



národní
úložiště
šedé
literatury

Scalar-Valued Score Functions and their use in Parametric Estimation

Fabián, Zdeněk
2022

Dostupný z <http://www.nusl.cz/ntk/nusl-512291>

Dílo je chráněno podle autorského zákona č. 121/2000 Sb.

Tento dokument byl stažen z Národního úložiště šedé literatury (NUŠL).

Datum stažení: 09.04.2024

Další dokumenty můžete najít prostřednictvím vyhledávacího rozhraní [nusl.cz](http://www.nusl.cz).



Institute of Computer Science
The Czech Academy of Sciences

Scalar-valued score functions and their use in parametric estimation

Zdeněk Fabián

Technical report No. V-1291

February 2023



Institute of Computer Science
The Czech Academy of Sciences

Scalar-valued score functions and their use in parametric estimation

Zdeněk Fabián¹

Technical report No. V-1291

February 2023

Abstract:

In the paper we present new paradigm in probabilistic and statistical reasoning, based on a finding that to any continuous univariate random variable X can be assigned a scalar-valued score random variable, expressing a relative influence of items generated from the distribution of X with respect to its typical value. The approach leads to a new description of standard distributions, finite typical values and finite measures of variability even of heavy-tailed distributions. The methodology is generalized for parametric families and used for solutions of some estimation problems.

Keywords:

score random variable, score mean, score variance, score distance, point estimation, generalized central limit theorem

¹Institute of Computer Science, The Czech Academy of Sciences, Pod Vodárenskou věží 2, 182 07 Prague 8, Czech Republic, E-mail: zdenek@cs.cas.cz.

1 Introduction

Probability theory of continuous random variables is deemed to be a relatively closed discipline with low probability of occurrence of new ideas of basic significance. However, the theory exhibits some curious facts.

By \mathcal{R} is denoted real line. The term random variable means in this text a univariate continuous random variable with a finite or infinite open interval support $\mathcal{X} = (a, b) \subseteq \mathcal{R}$ (“on \mathcal{X} ”), fully described by a couple (\mathcal{X}, F) , where $F(x) = \mathcal{P}(X < x)$ is the distribution function, the representative of a probability measure \mathcal{P} on \mathcal{X} .

Random variable X distributed according to F with density $f(x)$ is used to be characterized by moments $EX^k = \int_{\mathcal{X}} x^k f(x) dx$, especially by the mean EX as a measure of its central tendency and variance $Var X = E(X - EX)^2$ as a measure of its variability. Both first two moments are useful for description of distributions “near normal” and for studies of properties of asymptotically normal estimates of parameters of parametric distributions from large data samples, but not for a description of distributions themselves. They may not exist in cases of heavy-tailed distributions with densities going to zero too slowly. A well-known example is the symmetric Cauchy distribution with clearly expressed center, the mean of which does not exist. Fig. 1 shows two histograms of 250 items generated from the beta-prime distribution $BP(1, 2)$ (see Section 6) with infinite variance and generated values less than 15. Variability of the distribution should certainly be described by a more suitable measure.

All of the probability densities have integral over the sample space equal to one. What is the reason for which distributions are distinguished as “regular” and those with non-existing basic characteristics ?

The moments testify on an unsuccessful historical transition from discrete to continuous random variables, leading to introduction of the Euclidean geometry in the sample space apart from a large variety of probability measures on it.

2 Parametric approach

To get reliable information from observed data, mathematical statistics does not rely on sample values of moments and uses parametric approach.

The simplest parametric model is a location family $G(x - \mu)$ on \mathcal{R} . Given a

random sample $\mathbb{X}_n = (x_1, \dots, x_n)$, a realization of random variables X_1, \dots, X_n iid according to some $G_{\mu_0} \in G_\mu$, one obtains the best estimate of μ_0 from equation

$$\frac{1}{n} \sum_{i=1}^n U(x_i - \hat{\mu}_n) = 0, \quad (1)$$

where

$$U(x - \mu) = \frac{\partial}{\partial \mu} \log g(x - \mu) \quad (2)$$

is the generic likelihood score for μ , describing relative influence of an item x_i on the value of the estimate. To speak in terms of robust statistics, (2) is the best inference function for location distributions.

The situation is more complicated if $\boldsymbol{\theta} = (\theta_1, \dots, \theta_m)$ is a vector parameter. Hundred years ago, R. A. Fisher (1922) generalized (2) by introducing vector-valued score functions $U(y; \boldsymbol{\theta}) = (U_1, \dots, U_m)$ with components

$$U_j(x; \boldsymbol{\theta}) = \frac{\partial}{\partial \theta_j} \log f(x; \boldsymbol{\theta}),$$

now known as Fisher scores. It led to the maximum likelihood (ML) estimating equations

$$\frac{1}{n} \sum_{i=1}^n U_j(x_i; \boldsymbol{\theta}) = 0, \quad j = 1, \dots, m, \quad (3)$$

providing estimates with minimal asymptotic variance and hence considered as the best parametric estimates.

However, best only for data without outliers. Fisher scores are often unbounded. To suppress influence of outliers, Huber (1964) generalized equation (1) into

$$\frac{1}{n} \sum_{i=1}^n \psi(x_i - \hat{\mu}_n) = 0, \quad (4)$$

where $\psi(x)$ is a bounded inference function. Asymptotic variance of $\hat{\mu}_n$ from (4) is

$$\sigma_\psi^2 = \frac{E\psi^2}{n[E\psi']^2}. \quad (5)$$

Nowadays, a large number of bounded, scalar-valued, easily manipulated inference functions is known, more or less useful for practical estimation,

but often without a tight relation to the assumed model. To “robustify” equations (3) (by trimming or Huberizing unbounded Fisher scores) is too complicated even in cases of two-parameter distributions.

Recently, Fabián (2001) recognized a new, yet undetected way how to generalize function (2) and equation (1) for distributions with vector parameter. The aim of the present paper is to explain basic steps of the approach, to present main theoretical results, and to outline possibilities of its use in solution of statistical tasks.

The plan of the paper is the following. The score function of standard distributions on the entire \mathcal{R} , our starting point, is described in Section 3. The concept is generalized via specific transformations for distributions on arbitrary open interval support in Section 4. Here are also discussed new characteristics of distributions based on score random variables. The Central Limit Theorem for score random variables is proven in Section 5. Generalizations for parametric families and score moment estimates of parameters, outlined in Section 6 are followed by some examples (Section 7) and by a short discussion about possible use of scalar-valued score functions in other statistical tasks.

3 Score function of distributions on \mathcal{R}

Let Y be random variable with distribution G on \mathcal{R} and continuously differentiable density $g(y)$. Our starting point is the identity

$$\frac{\partial}{\partial \mu} \log g(y - \mu) = -\frac{1}{g(y - \mu)} \frac{d}{dy} g(y - \mu) \equiv S_G(y - \mu). \quad (6)$$

The generic log-likelihood score for location at $\mu = 0$ equals to the score function of a standard distribution G with *score function*

$$S_G(y) = -\frac{g'(y)}{g(y)}, \quad (7)$$

evidently characterizing the distribution itself.

Often, Fisher vector-valued function U or Fisher scores and even ψ -functions in robust statistics are sometimes called score functions. We use the term in the narrow sense of the word: score function of G on \mathcal{R} is the function (7).

To any Y can be assigned a *score random variable* $S_G(Y)$ having interesting properties:

i/ According to behavior of $S_G(y)$ when y is approaching to $\pm\infty$, distributions of Y on \mathcal{R} can be classified into 6 substantially different types:

UE	Unbounded Exponentially increasing
UP	Unbounded Polynomially increasing
BU	Bounded at $-\infty$ and Unbounded at ∞
UB	Unbounded at $-\infty$ and Bounded at ∞
BB	Bounded at $\pm\infty$
BR	Bounded Redescending

Basic types of score functions of standard distributions on \mathcal{R} and corresponding densities are given in Table 1.

ii/ Under mild regularity conditions, moments

$$ES_G^k = ES_G^k(Y) = \int_{-\infty}^{\infty} S_G^k(y)g(y) dy \quad (8)$$

are finite (Fabián, 2001), since the densities of light-tailed distributions are quickly decreasing to zero and score functions of heavy-tailed distributions are bounded.

iii/ Considering unimodal distributions only (the multimodal are usually taken as mixtures of unimodal ones), as a typical value of Y can be considered the modus y^* given by

$$S_G(y^*) = 0. \quad (9)$$

iv/ It follows from (6) that ES_G^2 equals the Fisher information for location. ES_G^2 of standard distributions (without parameters) may be taken as Fisher information for y^* or even of distribution G itself, cf. Cover and Thomas (2006). Moreover, in agreement with the sense of the Fisher information in classical statistics, as a measure of variability of G can be chosen $Var_S Y = 1/ES_G^2$.

v/ Function

$$w_G(y) = S'_G(y) = \frac{dS_G(y)}{dy}, \quad (10)$$

can be interpreted as the weight function of G , describing relative weight of an item $y \in \mathcal{X}$ with respect to the mode y^* .

Table 1 shows the simplest strictly increasing smooth function of each from 6 types, chosen as score functions, corresponding densities and distributions from (7) and their weight functions. They can be considered as basic types of standard distributions on \mathcal{R} .

Surprisingly, score random variables in the narrow sense are not discussed in probability or statistical textbooks. The reason is undoubtedly the fact, that score functions (7) have peculiar behavior in cases of distributions with a partial support $\mathcal{X} \neq \mathcal{R}$. For example, the “score function” (7) of the exponential distribution equals one and that of uniform distribution equals zero.

4 Score functions of distributions on $\mathcal{X} \neq \mathcal{R}$

This serious obstacle for using score functions has been surmounted and the relevant score functions of distributions with partial support has been found by Fabián (2001). The procedure was based on the idea that continuous random variables on $\mathcal{X} \neq \mathcal{R}$ are transformed random variables from \mathcal{R} .

Let random variable Y on the entire \mathcal{R} has standard distribution G with density $g(y)$. Due to a simple account, $g(y)$ is supposed unimodal and twice continuously differentiable. Let us call G a *prototype* of a transformed distribution $F(x) = G(\eta(x))$ on \mathcal{X} , where $\eta : \mathcal{X} \rightarrow \mathcal{R}$ be a strictly increasing smooth mapping. A distribution F of the transformed random variable $X = \eta^{-1}(Y)$ has on \mathcal{X} density

$$f(x) = g(\eta(x))\eta'(x), \quad (11)$$

where $\eta'(x) = d\eta(x)/dx$ is the Jacobian of the transformation.

As a transformation-based scalar-valued score $S_F(x)$ of F (“core function” in Fabián, 2001, “t-score” in his recent works) were identified the transformed score function (7) of prototype G ,

$$S_F(x) = S_G(\eta(x)). \quad (12)$$

By the use of (7) and (11) one obtains formula

$$S_F(x) = -\frac{1}{f(x)} \frac{d}{dx} \left[\frac{1}{\eta'(x)} f(x) \right], \quad (13)$$

eliminating the explicit dependence on G . Indeed,

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

$S_F(x)$ in (13) still depends on the mapping used. Various possible mappings $\eta : \mathcal{R} \rightarrow \mathcal{R}$ are clearly apparent from concrete formulas (11). It holds true even for many transformed distributions on $\mathcal{X} \neq \mathcal{R}$. For example, a distribution on $\mathcal{X} = (\pi/2, \pi/2)$ and density

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

has clearly normal prototype and $\eta(x) = \tan x$. If neither g nor $\eta'(x)$ are “visible” from (11), it is to use simple mappings by Johnson (1949), the part of the most of distributions in current use,

$$\eta(x) = \begin{cases} \log(x - a) & \text{when } \mathcal{X} = (a, \infty) \\ \log \frac{x - a}{(b - x)} & \text{when } \mathcal{X} = (a, b). \end{cases} \quad (14)$$

Consider the exponential distribution with density $f(x) = e^{-x}$. Since $e^{-x} = xe^{-x} \frac{1}{x}$, obviously $\eta(x) = \log x$ and $S_F(x) = x - 1$. By (13) and (14), the score function of the uniform distribution on $(0, 1)$ is $S_F(x) = 2x - 1$. A detailed discussion of this point can be found in Fabián (2016).

Other characteristics of transformed distributions are derived from those of their prototypes.

Definition 1 *Let F be a distribution on \mathcal{X} with density $f(x)$. Function $S_F(x)$ given by (13) is called the score function of distribution with a shorthand *sfd*. Any x^* such that $S_F(x^*) = 0$ is called the score mean.*

The score mean of transformed F with unimodal prototype G is a unique quantity. By (12) and (9), $0 = S_F(x^*) = S_G(\eta(x^*))$ and $x^* = \eta^{-1}(y^*)$ is the transformed mode of the prototype, which we take as the typical value of F , Fabián (2021).

By (12) and (11), score moments are

$$ES_F^k = \int_{\mathcal{X}} S_G^k(\eta(x)) g(\eta(x)) dx = ES_G^2,$$

particularly $ES_F = 0$.

Definition 2 *The score variance of F is defined by*

$$\omega_F^2 \equiv \text{Var}_S X = \frac{ES_F^2}{[S'_F(x^*)]^2}, \quad (15)$$

where $S'_F(x^*) = dS_F(x)/dx|_{x=x^*}$. Reasons for introduction of this measure of variability of distributions are explained in Fabián (2022).

Transformed weight function of F is, obviously, $w_F(x) = w_G(\eta(x))\eta'(x)$, Fabián (2021).

By means of newly introduced functions can be expressed the distance generated in the sample space by probability measure represented by F as

$$d_F(x_1, x_2) = |\delta_F(x_1, x_2)| \quad (16)$$

where

$$\delta_F(x_1, x_2) = \frac{S_F(x_2) - S_F(x_1)}{S'_F(x^*)} = \frac{1}{w_F(x^*)} \int_{x_1}^{x_2} w_F(x) f(x) dx. \quad (17)$$

Table 2 shows densities, t-scores and weight functions of transformed prototypes from Table 1 by using $\eta(x) = \log x$, which is a basic set of standard distributions on \mathcal{R}^+ . Transformed distribution is of the same type as its prototype, but its right tail is heavier. We omitted in Table 2 the log-Cauchy distribution since the Cauchy one is considered to be an extreme case of heavy-tailed behavior.

5 Central limit theorem for score random variables

A generalization of the central limit theorem for score random variables has been given by Fabián (2021).

Theorem 1 *Let X_1, \dots, X_n be random variables on \mathcal{X} iid according to F with scalar-valued score S_F . Random variables $S_F(X_1), \dots, S_F(X_n)$ are iid as well. Set*

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

For $n \rightarrow \infty$

$$\sqrt{n}\bar{S}_F \xrightarrow{\mathcal{D}} \mathcal{N}(0, ES_F^2). \quad (18)$$

Proof. Since $ES_F = 0$ and ES_F^2 is finite, the assertion follows directly from the Lindeberg-Lévy central limit theorem. \square

Given random sample $\mathbb{X}_n = (x_1, \dots, x_n)$ according to F , Theorem 1 can be used for simultaneous estimation of the score mean and score variance.

Theorem 2 *Let F be a standard distribution with strictly increasing differentiable scalar-valued score $S_F(x)$ and score mean x^* . Set*

$$\bar{x}_S = \frac{1}{n} \sum_{i=1}^n S_F(x_i) \quad \text{and} \quad \hat{x}_n^* = S_F^{-1}(\bar{x}_S)$$

If $n \rightarrow \infty$,

$$\sqrt{n}(\hat{x}_n^* - x^*) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \omega_F^2).$$

Proof. As $S'_F(x^*) \neq 0$ and for $n \rightarrow \infty$ $S_F(\hat{x}_n^*) \xrightarrow{\mathcal{D}} \mathcal{N}(S_F(x^*), ES_F^2/n)$,

$$S_F^{-1}(S_F(\hat{x}_n^*)) \xrightarrow{\mathcal{D}} \mathcal{N}(x^*, [S'_F(x^*)]^{-2} ES_F^2/n)$$

according to the delta method (cf. Serfling, 1980, pp. 118). By (15), $ES_F^2/[S'_F(x^*)]^2 = \omega_F^2$. \square

Sample score means of some standard distributions are given in Table 3.

By Theorem 2, the variance of the sample score mean multiplied by n is the estimate of the score variance of distribution F . This phenomenon, known as “happy coincidence” in the case of the normal distribution, occurs in terms score mean and score variance in any case of standard distributions. (This finding is, of course, of no practical use, but indicates a natural direction in which parametric estimation can move on).

6 Scalar-valued scores of distributions with vector parameter

Returning back to identity (6), that is to

$$\frac{\partial}{\partial \mu} \log g(y - \mu) = S_G(y - \mu),$$

Fabián (2001) suggested to generalize score function (7) for vector $\boldsymbol{\theta}$ by using formula (13) and $\eta(x)$ discussed in Section 3 as follows:

Definition 3 *Let $F(x; \boldsymbol{\theta})$ be a continuous parametric distribution on arbitrary open interval $\mathcal{X} \subseteq \mathcal{R}$ and density $f(x; \boldsymbol{\theta})$. Its scalar-valued score function (sfd: score function of distribution) is*

$$S_F(x; \boldsymbol{\theta}) = -\frac{1}{f(x; \boldsymbol{\theta})} \frac{d}{dx} \left[\frac{1}{\eta'(x)} f(x; \boldsymbol{\theta}) \right]. \quad (19)$$

Unlike the classical Fisher approach, one obtains, due to differentiation with respect to the variable, a parametric scalar-valued score function and characteristics described in the last section as functions of parameters.

The score mean and score variance of distributions with linear scalar-valued scores (normal, beta, gamma) is the mean and variance. Characteristics of three distributions on \mathcal{R}^+ discussed later we present in Table 3.

From the point of view of the score mean, there are two classes of distributions: with structural parameter, that is with location μ or transformed location $\tau = \eta^{-1}(\mu)$ (those on \mathcal{R}^+ are called by Marshall and Olkin (2007) the log-location distributions), and without such a parameter. Score mean x^* of distributions from the first class is the structural parameter and their estimates are ML estimates (Fabián, 2016, Theorem 1).

Example 4.1 Prototypes of the mutually reciprocal light-tailed Weibull and heavy-tailed Fréchet have a location and scale structure transformed to \mathcal{R}^+ by

$$\frac{y - \mu}{\sigma} \rightarrow \frac{\log x - \log \tau}{\sigma} = \left(\frac{x}{\tau}\right)^c, \quad \tau, c \in \mathcal{R}^+$$

so that transformed location τ is the structural and $1/c$ the scale parameter. Variance of the Fréchet one does not exist if $c \leq 2$. The left panel of Fig. 2 shows three at a first glance similar Fréchet densities with very different variances. The score variances of plotted densities are 0.17, 0.21 and 0.25, respectively. In the right panel are compared courses of $VarX = \sigma^2$ and $Var_S X = \omega_F^2$ with respect to increasing c , that is, to decreasing scale. Score variance is finite inside the parameter space.

Example 4.2. Heavy-tailed beta-prime $BP(p, q)$ (Johnson et al., 1995) is an example of a distribution without a structural parameter. The mean and variance of $EX = p/(q - 1)$ and $VarX = p(p + q - 1)/[(q - 1)^2(q - 2)]$ do not exist if $q < 1$ and $q < 2$, respectively. Some densities and sfds of $BP(p, p)$ are plotted in Fig. 3. Note that histograms on Fig. 1 correspond to the first one.

7 Estimates and distance in the sample space

Although the inference function (19) is scalar-valued, the true value of vector θ can be estimated by the use of the score moment method. By Fabián (2001, 2010, 2016) and Stehlík et al. (2010), given a random sample \mathbb{X}_n from

$F_\theta, \theta \in \Theta_m$, the score moment estimate $\hat{\theta}_n$ of “true” θ_0 is the solution of the system of equations

$$\frac{1}{n} \sum_{i=1}^n S_F^k(X_i; \theta) = E S_F^k(\theta), \quad k = 1, \dots, m. \quad (20)$$

Score moment (SM) estimates are the M -estimates so that they are consistent and asymptotically normal under well-known conditions, cf. Huber and Ronchetti (2009). SM estimates are sensitive to outliers on the light tail side and not sensitive on the heavy tail side of the considered family. Since in equations (20) occur powers of a unique function, in cases of heavy-tailed distributions are SM estimates robust for all components of the parametric vector. Scalar-valued scores of light-tailed distributions can be relatively easily “robustified”.

Having an estimate $\hat{\theta}_n$, sample \mathbb{X}_n is characterized by the sample score mean $\hat{x}_n^* = x^*(\hat{\theta}_n)$ as a *typical value of the data* and the square root of the sample score variance, $(\hat{\omega}_F)_n = \omega_F(\hat{\theta}_n)$ as a *characteristic radius* (strength) of the data, both acquiring finite values.

Particularly, a typical value \hat{x}_n^* of a sample \mathbb{X}_n taken from distribution $F(x; x^*)$ with explicitly expressed score mean, estimated from the first score moment equation

$$\frac{1}{n} \sum_{i=1}^n S_F(x_i; \hat{x}^*) = 0, \quad (21)$$

is identical with the sample score mean estimated by means of the generalized CLT (18).

The sample score mean of distributions with linear scalar-valued scores (normal, beta, gamma) is the mean.

Example 6.1. Average estimates of the score mean \hat{x}_n^* , its standard deviations (std), and estimates of the score variance ω_F^2 from 20000 samples of lengths $n = 500$ and $n = 50$, respectively, generated from (heavy-tailed) $BP(1, 0.5)$ and $BP(1, 1)$, are presented in Table 5. By ML are denoted the maximum likelihood and by SM the score moment estimates. STD estimates are based on Theorem 2: \hat{x}_n^* determined from the first SM equation

$$\sum_{i=1}^n \frac{x_i - x^*}{x_i + 1} = 0 \quad (22)$$

and $\hat{\omega}_F^2 = n[\text{std}(\hat{x}_n^*)]^2$. The values of the score mean are practically the same, standard deviations of ML estimates (computed, however, using two ML equations for p and q) are slightly lower.

Under contamination results became different Fig. 4 shows SM and ML estimates of the score mean of samples from the contaminated $BP_{cont} = 0.9BP(1, 4) + 0.1BP(p, 2)$ for increasing contamination expressed by increasing p , that is, with increasing score mean of contaminating distribution. The SM method is naturally better if strong contamination takes a part even in cases of the data from heavy-tailed distributions.

The distance in the sample space of a parametric distribution is obtained by generalizing the distance (16).

Example 6.2. Distances $d_F(x, 5)$ in the sample space of Weibull and Fréchet distributions for some values of parameter c are plotted in Fig. 5. Distances on the heavy-tail side are bounded.

The most interesting distance in the sample space is certainly that one between the hypothetical value x^* and its estimate \hat{x}_n^* from (21).

Theorem 3 *Random variable*

$$\delta_F(\hat{x}_n^*, x^*) = \frac{\sum_{i=1}^n S_F(x_i; x^*)}{nS'_F(x^*)}$$

is $AN(0, \omega_F^2/n)$.

Proof. By (16)

$$\delta_F(\hat{x}_n^*, x^*) = \frac{S_F(\hat{x}_n^*) - S_F(x^*)}{S'_F(x^*)}.$$

By Theorem 2 $\hat{x}_n^* - x^*$ is $AN(0, \omega_F^2/n)$, by Theorem A, Seffling (1980, pp. 118) $\delta_F(\hat{x}_n^*, x^*)$ is $AN(0, \omega_F^2/n)$ as well. As $S_F(x^*) = n^{-1} \sum_{i=1}^n S_F(x_i, x^*)$ and $S_F(\hat{x}_n^*) = 0$, the assertion holds true.

For distributions with a structural parameter the distance (16) is identical with that considered in the Rao scores tests and used for construction of Rao confidence intervals, cf. Lehman (2001, pp. 529). For distributions without a structural parameter it is a generalized Rao distance with respect to the

score mean. The Rao two-sided confidence interval for the sample score mean is thus determined from

$$d_F(\hat{x}_n^*, x^*) \leq \frac{u_{\alpha/2}}{\sqrt{n}} \omega_F \quad (23)$$

where $u_{\alpha/2}$ is the upper $\alpha/2$ point of the standard normal distribution.

Example 7.2. Given a sample \mathbb{X}_n according to $BP(p, q)$ and \hat{x}_n^* determined from (22),

$$\delta_F(\hat{x}_n^*, x^*) = \frac{1}{1+x^*} \frac{1}{n} \sum_{i=1}^n \frac{x_i - x^*}{x_i + 1}.$$

The two-sided generalized Rao confidence interval or \hat{x}^* follows from (23) and Table 4.

8 Other possible applications

Besides the point estimates, the simplicity of score-valued score functions (sdfs) predestines them for solution of various statistical problems in situations at which adequate models of continuous distributions are far from normal.

The first widely accepted application is the robust t-Hill estimator of the extremal value index of heavy-tailed distributions. The original procedure by Hill (1975) was considered by Stehlík and Fabián (2008) as an algorithm, modified by using sdfs of Pareto distribution instead of likelihood scores. The score mean of the distribution, the harmonic mean, has direct relation to this index. Properties of the t-Hill estimator were further studied by Stehlík et al. (2012), Beran and Shell (2012), Jordanova et al. (2013, 2016). Stehlík et al. (2020) suggested t-lgHill estimator using sfd of the log-gamma distribution.

Preliminary studies of a use of scalar-valued scores in other statistical tasks concerns the linear regression with positive data, Stehlík et al. (2019), estimation of the score correlation coefficient of samples from heavy-tailed distributions, Fabián (2010c) and the estimation of the power spectra of positive time series with outliers, Fabián (2010b). While in robust statistics are studied data influenced by outliers, our studies were oriented to data from heavy-tailed distributions where outlying data are regular values generated from the distribution. Results are showing that a use of a proper sdfs in

estimation procedures can be beneficial in situations with underlying skewed heavy-tailed distributions with great kurtosis.

9 Conclusions

In the paper we have described a new tool for study of continuous random variables. The suggested approach consists in dealing, instead with random variables X , with their treated forms, the scalar-valued score random variables (sfd), expressing relative influence of items from the distribution of X with respect to the typical value. Moments of score random variables are finite even in cases of heavy-tailed distribution, which enables to establish finite measures of central tendency and variability of continuous distributions. The approach provides a unified point of view on distributions both with and without structural parameter and on both the light-tailed and heavy-tailed ones. Another important result is establishing a one-to-one relation between the probability measure and the corresponding metric in the sample space.

As the scalar-valued scores are constructed by means of differentiation with respect to the variable, procedures developed for characteristics of standard distributions are immediately extendable to parametric distributions. While scalar-valued, the sfd can be used for estimation of parameters by a generalized moment method. In contrast with maximum likelihood estimates, the score moment estimates are direct consequences of the theory for standard distributions.

The concepts described here belongs within the range of classical statistics, but by choosing heavy-tailed parametric models one obtains robust results. For light-tailed models are robust considerations necessary, but perhaps feasible since inference functions are scalar-valued.

At the end we noted that knowing (assuming) a suitable parametric model underlying the data, the sufficiently general but simple scalar-valued score functions could be used in solutions of other statistical problems.

Acknowledgments

Author is indebted to prof. J. Antoch for valuable comments. The work was supported by Institutional Subsidy No. RVO:67985807.

References

- [1] Beran, J. and Schell, D. (2012). On robust tail index estimation. *Comp. Stat. Data Anal.*, **56**, 11, 3430-3443.
- [2] Cover, T. M. and Thomas, J. A. (2006). *Elements of Information Theory*, Second Edition, Wiley, New York.
- [3] Fabián, Z. (2001). Induced cores and their use in robust parametric estimation. *Communs. Stat. Theory Methods*, **30**, 537-556.
- [4] Fabián, Z. (2008). New measures of central tendency and variability of continuous distributions. *Communs. Stat. Theory Methods*, **37**, 2, 159-174.
- [5] Fabián, Z. and Stehlík, M. (2009). On robust and distribution sensitive Hill like method. *Tech. rep. of IFAS Reasearch Paper Series*, **43**, 4.
- [6] Fabián Z. Score moment estimators. In *Proc. of COMPSTAT 2010*, Physica-Verlag, 975-982.
- [7] Fabián Z. Spectral estimation of non-Gaussian time series. *Neural Network World*, **20**, 4, 491-499. ISSN 1210-0552.
- [8] Fabián Z. Score Correlation. *Neural Network World*, **20**, 6, 793-798. ISSN 1210-0552.
- [9] Fabián, Z. (2016). Score function of distribution and revival of the moment method. *Communs. Stat. Theory Methods*, **45**, 1118-1136.
- [10] Fabián, Z. (2021). Mean, mode or median? The score mean. *Communs. Stat. Theory Methods*, **50**, 2360-2370.
- [11] Fabián, Z. (2022). A measure of variability within parametric families of continuous distributions. *Communs. Stat. Theory Methods*, doi.org/10.1080/03610926.2022.2155792.
- [12] Fisher, R. A. (1922). On the mathematical foundations of theoretical statistics. *Philos. Tans. Roy. Soc. London*, A 222, 309-368.
- [13] Hill, B. M. (1975). A simple general approach to inference about the tail of a distribution. *Ann. Stat.*, **3**, 1163-1174.

- [14] Huber, P. J. (1964). Robust estimation of a location parameter. *Ann. Math. Statist.* **35**, 73-101.
- [15] Huber, P. J. and Ronchetti, E. M. (2009). *Robust statistics*. Second edition, New York, Wiley.
- [16] Johnson, N. L. (1949). Systems of frequency curves generated by methods of translations. *Biometrika*, **36**, 149-176.
- [17] Johnson, N. L., Kotz, S. and Balakrishnan, N. (1995). *Continuous Univariate Distributions 2*. New York, Wiley.
- [18] Jordanova, P. Stehlík, M. and Fabián, Z. (2013). On estimation and testing for Pareto tails. *Pliska Studia Mathematica Bulgarica*, **22**, 1, 89-108. ISSN 0204-9805.
- [19] Jordanova, P. et al. (2016). Weak properties and robustness of t-Hill estimators. *Extremes*, **19**, 591-626.
- [20] Marshall, A. W. and Olkin, I. (2007). *Life distributions*. Springer, New York.
- [21] Serfling, R. J. (1980). *Approximation Theorems of Mathematical Statistics*. New York, Wiley.
- [22] Stehlík, M. et al. (2010). On the favorable estimation of fitting heavy tailed data. *Comp. Statist.*, **25**, 485-503.
- [23] Stehlík, M. et al. (2019). On Ecological Aspects of Dynamics for Zero Slope Regression for Water Pollution in Chile. *Stochastic Analysis and Applications*, **37**, 4, 574-601. ISSN 0736-2994.
- [24] Stehlík, M. et al. (2020). Priority statement and some properties of t-lgHill estimator. *Extremes*, **23**, 3, 493-499.

Table 1: Score functions and densities of the basic set of standard distributions on R . K_0 is the Bessel function of the third kind.

$S_G(y)$	type	$g(y)$	distribution	$w_G(y)$
$\frac{e^y - e^{-y}}{2}$	UE	$\frac{1}{2K_0(1)}e^{-\cosh y}$	hyperbolic	$\frac{e^y + e^{-y}}{2}$
y	UP	$\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}y^2}$	normal	1
$e^y - 1$	BU	$e^y e^{-e^y}$	Gumbel	e^y
$1 - e^{-y}$	UB	$e^{-y} e^{-e^{-y}}$	extreme value	e^{-y}
$\frac{e^y - 1}{e^y + 1}$	BB	$\frac{e^y}{(1 + e^y)^2}$	logistic	$\frac{2e^y}{(e^y + 1)^2}$
$\frac{2y}{1 + y^2}$	BR	$\frac{1}{\pi(1 + y^2)}$	Cauchy	$\frac{2(1 - y^2)}{(1 + y^2)^2}$

Table 2: Densities, t-scores and weight functions of the basic set of distributions with support \mathcal{R}^+ . GIG means the Generalized Inverse Gaussian.

type	distribution	$f(x)$	$T_F(x)$	$w_F(x)$
UE	GIG	$\frac{1}{2K_0(1)x}e^{-\frac{1}{2}(x+1/x)}$	$\frac{1}{2}(x - 1/x)$	$\frac{1}{2}1/x^2$
UP	lognormal	$\frac{1}{\sqrt{2\pi x}}e^{-\frac{1}{2}\log^2 x}$	$\log x$	$1/x$
BU	exponential	e^{-x}	$x - 1$	1
UB	Fréchet	$\frac{1}{x^2}e^{-1/x}$	$1 - 1/x$	$1/x^2$
BB	loglogistic	$\frac{1}{(x+1)^2}$	$\frac{x-1}{x+1}$	$\frac{2}{(x+1)^2}$

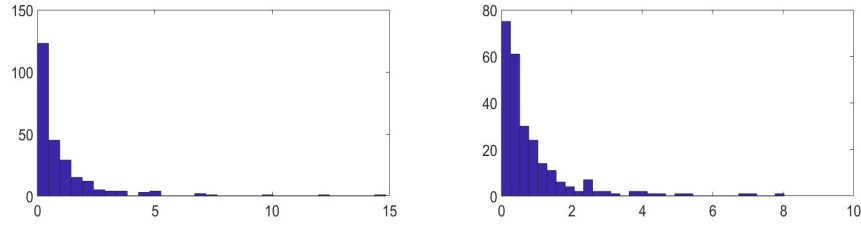


Figure 1: Histograms of random samples generated from beta-prime distribution with infinite variance

Table 3: The sample score mean of some standard distributions. \bar{x}_G means the geometric mean, \bar{x}_H harmonic mean and GIG means generalized inverse Gaussian.

\mathcal{X}	F	$f(x)$	$S_F(x)$	\hat{x}_F^*
$(0, 1)$	uniform	1	$2x - 1$	\bar{x}
\mathcal{R}	normal	$\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$	x	\bar{x}
\mathcal{R}^+	lognormal	$\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\log^2 x}$	$\log x$	\bar{x}_G
\mathcal{R}^+	exponential	e^{-x}	$x - 1$	\bar{x}
\mathcal{R}^+	inv. exponential	$\frac{1}{x^2} e^{-1/x}$	$1 - 1/x$	\bar{x}_H
\mathcal{R}^+	GIG	$\frac{1}{K_0(1)x} e^{-\frac{1}{2}(x+1/x)}$	$\frac{1}{2}(x - 1/x)$	$(\bar{x}\bar{x}_H)^{1/2}$

Table 4: Densities, score functions, typical values and score means of some distributions on \mathcal{R}^+ .

F	$f(x)$	$S_F(x)$	x^*	ω_F^2
Weibull	$\frac{c}{x}(x/\tau)^c e^{-(x/\tau)^c}$	$\frac{c}{\tau}[(x/\tau)^c - 1]$	τ	τ^2/c^2
Fréchet	$\frac{c}{x}(\tau/x)^c e^{-(\tau/x)^c}$	$\frac{c}{\tau}[1 - (\tau/x)^c]$	τ	τ^2/c^2
beta-prime	$\frac{1}{B(p,q)} \frac{x^{p-1}}{(x+1)^{p+q}}$	$\frac{q}{p} \frac{qx-p}{x+1}$	$\frac{p}{q}$	$\frac{p(p+q)^2}{q^3(p+q+1)}$

Table 5: Estimates of the score mean and score variance of two beta-prime distributions.

BP	$x^* = 2$	$n = 500$	$\omega_F^2 = 7.2$		$x^* = 2$	$n = 50$	$\omega_F^2 = 4/3$
$(1, 0.5)$	\hat{x}^*	std	$\hat{\omega}_F^2$		\hat{x}^*	std	$\hat{\omega}_F^2$
ML	2.0064	0.113	7.268	ML	2.0508	0.380	7.788
SM	2.0064	0.119	7.283	SM	2.0495	0.402	7.930
CLT			7.383	CLT			7.924
BP	$x^* = 1$	$n = 500$	$\omega_F^2 = 7.2$		$x^* = 1$	$n = 50$	$\omega_F^2 = 4/3$
$(1, 1)$	\hat{x}^*	std	$\hat{\omega}_F^2$		\hat{x}^*	std	$\hat{\omega}_F^2$
ML	1.0013	0.050	1.338	ML	1.0122	0.162	1.377
SM	1.0014	0.052	1.340	SM	1.0131	0.168	1.396
CLT			1.343	CLT			1.386

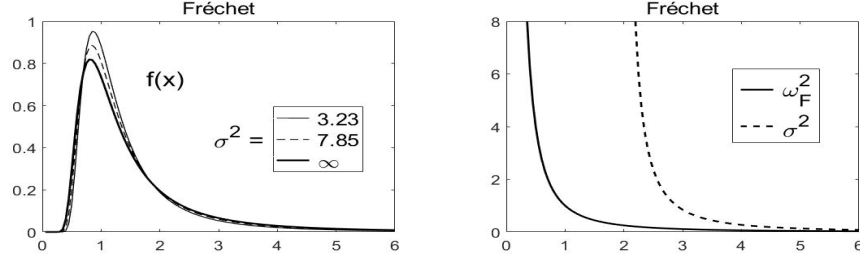


Figure 2: Similar Fréchet densities with very different variances and dependence of $\text{var}X$ and $\text{Var}_S X$ on increasing c

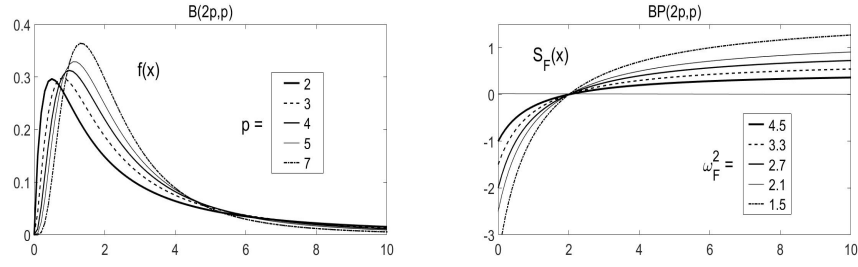


Figure 3: Densities and t-scores of $BP(2p, p)$ distribution

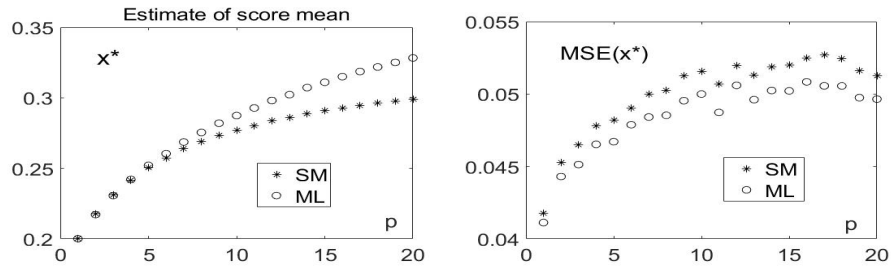


Figure 4: Estimates of the score mean and its mean square error in a contaminated $BP(1, 4)$ model

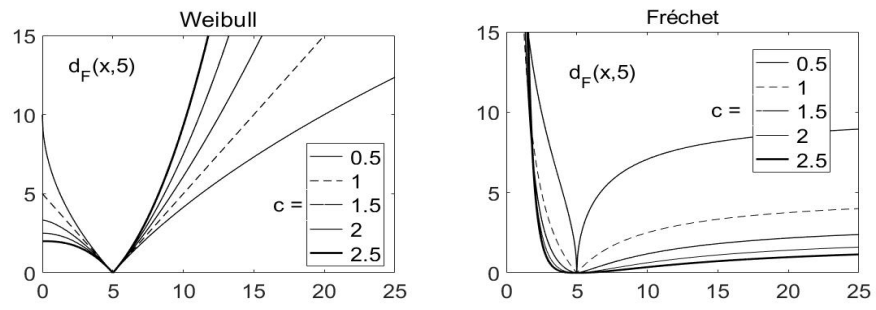


Figure 5: Distances $d_F(x, 5)$ in sample spaces of Weibull and Fréchet distributions