On the Relation Between Gnostical and Probability Theories

Fabián, Zdeněk
1996

INSTITUTE OF COMPUTER SCIENCE

ACADEMY OF SCIENCES OF THE CZECH REPUBLIC

ON THE RELATION BETWEEN GNOSTICAL AND PROBABILITY THEORIES

Zdeněk Fabián

Technical report No. 671

April 23, 1996

Institute of Computer Science, Academy of Sciences of the Czech Republic
Pod vodárenskou věží 2, 182 07 Prague 8, Czech Republic
phone: (+422) 66414244 fax: (+422) 8585789
e-mail: uivt@uivt.cas.cz

INSTITUTE OF COMPUTER SCIENCE

ACADEMY OF SCIENCES OF THE CZECH REPUBLIC

ON THE RELATION BETWEEN GNOSTICAL AND PROBABILITY THEORIES

Zdeněk Fabián

Technical report No. 671
April 23, 1996

Abstract

A description of continuous probability distributions by means of influence and weight functions of distribution has been developed. The applicability of the new concepts is briefly discussed. It is shown that in the case of special probability distribution these functions correspond to "irrelevance" and "fidelity" of the gnostical theory. Gnostical model of uncertainty, claimed by its author to be independent of probabilistic concepts, can be thus replaced by a special case of the classical probabilistic model.

Keywords
Probability, gnostical theory, estimates

1 Introduction of the influence function of a continuous distribution

R denotes real line. Let $T \subset R$ be an open interval and $\mathcal{B}_T$ the σ-field of its Borel subsets. Let $U_T : T \to R$ be a (continuous) random variable with distribution P_T, $P_T\{T\} = 1$. F_T denotes the distribution function and p_T the density of random variable U_T. Π_T denotes the set of all absolutely continuous distributions on $(T, \mathcal{B}_T)$ with densities twice continuously differentiable a.e.

Recall the concept of the score function of random variable $U_R : R \to R$, defined by

$$h_R(x) = \frac{d}{dx}(-\log p_R(x)) = -\frac{p_R'(x)}{p_R(x)}. \tag{1.1}$$

It is known from robust statistics (cf. [8]) that, for $T = R$ and the location model, the score function is proportional to the influence function of the maximum likelihood estimator.

A generalization of the score function for distributions defined on $(T, \mathcal{B}_T)$, where $T \neq R$, has been proposed in [4], [5]. It has been assumed that the set Π_T is an image of the set Π_R by using of some suitable diffeomorphism $\varphi : R \to T$. In other words, any U_T on $(T, \mathcal{B}_T)$ has a "prototype" U_R on $(R, \mathcal{B}_\mathcal{R})$ given by $U_R = \varphi^{-1}(U_T)$.

Such U_T and U_R and their distributions we call $\varphi-$related. The relation between their distribution functions is obviously

$$F_T(u) = F_R(\varphi^{-1}(u)). \tag{1.2}$$

The generalized score function belonging to U_T with distribution $P_T \in \Pi_T$, here called the influence function of the distribution of U_T, is defined as an image in T of the score function of its prototype under the mapping φ.

Definition 1. Let U_R be a random variable with distribution $P_R \in \Pi_R$ and with score function h_R. Let $T \subset R$ be an open interval. Let a mapping $\varphi : R \to T$ be strictly increasing diffeomorphism and let $U_T = \varphi(U_R)$. Real-valued function $h_T : T \to R$, given by

$$h_T(u) = h_R(\varphi^{-1}(u)), \tag{1.3}$$

will be called the influence function of random variable U_T or the influence function of its distribution (IFD) P_T.

Due to properties of the mapping φ, h_T exists. By the choice of the identical mapping $\varphi : R \to R$, $\varphi(u) = u$, the influence function of random variable U_R is the score function (1.1). An explicit form of the influence function of random variable U_T, $T \neq R$ is given by the following theorem.

Theorem 1. The influence function of random variable U_T specified in Definition 1 is given by

$$h_T(u) = \frac{1}{p_T(u)} \frac{d}{du}(-L(u)p_T(u)), \tag{1.4}$$

where

$$L(u) = \frac{d(\varphi(u))}{du}. \tag{1.5}$$

Proof. Denote $v = \varphi^{-1}(u)$. According to (1.2), the density of U_T is

$$p_T(u) = \frac{dF_T(u)}{du} = \frac{dF_R(v)}{dv}\frac{dv}{du} = p_R(v)L^{-1}(u) \tag{1.6}$$

by the formula for the inverse function derivative. By (1.3) and (1.1)

$$h_T(u) = h_R(v) = \frac{1}{p_R(v)}\,\frac{d}{dv}(-p_R(v)) = \frac{1}{L(u)p_T(u)}\frac{d}{du}(-L(u)p_T(u)) \cdot L(u)$$

□

The relation inverse to (1.4) is

$$p_T(u) = c^{-1}\exp\left(-\int L^{-1}(u)[h_T(u) + L'(u)]\,du\right) \tag{1.7}$$

(supposing that $c = \int_T p_T(u)du$ exists).

We specify the mapping φ for the case $T = R^+ = (0,\infty)$. We set $X = U_R$, $Z = U_{R^+}$ and

$$Z = \varphi(X) = e^X, \tag{1.8}$$

so that $X = \varphi^{-1}(Z) = \ln Z$. Denote by $p(z), h(z)$, $z \in R^+$ the corresponding density and influence function of Z. Then

$$L(z) = z, \tag{1.9}$$

IFD of Z is given by

$$h(z) = -1 - zp'(z)/p(z). \tag{1.10}$$

and (1.7) reduces to

$$p(z) = c^{-1}\exp\left(-\int z^{-1}[h(z)+1]\,dz\right). \tag{1.11}$$

This choice of the mapping φ for $T = R^+$ seems to be in the spirit of statistics. Positive data are often logarithmically transformed and some pairs of distributions on R^+ and R (the lognormal and normal, the log-Cauchy and Cauchy etc.) are known to be "logarithmically related". We only generalize this concept to all $P_{R^+} \in \Pi_{R^+}$.

Let now $\Theta \subset R^m$ be an open convex set. Let $\mathcal{P}_T = \{P_\theta | \theta \in \Theta\}$ be a parametric family of distributions on $(T, \mathcal{B}_T)$, dominated by Lebesgue measure, with densities $\{p_T(u|\theta) | \theta \in \Theta\}$. A generalization of the influence function of the distribution for the parametric set $\mathcal{P}_T$ is straightforward: it is a parametric set $\{h_T(u|\theta) | \theta \in \Theta\}$, where

$$h_T(u|\theta) = \frac{1}{p_T(u|\theta)}\,\frac{d}{du}(-L(u)p_T(u|\theta)) \tag{1.12}$$

and where $L(u)$ is given by (1.5).

Consider a special case of a parametric family with location and scale structure. When $T = R$, the location parameter $x_0 \in R$ represents a shift of the mode of the density p_R along real line and $p_R(x|x_0,\sigma) = \sigma^{-1}\tilde{p}((x-x_0)/\sigma)$ where $\tilde{p}$ is the parent ("prototype") density and $\sigma \in R^+$ is the scale parameter. The score function is, by (1.1), $h_R(x|x_0,\sigma) = \sigma^{-1}\tilde{h}((x-x_0)/\sigma)$ where $\tilde{h}$ is the "prototype" score function.

We define the generalized location parameter $u_0 \in T$ of the φ-related distribution P_T on $(T, \mathcal{B}_T)$ by the relation $u_0 = \varphi(x_0)$. Densities in the location and scale model on $(R^+, \mathcal{B}_{R^+})$ are, by (1.6),

$$p(z|z_0,\sigma) = z^{-1}\sigma^{-1}\tilde{p}(\sigma^{-1}(\ln z \ - \ln z_0)) = z^{-1}\sigma^{-1}\tilde{p}(\ln(z/z_0)^{1/\sigma}). \qquad (1.13)$$

and the corresponding IFDs

$$h(z|z_0,\sigma) = \sigma^{-1}\tilde{h}(\ln\,(z/z_0)^{1/\sigma}). \qquad (1.14)$$

2 Some properties of the influence function of the distribution

We briefly mention some properties of IFD's, discussed in more details in [5]-[7].

i/ IFD represents an equivalent and usually simpler description of the distribution as density.

Due to assumptions, (1.4) and (1.7) represent a one-to-one correspondence between density and IFD of a continuous probability distribution. The simplicity of IFD's is apparent from some examples given in Table 1.

TABLE 1 *IFDs and densities of distributions on* $(R, \mathcal{B}_R)$ *and of* e^x-*related distributions on* $(R^+, \mathcal{B}_{R^+})$

$h_R(x)$	$p_R(x)$	$h(z)$	$p(z)$
x	$\frac{1}{\sqrt{2\pi}}e^{-x^2/2}$	$\ln z$	$\frac{1}{\sqrt{2\pi}\,z}e^{-\frac{1}{2}\ln^2 z}$
e^x-1	$e^x e^{-e^x}$	$z-1$	e^{-z}
$\mathrm{tgh}(x/2)$	$\frac{1}{4}\cosh^{-2}(x/2)$	$(z-1)/(z+1)$	$1/(z+1)^2$
$\sinh x$	$\frac{1}{2K_0(1)}e^{-\cosh x}$	$\frac{1}{2}(z-1/z)$	$\frac{1}{2K_0(1)\,z}e^{-\frac{1}{2}(z+1/z)}$

Here K_0 is the Bessel function of the III. kind. In the first three rows of Table 1 are standardized forms of pairs of e^x-related distributions: normal and lognormal, double exponential and exponential, logistic and log-logistic. The distribution with density p_R in the fourth row is not encountered in statistics, the e^x-related one is Wald-type.

ii/ In a generalized location model, IFD is proportional to the partial maximum likelihood score function.

Recall that the partial maximum likelihood score functions are defined as

$$r_j(u|\theta) = \frac{\partial}{\partial \theta_j}(\log p(u|\theta)), \qquad j = 1, ..., m.$$

Theorem 2. Let u_0 be the location parameter of a parametric family $\{P_{T\theta}|\theta \in \Theta\}$ on $(T, \mathcal{B}_T)$, where $\theta = (u_0, \alpha)$, $\alpha = \theta_2, ..., \theta_m$. Let the partial maximum likelihood score function $r_1(u|u_0, \alpha)$ exists. Then

$$h_T(u|u_0, \alpha) = L(u_0) r_1(u|u_0, \alpha).$$

Proof. Let $P_{T\theta} = \varphi(P_{R\theta'})$ where $\theta' = (x_0 = \varphi^{-1}(u_0), \alpha)$ and denote $v = \varphi^{-1}(u) - \varphi^{-1}(u) = x - x_0$. Analogically to (1.6), $p_T(u|\theta) = L^{-1}(u) p_R(v|\alpha)$. Then

$$r_1(u|\theta) = \frac{1}{p_T(u|\theta)} \frac{\partial p_T(u|\theta)}{\partial u_0} =$$

$$= \frac{L(u)}{p_R(v|\alpha)} \frac{\partial (L^{-1}(u) p_R(v|\alpha))}{\partial v} \frac{dv}{du_0} = -\frac{p_R{}'(v|\alpha)}{p_R(v|\alpha)} L^{-1}(u_0) =$$

$$= h_R(v|\alpha) L^{-1}(u_0) = L^{-1}(u_0) h_T(u|\theta).$$

□

iii/ *The IFD-moments are better numerical characteristics of continuous random variables than the classical moments.*

Consider T and φ specified in Definition 1. Let p_T be the density and h_T the influence function of random variable U_T with distribution $P_{T\theta} \in \Pi_T$. Let $k \in \mathcal{N}$. The k-th IFD-moment of random variable U_T has been defined in [5,7] by the integral

$$M_k(\theta) = \int_T h_T^k(u|\theta) p_T(u|\theta)\, du. \tag{2.1}$$

It has been shown in [7] that the IFD-moments exist even in cases of distributions, of which the usual moments do not exist (Cauchy and log-logistic distributions, for instance).

Let $c_1 = \inf\{u : u \in T\}, c_2 = \sup\{u : u \in T\}$. By (1.4) and (1.6),

$$M_1 = -L(u) p_T(u)|_{-c_1}^{c_2} = -p_R(x)|_{-\infty}^{\infty} = 0. \tag{2.2}$$

All the other IFD-moments are expressed by means of parameters only and not by some non-elementary functions of parameters, as it often appears in the case of usual moments. Estimate $\hat{\theta}$ of the true parameter θ^0 given by equations

$$\hat{\theta}: \qquad \sum_{i=1}^n h_T^k(u_i|\hat{\theta}) = M_k(\hat{\theta}), \qquad k = 1, ..., m, \tag{2.3}$$

where $u_1, ..., u_n$ are observed values of independent, identically distributed (i.i.d.) random variables with distribution P_{θ^0}, are consistent and asymptotically normal.

The special form of (2.3) for distributions with the location and scale structure on $(R^+, \mathcal{B}_{R^+})$ is, by (2.2) and (1.14),

$$\sum_{i=1}^{n} \tilde{h}(\ln\,(z_i/\hat{z}_0)^{1/\hat{\sigma}}) = 0 \tag{2.4}$$

$$\sum_{i=1}^{n} \tilde{h}^2(\ln\,(z_i/\hat{z}_0)^{1/\hat{\sigma}}) = \hat{\sigma}^2 M_2(\hat{z}_0, \hat{\sigma}). \tag{2.5}$$

According to Theorem 2, the first moment equation (2.4) is identical with the maximum likelihood equation for the location parameter. It has been shown in [7] that in cases of distributions with bounded IFDs, the asymptotic variances of estimates (2.3) are near to the Cramer Rao bound. Simultaneosly, IFD-moment estimates of both location and scale parameters are robust, whereas the ML estimates of the scale parameter are sensitive to outlier values.

iv/ The second IFD-moment is a generalization of the Fisher information.

The Fisher information $FI(\theta)$ is usually defined and interpreted with respect to parameters of parametric distributions. The definition of the Fisher information of the distribution (without parameters) we found in [1, pp.494]. It is not sufficiently general, however. The mean value of h_T^2,

$$M_2 = \int_T h_T^2(u)\; p_T(u) du \tag{2.6}$$

appears to be a correct generalization of the Fisher information for distributions defined on arbitrary $(T, \mathcal{B}_T)$, since M_2 is, according to Theorem 2, proportional in the location model to the Fisher information for the location parameter, $M_2(u_0) = L^{-1}(u_0)FI(u_0)$, even when $T \neq R$.

It is, hopefully, apparent from this short discussion that the IFD can be understood as a fundamental concept connected to a continuous distribution.

3 Weight function of a distribution

Definition 2. Let h_T be the influence function of a distribution $P_{T\theta} \in \mathcal{P}_T \subset \Pi_T$. A real-valued function $g_T : T \to R$, given by

$$g_T(u|\theta) = \frac{dh_T(u|\theta)}{du} \tag{3.1}$$

will be called the weight function of the distribution (WFD) $P_{T\theta}$.

By means of the IFD, a distance in the sample space T can be introduced by the formula

$$\rho(u_1, u_2|\theta) = |h(u_2|\theta) - h(u_1|\theta)| = \int_{u_1}^{u_2} g_T(u|\theta)\; du \qquad u_1, u_2 \in \Theta. \tag{3.2}$$

By Theorem 2, (3.2) is proportional to the distance, which is in fact introduced in the sample space by the maximum likelihood estimator of the (generalized) location parameter.

(3.2) appears to be a metric for a continuous, strictly increasing h_T. The space (T, g) is in such a case a one-dimensional Riemannian metric space. In accordance to concepts of the Riemannian geometry (we refer to [9]), g_T represents the weight introduced in the space T.

It follows from the direct differentiation of (1.12) and Theorem 1 that weight functions of $\varphi-$related distributions are related by

$$g_T(u|\theta) = L^{-1}(u) g_R(\varphi^{-1}(u|\theta)). \tag{3.3}$$

The term

$$w_T(u|\theta) = g_R(\varphi^{-1}(u|\theta)) \tag{3.4}$$

is called in a further account the proper weight function of the distribution P_T. We suppose that the proper WFD could represent a relative weight (i.e. a relative importance) of a point $u \in T$ (or of the observed value u) under the assumption of the distribution $P_{T\theta}$. We do not know, however, contrary to the case of the IFD, a direct application of the WFD. The generalization of the WFD in a case of a parametric family $\mathcal{P}_T$ is the metric tensor of the the statistical variety $\mathcal{P}_T$ (see [1]).

Consider for the sake of simplicity a distribution without parameters with density $p_T(u)$ and IFD $h_T(u)$, so that the WFD is given by $g_T(u) = dh_T(u)/du$. Taking derivatives of (1.1), (1.10) with respect to x and z, respectively, we obtain WFD's on $(R, \mathcal{B}_R)$ and $(R^+, \mathcal{B}_{R+})$ expressed by densities as

$$g_R(x) = \left(\frac{p_R'(x)}{p_R(x)}\right)^2 - \frac{p_R''(x)}{p_R(x)}, \qquad g(z) = -\frac{p'(z)}{p(z)} + z\left[\left(\frac{p'(z)}{p(z)}\right)^2 - \frac{p''(z)}{p(z)}\right].$$

Example. In Table 2, the proper weight functions $w(z)$ of distributions from Table 1 are given.

TABLE 2 *Densities p_R, weight functions g_R of distributions on $(R, \mathcal{B}_R)$ from Table 1 and proper densities q, weight functions g and proper weight functions w of e^x-related distributions on $(R^+, \mathcal{B}_{R+})$*

$p_R(z)$	$g_R(x)$	$p(z)$	$g(z)$	$w(z)$
$\frac{1}{\sqrt{2\pi z}} e^{-\frac{1}{2}x^2}$	1	$\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\ln^2 z}$	$1/z$	1
$e^x e^{-e^x}$	e^x	e^{-z}	1	z
$\frac{1}{4}\cosh^{-2} x$	$\frac{1}{2}\cosh^{-2} x$	$1/(z+1)^2$	$2/(z+1)^2$	$(\sqrt{2}/(z^{1/2}+z^{-1/2}))^2$
$\frac{1}{2K_0(1)} e^{-\cosh x}$	$\cosh x$	$\frac{1}{2K_0(1)\, z} e^{-\frac{1}{2}(z+1/z)}$	$\frac{1}{2}(1+1/z^2)$	$\frac{1}{2}(z+1/z)$

The weight function of the distribution with the location and scale structure on $(R^+, \mathcal{B}_{R+})$ is, using (3.1), (1.14) and (3.3), given by

$$g(z|z_0, \sigma) = \sigma^{-1} d\tilde{h}(\ln(z/z_0)^{1/\sigma})/dz = \sigma^{-2} z^{-1} \tilde{g}(\ln(z/z_0)^{1/\sigma}) \tag{3.5}$$

where $\tilde{g} = \tilde{h}'$ is the "prototype" weight on $(R, \mathcal{B}_R)$.

4 Gnostical theory

A very unusual theory of data treatment was presented by Kovanic [10]-[13]. The aim of his "gnostical" theory is the same as that of statistics: to make inferences from data observed under the influence of uncertainty. The theory is believed by its author to be completely independent of the probabilistic model and of basic concepts in probability theory.

Kovanic introduced a mathematical model of an individual uncertainty which is contained in a single positive data item z in the form

$$z = z_0 e^{s\Omega} \tag{4.1}$$

where $z_0 \in R^+$ is an "ideal value" of z and $\Omega \in R$ the uncertainty, scaled (in [13]) by parameter $s \in R^+$. Since (4.1) seems to be a general parametric model of positive data items and any real measured data are in fact positive Kovanic considered that (4.1) is a universal mathematical model of data which are "suffered from uncertainty". Based on this model, he derived two individual "gnostical" data characteristics that depend on uncertainty. They are "fidelity" , given by the expression

$$f(z|z_0, s) = \cosh^{-1}(2\Omega) = 2/[(z/z_0)^{2/s} + (z/z_0)^{-2/s}], \tag{4.2}$$

and "irrelevance" , given by

$$h_e(z|z_0, s) = -\mathrm{tgh}(2\Omega) = -\frac{(z/z_0)^{2/s} - (z/z_0)^{-2/s}}{(z/z_0)^{2/s} + (z/z_0)^{-2/s}}, \tag{4.3}$$

with mutual relation to

$$h_e^2(z|z_0, s) = 1 - f^2(z|z_0, s). \tag{4.4}$$

These are the two basic gnostical characteristics of one data item when the model (1) is known.

On the other hand, having a sample $\mathbf{Z}_n = (z_1, \ldots, z_n)$ of data from one source (4.1), each data item z_i can be characterized, after Kovanic, by its fidelity and irrelevance. They are in latent form because of unknown parameters z_0, s which can, however, be estimated from the data sample $\mathbf{Z}_n$. The simplest gnostical estimate of the ideal value z_0 is obtained by Kovanic's requirement of zero average irrelevance of the sample $\mathbf{Z}_n$. This gives the estimation equation

$$\hat{z}_0 : \qquad \frac{1}{n} \sum_{i=1}^{n} h_e(z_i|\hat{z}_0, \hat{s}_a) = 0, \tag{4.5}$$

where $\hat{s}_a$ is a prior estimate of the scale parameter s. The function h_e in (4.3) is bounded,
$|h_e(z|z_0)), s)| \leq 1$. A consequence of this fact is the insensitivity of estimates (4.5) to

outlying values in data, without introducing any of the robustifying functions of robust statistics. The fact that the gnostical estimator (4.5) can be useful, was demonstrated by its comparison with a large set of robust statistical estimators. They were all applied to the well-known collection of Stiegler's data [17]. The gnostical estimator, giving quite realistic estimates, was found ([14]) to have the smallest mean square error.

Other gnostical data characteristics and estimation procedures take various forms, some of them being restatements of well-known statistical principles with one basic difference: instead of raw data, the irrelevances are substituted into formulas. As an example, the "gnostical correlation coefficient" is $C_e(k) = \frac{1}{n-k}\sum_{i=1}^{n-k} h_e(z_i|z_0,s)h_e(z_{i+k}|z_0,s)$.

The more advanced gnostical estimation procedures, which we do not discuss in the present paper, are based on the "data composition law" of the gnostical theory, which states that the "composite event" z_c of a data sample $\mathbf{Z}_n$ is given by

$$h_e(z_c|z_0,s) = \sum_{i=1}^{n} h_e(z_i|z_0,s)/w_e, \tag{4.6}$$

where $w_e = ([\sum_{i=1}^n f(z_j|z_0,s)]^2 + [\sum_{i=1}^n h_e(z_j|z_0,s)]^2)^{1/2}$, i.e. that the irrelevance of the composite event is the weighted sum of individual irrelevances.

Kovanic believes that the "gnostical data processing" differ from that of statistics in substance ([13], pp. 657). He also believes that the gnostical model of individual uncertainty contained in data, given by (4.1), is generally applicable for small data samples. He asserts that it can be used even in situations when a probabilistic model of the data is unknown and cannot be guessed ("Let data speak for themselves", [13], pp.658).

The first serious argument against these assertions was given in [3]. The author of the present paper noticed that the square of fidelity (4.2) is similar to the density of a certain probability distribution, later identified as log-logistic. He also showed that gnostical estimators are identical to the maximum-likelihood estimator or to α-estimators introduced by Vajda [20], in the case of this distribution. Based on this result, Vajda [21], [22] and Novovičová [15] studied properties of gnostical estimators. Apart from Kovanic's further attempts to consider only finite n-point "data varieties", they proved that gnostical estimators are the usual statistical M-estimators, strongly consistent and asymptotically normal. They also derived their asymptotic variances.

The success of the estimator (4.5) applied to the Stiegler data sets can be explained simply. The influence function of the robust estimator (4.5) is, contrary to usual robust estimators, non-symmetrical. This coincides with the clear non-symmetry of Stiegler's data. Nevertheless, some questions concerning gnostical theory remain unanswered. What does it the "fidelity" and "irrelevance" of one data item really mean ? Why the gnostical estimator (4.5) belongs to the class of statistical M-estimators, although the maximum likelihood principle is not postulated in gnostical theory ?

5 Gnostical irrelevance and fidelity as the influence function and the square root of the weight function of special probability distributions

In the previous section we mentioned only one of Kovanic's irrelevances. In fact, there are two. By means of "estimating irrelevance" , given by (3), there are constructed robust gnostical estimates. The second type is the "quantifying irrelevance", given by

$$h_q(z|z_0,s) = \sinh(2\Omega) = \frac{1}{2}[(z/z_0)^{2/s} - (z/z_0)^{-2/s}]. \qquad (5.1)$$

The requirement of zero average of quantifying irrelevances of a data sample provides sensitive gnostical estimates [13].

Theorem 3. Probability densities corresponding to two types of Kovanic's irrelevances (4.3), (5.1) are

$$p_1(z|z_0,s) = \frac{2\sqrt{2\pi}}{zs\Gamma^2(1/4)}\,\frac{1}{[(z/z_0)^{2/s} + (z/z_0)^{-2/s}]^{1/2}} \qquad (5.2)$$

$$p_2(z|z_0,s) = \frac{1}{zsK_0(1/2)}\;e^{-\frac{1}{4}[(z/z_0)^{2/s}+(z/z_0)^{-2/s}]}, \qquad (5.3)$$

respectively.

Proof. Let

$$h_{R1}(u) = \mathrm{tgh}(2u), \qquad h_{R2}(u) = \sinh(2u) \qquad (5.4)$$

be score functions of some distributions. The corresponding densities are

$$p_{R1}(u) = c_1^{-1} e^{-\int \mathrm{tgh}(2u)\,du} = c_1^{-1}\cosh^{-1/2}(2u) \qquad (5.5)$$

$$p_{R2}(u) = c_2^{-1} e^{-\int \sinh(2u)\,du} = c_2^{-1} e^{-\frac{1}{2}\cosh(2u)}. \qquad (5.6)$$

By the use of integrals

$$\int_0^\infty \cosh^{-\nu} ax\,dx = \frac{2^{\nu-1}}{a\Gamma(\nu)}\Gamma^2(\nu/2), \qquad \int_0^\infty z^{\alpha-1} e^{-(pz\ +q/z)} dz = 2(q/p)^{\alpha/2} K_\alpha(2\sqrt{pq})$$

where Γ is the gamma function and K_α the modified Bessel function of the third kind (see [15]), norming constants are $c_1 = \Gamma^2(1/4)/2\sqrt{2\pi}$, $c_2 = K_0(1/2)$. By the substitution

$$u = \ln(z/z_0)^{1/s} \qquad (5.7)$$

into (5.4) and using (1.14), one obtains influence functions of searched distributions in the form

$$h_1(z|z_0,s) = s^{-1}\mathrm{tgh}(\ln(z/z_0)^{2/s}) = -s^{-1} h_e(z|z_0,s) \qquad (5.8)$$

$$h_2(z|z_0,s) = s^{-1}\sinh(\ln(z/z_0)^{2/s}) = s^{-1} h_q(z|z_0,s), \qquad (5.9)$$

where $-h_e$ and h_q are gnostical irrelevances given by (4.3) and (5.1). Substituting (5.7) into (5.5) and (5.6) and using transformation relations (1.6), one obtains the searched densities (5.2) and (5.3).

The opposite sign of the estimating irrelevance with respect to IFD, as well as the constant factor, plays no role in practical applications of gnostical algorithms. Also, considering the one-to-one correspondence of IFDs and densities, the assertion holds. □

Theorem 4. Square of the gnostical fidelity is, apart from the constant, the weight function of the distribution of the family given by (5.2).

Proof. Weight functions of distributions with densities (5.5), (5.6) are, using (3.1) and (5.4)

$$g_1(u) = 2\cosh^{-2}(2u), \qquad g_2(u) = 2\cosh(2u). \tag{5.10}$$

After substitution (5.7) and by the use of (3.5),

$$g_1(z|z_0, s) = 2s^{-2}z^{-1}\cosh^{-2}(\ln(z/z_0)^{2/s}) = 2s^{-2}z^{-1}f^2(z|z_0, s) \tag{5.11}$$

$$g_2(z|z_0, s) = 2s^{-2}z^{-1}\cosh(\ln(z/z_0)^{2/s}) = 2s^{-2}z^{-1}f^{-1}(z|z_0, s),$$

where f is the fidelity (4.2). After comparing to (3.5) and (3.4) and apart from the constant factor, f^2 is the proper weight function of the distribution (5.2) (and, similarly, f^{-1} is the proper weight function of the distribution (5.3)). □

6 Conclusion

Given a model of a statistical experiment in the form of a parametric set $\mathcal{P}_T$, the observed values $u_1, ..., u_n$, the realizations of i.i.d. random variables $U_1, \ldots U_n$ with distribution $P_{\theta^0} \in \mathcal{P}_T$ are no longer merely an observed collection of data items. For each data item u_i are, by the assumed model, prescribed the a priori data characteristics: the value of the influence function of the distribution, $h_T(u_i|\theta^0)$, and the value of the proper weight function of the distribution, $w_T(u_i|\theta^0)$. They are, similarly as with the likelihood, in latent form because of an unknown θ^0. They can be approximately determined after an estimate $\hat{\theta}$ of the true value θ^0 is found.

Bearing it in mind, theorems in the previous section give a possible statistical explanation of gnostic characteristics of "one data item". The "ideal value" z_0 can be understood as the (generalized) location parameter, the scale s as the usual scale parameter and the "irrelavance" and "fidelity" as the influence function and the square root of the weight function of distributions (5.2) and (5.3). We thus give an explanation of Kovanic's "non-statistical" notions of irrelevance and fidelity of individual data in a rather unexpected fashion by including their general equivalents into the probability theory.

The Kovanic's heuristic estimate given by (4.5) appears to be the first IFD-moment estimate in the special model (5.2). By iii/ of section 2, (2.4) yields the maximum likelihood estimate of the location parameter without a need of the maximum likelihood principle (e.g. without a need of differentiation with respect to the location parameter). To this fact, together with the boundedness of the influence function of the distribution (5.2)), the success of the gnostical estimator of the location parameter can be attributed. The difficulties with gnostical estimation of the scale parameter (which we did not mentioned in our account) could be removed by the use of the second IFD-moment estimation equation (2.5).

It should be noted that we did not explain Kovanic's estimation procedures based on his "data composition law". We suppose that, in probabilistic terms, the composition law (4.6) can be considered to be a "finite equivalent" of some limit theorem concerning sums of i.i.d. random variables, weighted by a special way. "Qualitatively", (4.6) asserts that the weighted sum of i.i.d. random variables is distributed according to the original probability law. This idea might be interesting, but it should be proved or disproved.

Bibliography

[1] S. Amari: Differential-Geometrical Methods in Statistics. Lecture Notes in Statistics *28*, Springer-Verlag, Heidelberg, 1985.

[2] T.M.Cover, J.A.Thomas: Elements of Information Theory, Wiley, New York, London, 1991.

[3] Z. Fabián: Point estimation in case of small data sets. Trans. 10th Prague Conf. on Information Theory, held at Prague 1986, Academia, Prague, 1988, 305-312.

[4] Z. Fabián: Metric function and its use in neural networks. Neural Network World *2* (1994), 133-140.

[5] Z. Fabián: Generalized score function and its use. Trans. 12th Prague Conf. on Information Theory, Academia, Prague 1994.

[6] Z. Fabián: Information and entropy of continuous random variables. Techn. rep. V-633, ICS AS CR, Prague 1995.

[7] Z. Fabián: Geometric moments and geometric moment estimates. Techn. rep. V-656, ICS AS CR, Prague 1995.

[8] F.R. Hampel, P.J. Rousseeuw, E.M. Ronchetti and W.A. Stahel: Robust Statistic. The Approach Based on Influence Functions. Wiley, New York 1987.

[9] S.Kobayashi and K.Nomizu: Foundations of differential geometry. Intersci Publishers, New York, London 1963.

[10] P.Kovanic: Gnostical theory of individual data. Problems of Control and Information Theory *13(4)* (1984), 259-274.

[11] P. Kovanic: Gnostical theory of small samples of real data. Problems of Control and Information Theory *13(5)* (1984), 303-319.

[12] P. Kovanic: On relation between information and physics. Problems of Control and Information Theory *13(6)* (1984), 383-399.

[13] P.Kovanic: A new theoretical and algorithmical tool for estimation, identification and control. Automatica *22*, 6 (1986), 657-674.

[14] P. Kovanic and J. Novovičová: Comparizon of statistical and gnostical estimates of parameter of location on real data. Proc. of conf. ROBUST. JČMF, Prague 1986 (in Czech).

[15] J. Novovičová: M-estimators and gnostical estimators of location. Problems of Control and Information Theory *18(6)* (1989), 397-407.

[16] A.P. Prudnikov, J.A. Bryčkov and O.I. Maričev: Integrals and series. Moskva, Nauka 1981 (in Russian).

[17] S.M. Stiegler: Do robust estimators work with real data ? Ann. Statist. *6* (1977), 1055-1098.

[18] J. Šindelář: On L-estimators viewed as M-estimators. Kybernetika *30* (1994), 551-562.

[19] J. Štěpán: Probability theory. Academia, Praha 1987 (in Czech).

[20] I. Vajda: Efficiency and robustness control via distorted maximum likelihood estimation. Kybernetika *22* (1986), 47-67.

[21] I. Vajda: Minimum-distance and gnostical estimators. Problems of Control and Information Theory *17(5)* (1987), 253-266.

[22] I. Vajda: Comparison of asymptotic variances for several estimators of location. Problems of Control and Information Theory *18(2)* (1989), 79-87.