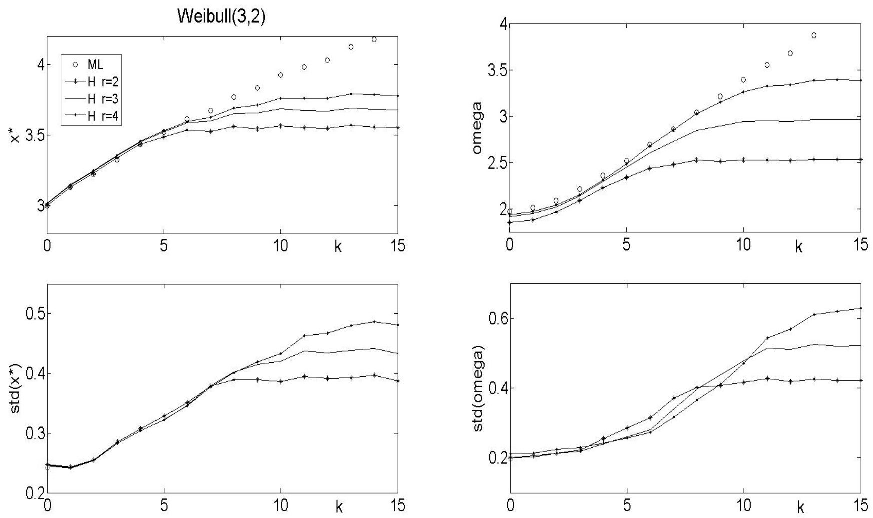

Fig. 6. Average estimate of typical value and score deviation of contaminated Weibull distribution and their standard deviations.

Average efficiencies of huberized moment estimates for various combinations of τ and ω from simulation experiments 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 MADN) 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 necessary to use other input values, perhaps the mean and variance. 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. Gamma distribution (Table I) with $x^* = \alpha/\gamma$ and $\omega = \alpha/\gamma^2$ has score function $S(x) = \frac{x-x^*}{\omega^2}$. By setting $u = 0$ and $v = x_0^* + k\omega_0$, the huberized score function 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 the simplified equations

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

where $\tilde{x}_i$ are given by (0.34). Surprisingly, even biased solutions of these simple equations are reasonably efficient (Fig. 7).

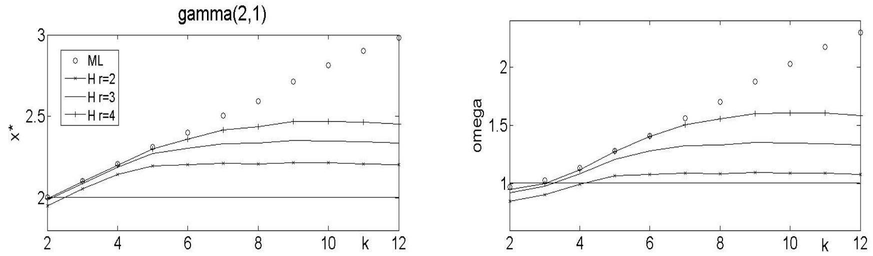

Fig. 7. Average estimate of typical value and score deviation of contaminated gamma distribution.

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 is documented by the behavior of standard deviations.

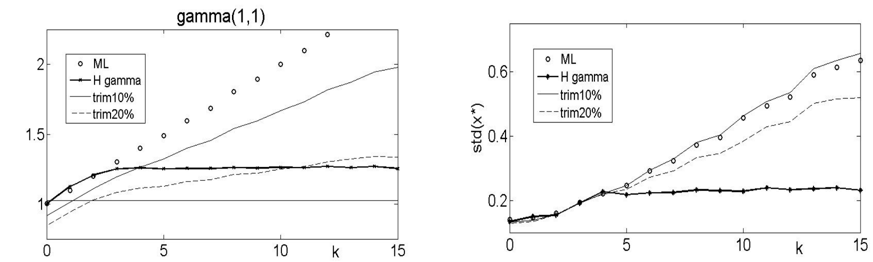

Fig. 8. Robust estimates under increasing contamination.

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.

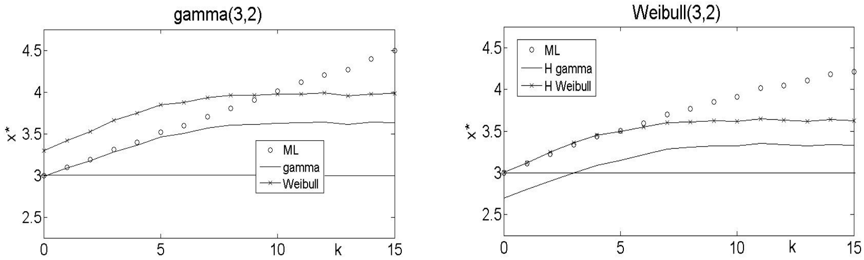

Fig. 9. Using the proper and improper estimator. Data are generated from a distribution stated in the headline.

6. COMPARISON OF RESULTS FOR DIFFERENT MODELS

Based on simulation experiments, one can expect that the area of implementation of estimators based on score functions of distributions and their huberized versions is the situation at which the model is known (can be assumed on the base of some previous or theoretical knowledge) and data are highly contaminated.

In the light of the previous account, an estimate $\hat{\theta}_n$ of θ need not be a final result of the inference procedure. The more interesting values are the sample typical value and sample variance

$$x^* = x^*(\hat{\theta}_n), \qquad \hat{\omega}^2 = \omega^2(\hat{\theta}_n)$$

(and, perhaps, the higher score moments), which make possible to compare results of estimation under various assumptions of differently parametrized models. To utilize the developed theory, such a comparison of models F and G can be based on the score divergence, suggested (in a slightly different form) by Fabián and Vajda (2003) as

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

where S_F and S_G are the corresponding score functions.

EXAMPLE 6.1. 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 (uncontaminated) 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 when data originate from the gamma.

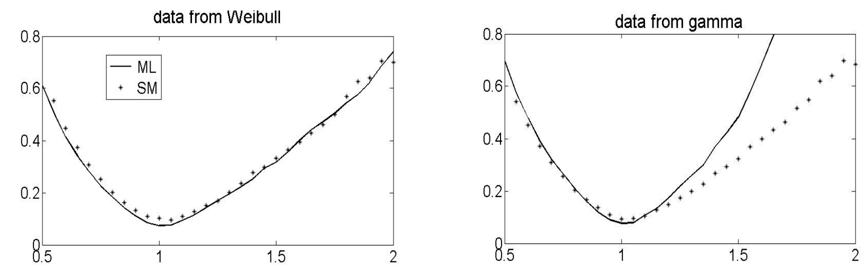

Fig. 10. Observed score divergence of the Weibull and gamma distributions as functions of ω.

7. CONCLUSIONS

By Definition 1 and 2, the concept of the score function of distribution, studied in the author's previous works (under names "core function", "Johnson score" or "scalar score") is generalized and clarified. In contrast to the vector-valued score function, score function of distribution is a scalar-valued function even if the parameter space is a vector one. For any $\mathcal{X} \subseteq \mathbb{R}$ exists a particular class of distributions $\Pi_{\mathcal{X}} \in \mathcal{P}_{\mathcal{X}}$, for which both the score function of distribution and (likelihood) score function with respect to certain (transformed location) parameter are identical, which, we think, justifies taking score function of distribution as a significant function of any F. An attractive feature of function (0.13) is that it does not contain either the normalizing constant or terms arising from its differentiation according parameters, so that it is given in many cases by a simple expression. The most important property of all types of score functions is their behavior at the end-points of the support interval (they can be unbounded, bounded, semi-bounded and even "redescending".

The typical value of distribution $F \in \mathcal{P}_{\mathcal{X}}$, the zero of the score function of distribution, can be used instead of (or in addition to) the mean value. Score function of distribution appears to be the *score function for the typical value* x^* of the distribution. On the one hand, the score function of distribution is a generalization of the score function for an important quantity which may not be a parameter of the distribution. On the other hand, it represents an important simplification, since it is a scalar-valued function. Score moments appear to be new relevant numerical characteristics of regular distributions, they exist even for heavy-tailed distributions and have a reasonable sense. Particularly, ES^2 is the Fisher information about x^* and its reciprocal value measures the variability of the distribution instead of (in addition to) the variance.

Moment estimators given by (0.2) with the score function of distribution are new variants of estimators of classical statistics. They are not efficient, but they are robust if the score function of the model is bounded. In other cases, the score function can be easily modified to be bounded by methods of robust statistics. Sample characteristics based on the typical value and score variance of the model can be useful in description of data samples in situations when a model is known and data are strongly contaminated.

ACKNOWLEDGEMENT. This work is done with institutional support RVO:67985807.

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.