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

Z. Fabián

Technical report No. 624

4.3. 1995

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: zdenek@uivt.cas.cz

INSTITUTE OF COMPUTER SCIENCE

ACADEMY OF SCIENCES OF THE CZECH REPUBLIC

On the relation between gnostical and probability theories

Z. Fabián

Technical report No. 624
4.3. 1995

Abstract

Taking into account transformation relations between distributions defined on different supports, an alternative description of continuous probability distributions by means of influence functions of distribution and weight functions has been developed. They correspond in cases of special probability distributions to “irrelevances” and “fidelity” of gnostical theory. Gnostical theory, claimed to be quite independent of probabilistic concepts, appears thus to be the special case of the classical probabilistic model.

Keywords

Research report, probability, gnostic theory, influence function of the distribution, LaTeX

1 Introduction: The gnostical theory

A very unusual theory of data treatment was presented by Kovanic [7]-[10]. 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, however, to be completely independent of the probabilistic model and of basic concepts of probability. Let us give a concise exposition of its main ideas.

By Axiom 1, Kovanic introduced a mathematical model of an individual uncertainty contained in a single positive data item z in the form

$$z = z_0 e^{s\Omega} \tag{1.1}$$

where z_0 is an "ideal value" of z and Ω the uncertainty, scaled in [10] by parameter s. Since (1.1) seems to be a general parametric model of positive data items, and any real measured data are in fact positive, Kovanic considered that (1.1) is a universal mathematical model of data "suffered from uncerntainty". By an ingenious reasoning, based on analogies from thermodynamic and relativistic mechanics (description of which we omit, referring to original papers [7], [10]), he derived from (1.1) two individual data characteristics depending on uncertainty. They are "fidelity" , given by the expression

$$f(z|z_0, s) = \cosh^{-1}(2\Omega) = 2/[(z/z_0)^{2/s} + (z/z_0)^{-2/s}], \tag{1.2}$$

and "irrelevance" , given by

$$h_e(z|z_0, s) = -\mathrm{tgh}(2\Omega) = -\frac{(z/z_0)^{2/s} - (z/z_0)^{-2/s}}{(z/z_0)^{2/s} + (z/z_0)^{-2/s}}, \tag{1.3}$$

with mutual relation

$$h_e^2(z|z_0, s) = 1 - f^2(z|z_0, s). \tag{1.4}$$

These two mysterious gnostical characteristics of each data item are in "latent form" because of unknown parameters z_0, s of the source of data (1.1), and represent, after Kovanic, a (latent) weight (fidelity) and (latent) relative error (irrelevance) of the data item. Having a sample $\mathbf{Z}_n = (z_1, \ldots, z_n)$ of data of the same origin, the unknown parameters z_0, s in (1.2) and (1.3) can be estimated either by the simple requirements similar to some statitistical ones (see below), or by composition law, presented as Axiom 2 of the theory, which states that the "composite event" z_c of a data sample $\mathbf{Z}_n$ is given by

$$h_e(z_c|z_0, s) = \sum_{i=1}^{n} h_e(z_i|z_0, s)/w_e, \tag{1.5}$$

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 irrelevance of the composite event is the weighted sum of individual irrelevances.

Gnostical estimation procedures take various forms, from pure heuristics to restatements of well-known statistical principles, with one basic difference: instead of raw data, the irrelevances are substituted into computational formulas. For instance, the "gnostical correlation coefficient" is given by $C_e(k) = \frac{1}{n-k}\sum_{i=1}^{n-k} h_e(z_i|z_0, s)h_e(z_{i+k}|z_0, s)$.

The simplest gnostical estimate of the ideal value z_0 is obtained by the requirement of zero average irrelevance of the sample $\mathbf{Z}_n$. This gives an estimation equation of the ideal value z_0 in the form

$$\hat{z}_0 : \qquad \frac{1}{n}\sum_{i=1}^{n} h_e(z_i|\hat{z}_0, \hat{s}_a) = 0, \tag{1.6}$$

where $\hat{s}_a$ is some a priory estimate of the scale parameter s. The function h_e in (1.3) is bounded, $|h_e(z|z_0)), s)| \leq 1$. An immediate consequence of this fact is the B-robustness of estimates (1.6). Furthemore, the distance between two data items in "data variety", measured by means of irrelevances, given by

$$\rho_e(z_1, z_2) = |h_e(z_2|z_0, s) - h_e(z_1|z_0, s)|, \tag{1.7}$$

is bounded too, as $rho_e(z_1, z_2) < 2$.

Summarizing Kovanic's approach, there are three interesting ideas in the context of data processing:

(I) The gnostical model of uncertainties contained in data, given by (1.1), is generally applicable and independent of probabilistic models. Gnostical data processing is an alternative to statistical data processing.

(II) Each data item contains its own (gnostical) "weight" and "relative error", which can be approximately expressed after estimating z_0, s from a set of data of the same origin.

(III) Estimation procedures are to be based on the geometry in "data variety", which is non-Euclidean.

2 Case of history

At the time when the theory appeared, it was, during discusions on author's lectures, completely rejected by statisticians. On the other hand, results of the application of gnostical estimators and procedures to artificial as well as to real data were outstanding. From the comparison of the gnostical estimator (1.6) with a large set of robust statistical estimators [11], applied on the well-known collection of Stiegler's data [15], the gnostical estimator was found to be the best, since estimated values of the *known* location parameters of Stiegler's data sets, using (1.6), were the most accurate.

The first serious doubt of the validity of (**I**) was given in Fabián [1]. He noticed that the square of fidelity (1.2) 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 (ML) estimator or to α-estimators introduced by Vajda [18] in the case of this distribution. Based on this result, Vajda [19], [20] and Novovičová [13] studied properties of gnostical estimators. Apart from Kovanic's further attempts to consider only finite n-point "data varietes", they proved that gnostical estimators are regular statistical M-estimators, strongly consistent and asymptotically normal, and derived their asymptotic variances.

3 Problem statement

The success of the estimator (1.6) applied to the Stiegler data sets can be simply explained. The influence function of the robust estimator (1.6) is, contrary to usual robust estimators, non-symmetrical. This coincides with a clear non-symmetry of Stiegler's data. Nevertheless, some questions concerning gnostical theory remain unanswered. The incorrectness of (**I**) has not yet been rigorously proved and gnostical theory is still considered by some statisticians (see [16]) as a nonstatistical approach to uncerntainty. Can Kovanic's ideas (**II**) and (**III**) be proved incorrect, too ?

In the presented paper, we show that:

(i) There is a possible interpretation of Kolmogorov probability theory, in which description of continuous random variables exhibit some features similar to features of the gnostical theory.

(ii) Within the framework of this (geometric, say) probabilistic approach, basic notions of the gnostical theory appear to be probabilistic notions in cases of two special distributions.

(iii) Within these special data models, Kovanic's ideas (**II**) and (**III**) are right and gnostical estimation procedures, consisting in setting of irrelevances into "statistical formulas", are correct (if data follow one of these special models).

4 Geometric description of continuous random variables

R denotes real line, and $T \subset R$ a finite or infinite open interval. By $\mathcal{B}_T$ is denoted the σ-field of Borel subsets of T. Let $U : T \to R$ be a continuous random variable on $(T, \mathcal{B}_T)$ with distribution P. Let $P\{T\} = 1, P\{R - T\} = 0$, so that T is the support set of the distribution P. By F is denoted the distribution function $F(u) = P\{\zeta \in T : \zeta < u\}$ and by p the density $p(u) = dF(u)/du$ of random variable U. If the support is the whole real line, $T = T' = R$, we denote random variables by X and their distributions by P_R, distribution functions by F_R and densities by p_R.

Let $\varphi : T' \to T$ be a continuously differentiable homeomorphic mapping. Consider the random variable

$$U = \varphi(X). \tag{4.1}$$

The distribution function of U is given by

$$F(u) = F_R(\varphi^{-1}(u)) \tag{4.2}$$

and the density by

$$p(u) = \frac{dF(u)}{du} = \frac{dF_R(v)}{dv}|_{v=\varphi^{-1}(u)} \cdot \frac{d(\varphi^{-1}(u))}{du}. \tag{4.3}$$

Density of the random variable U consists of two terms,

$$p(u) = q(u)J(u). \tag{4.4}$$

The first one,

$$q(u) = p_R(\varphi^{-1}(u)) = \frac{dF_R(v)}{dv}|_{v=\varphi^{-1}(u)}, \tag{4.5}$$

is the "image" of the basic density p_R on R. It will be called the *proper density.* The second term,

$$J(u) = \frac{d(\varphi^{-1}(u))}{du}, \tag{4.6}$$

is of geometric origin. It is the Jacobian of the inverse mapping $\varphi^{-1} : T \to T'$ and does not actually depend on the basic density p_R, but only on T and φ. We call it the *geometric term.* Equation (4.4) gives the transformation relation of densities of random variables related by (4.1). Such random variables and their distributions will be called *φ-related.*

Let $\Theta \subset R^m$ be an open convex set. We denote by $\mathcal{P}_T = \{P_\theta | \theta \in \Theta\}$ a parametric family of distributions on $(T, \mathcal{B}_T)$, dominated by Lebesgue measure, with the parent P and with corresponding distribution functions F_θ and densities $p(u|\theta)$. Let $U_1, ..., U_n$ be observations, independent and identically distributed (i.i.d.) according to F_{θ^0} and let $\mathcal{T}_n(U_1, ..., U_n)$ be a functional $\mathcal{T}_n : T \to R$ such that $\mathcal{T}_n \to \mathcal{T}(F_{\theta^0})$ for $n \to \infty$.

Let $\mathcal{T}$ be an M-estimator, given by a suitable "$\psi-$function" ψ and by the equation

$$\sum_{j=1}^{n} \psi(X_j|\mathcal{T}_n) = 0.$$

Its influence function at $F = F_{\theta^0}$ is given by

$$IF(u; \mathcal{T}, F) = c^{-1}\psi(u|\mathcal{T}(F)),$$

where c is constant. It has been established in robust statistics (we refer to Hampel et all [5]) that the influence function of estimators determines the properties of estimates. Thus, estimates are sensitive or insensitive (robust) to outlying values in data if the influence function of the estimator is unbounded or bounded on T.

Consider the maximum likelihood (ML) estimator $\mathcal{T}$ in the location model $T = R, \Theta = R, p_R(x|\theta) = p_R(x - \theta)$. Its ψ-function is the partial score function $r_\theta(x)$,

$$\psi(x|\theta) = \psi(x - \theta) = \frac{\partial}{\partial\theta}(\ln p_R(x - \theta)) = r_\theta(x). \tag{4.7}$$

By the use of relations

$$r_\theta(x) = \frac{\partial}{\partial\theta}(\ln p_R(x - \theta)) = -\frac{d}{dx}(\ln p_R(x - \theta)) = -\frac{p_R'(x - \theta)}{p_R(x - \theta)} = h_R(x - \theta),$$

and by setting $\theta = 0$, one can realize that properties of ML estimates depend in fact of the score function

$$h_R(x) = -p_R'(x)/p_R(x) \tag{4.8}$$

(see [5], Example 1, pp. 104). Thus, the influence functions of ML estimators seems to be rather an entity connected with distributions than with ML estimators.

Such a point of view is, however, possible only in the case $T = R$. Score functions of distributions defined on Borel sets of $T \neq R$ are different from partial score functions of corresponding ML estimators. Consider, for instance, the lognormal distribution defined on $(R^+, \mathcal{B}_{R^+})$ where $R^+ = (0, \infty)$ by the density $p(z|z_0) = (\sqrt{2\pi} z)^{-1} \exp(-0.5 \ln^2(z/z_0))$. Its score function is for $z > 0$ given by $h_R(z|z_0) = z^{-1}(1+\ln(z/z_0))$, whereas the ψ-function (partial score function) of the lognormal ML estimator for parameter z_0 is $r_{z_0}(z|z_0) = z_0^{-1} \ln(z/z_0)$.

A generalization of the concept of the score function into a function expressing properties of ML estimates of distributions defined on arbitrary T has been obtained by Fabián ([2]-[4]). We show here that it will do for it to require the "natural transformation law" (identical to (4.2)) for score functions of distributions defined on R.

Definition 1 *Let a mapping $\varphi : T' = R \to T$ be continuously differentiable and homeomorphic. Let the score function h_R of a distribution P_R defined on $(T', \mathcal{B}_R)$ exists. A real-valued function $h : T \to R$, given by transformation*

$$h(u) = h_R(\varphi^{-1}(u)), \tag{4.9}$$

*will be called the influence function (*IFD*) of the distribution P .*

Explicit form of an IFD on arbitrary T is given by Theorem 1.

Theorem 1 *Influence function of an absolute continuous distribution P on arbitrary $(T, \mathcal{B}_T)$ is given by the relation*

$$h(u) = \frac{1}{p(u)} \frac{d}{du}(-J^{-1}(u)p(u)), \tag{4.10}$$

where J is given by (4.6).

Proof Denote $v = \varphi^{-1}(u)$. Using (4.9) and (4.8)

$$h(u) = \frac{1}{p_R(v)} \frac{d}{dv}(-p_R(v)) = \frac{1}{p_R(v)} \frac{d}{du}(-p_R(v)) \frac{du}{dv}.$$

Substituting $p_R(v) = p(u)\ J(u)^{-1}$ from (4.4) and using (4.6),

$$h(u) = \frac{J(u)}{p(u)} \frac{d}{du}(-J^{-1}(u)p(u)) \cdot J^{-1}(u) = \frac{1}{p(u)} \frac{d}{du}(-J^{-1}(u)p(u)).$$

□

Specifying general relations for the case $T = R^+ = (0, \infty)$, we denote $z = u \in R^+$ and $Z = U$ on R^+. Let $Z = \varphi(X)$, and choose the simplest possible mapping $\varphi : R \to R^+$, by

$$z = \varphi(x) = e^x. \tag{4.11}$$

Then

$$x = \varphi^{-1}(z) = \ln z, \qquad J(z) = 1/z. \tag{4.12}$$

Using (4.11), (4.12), IFD (4.10) on R^+ is expressed by

$$h(z) = -1 - zp'(z)/p(z). \tag{4.13}$$

Knowing the score function h_R of a distribution P_R on $R, \mathcal{B}_R$, one can compute the corresponding density by a formula inverse to (4.8)

$$p_R(x) = \frac{1}{c_x} \exp\,(-\int h_R(x)\,dx) \tag{4.14}$$

where c_x is a norming constant (to be computed). In the case of an influence function $h(z)$ on R^+, the formula inverse to (4.13) is

$$p(z) = \frac{1}{c_z} \exp\,(-\int z^{-1}[1 + h(z)]\,dz). \tag{4.15}$$

Since p_R is assumed continuously differentiable, p is continuously differentiable, too, and relations (4.8), (4.14) and (4.13), (4.15) represent one-to-one correspondences between densities and IFD's of continuous probability distributions.

Example IFD's and densities of some couples of e^x-related distributions on R and R^+ are given in Table 1. Tabeled influence functions are plotted on fig.1.

TABLE 1 *IFD's $h_R(x)$ and densities $p_R(x)$ of some distributions on R and IFD's $h(z)$ and densities $p(z)$ of e^x-related distributions on 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}
$\mathrm{tgh}(x/2)$	$\frac{1}{4}\cosh^{-2}(x/2)$	$(z-1)/(z+1)$	$1/(z+1)^2$
$\sinh x$	$\frac{1}{2K_0(1)} e^{-\cosh x}$	$\frac{1}{2}(z - 1/z)$	$\frac{1}{2K_0(1)\,z} e^{-\frac{1}{2}(z+1/z)}$

where K is the MacDonald function (Bessel function of the III. kind). In the first three rows of Table 1 are standardized forms of couples of e^x-related distributions: normal and lognormal, double exponential and exponential, logistic and log-logistic (the distribution on R in the fourth row is not encountered in statistics, the related one on R^+ is Wald-type).

In what follows, we will consider that the influence functions of distributions on R^+ are defined with respect to the mapping φ given by (4.11).

5 Influence function of a parametric distribution

Generalization of the influence function of a distribution (4.10) for the case of a parametric family $\mathcal{P}_T$ is straightforward: a parametric set of IFD's is $\{h(u|\theta)|\theta \in \Theta\}$, where

$$h(u|\theta) = -\frac{1}{p(u|\theta)}\frac{d}{du}(-J^{-1}(u)p(u|\theta)). \tag{5.1}$$

Consider the simplest case of a parametric family with one location parameter. In the case of distributions defined on R, the location parameter $x_0 \in R$ means the shift of the mode of the parent distribution P_R along the real line, and corresponding densities are $p(x|x_0) = p_R(x - x_0)$.

The notion of the location parameter will be generalized for cases of distributions defined on T as follows.

Definition 2 *Let x_0 be the location parameter of a distribution P_R. The location parameter u_0 of a φ-related distribution P on $(T, \mathcal{B}_T)$ is given by the relation*

$$u_0 = \varphi(x_0).$$

The form of densities of location type on R^+ is therefore

$$p(z|z_0) = z^{-1}p_R(\ln z - \ln z_0) = \frac{1}{z}p_R(\ln\frac{z}{z_0}) \tag{5.2}$$

and corresponding IFD's are, by (4.9),

$$h(z|z_0) = h_R(\ln (z/z_0)).$$

We now show that, in the case of a distribution with a (generalized) location parameter, its IFD is proportional to the partial score function for the location parameter.

Theorem 2 *Let u_0 be the location parameter of a parametric family $\{P_{u_0}|u_0 \in \Theta\}$ on $(T, \mathcal{B}_T)$. Let the partial score function $r_1(u|u_0)$ for u_0 exists. Then*

$$h(u|u_0) = J^{-1}(u_0)r_1(u|u_0).$$

Proof Denote $x = \varphi^{-1}(u), x_0 = \varphi^{-1}(u_0)$. Using (4.4), (4.6), Theorem 1 and (4.9),

$$r_1(u|u_0) = \frac{d}{du_0}(\ln p(u|u_0)) = \frac{1}{p(u|u_0)}\frac{d}{du_0}p(u|u_0) =$$

$$= \frac{1}{J(u)p_R(x - x_0)}\cdot\frac{d}{dx_0}(J(u)p_R(x-x_0))\frac{d(\varphi^{-1}(u_0))}{du_0} = h_R(x-x_0)J(u_0) = J(u_0)h(u|u_0).$$

□

According to the Theorem, and referring to the meaning of the ML scores function in statistics, we suppose that a general sense of the influence function of a distribution is the following: $h(u)$ is the value which is to enter into statistical inference mechanisms. In other words, it is the "inference value" of $u \in T$ under the assumption of the distribution P_θ. We do not discusse this question in more details here. Some further results supporting this idea can be found in [4].

Consider the location and scale model on R, given by a parametric family $\mathcal{P}_R = \{P_{x_0,\sigma}|x_0 \in R, \sigma \in R^+\}$ on $(R, \mathcal{B}_R)$, where $P_{x_0,\sigma} = (P_R)_{x_0,\sigma}$, with densities in the form $p(x|x_0,\sigma) = \sigma^{-1}p_R((x-x_0)/\sigma)$. Using (5.2) we obtain densities of related parametric distributions $\mathcal{P}_{R+} = \{P_{z_0,\sigma}|z_0 \in R^+, \sigma \in R^+\}$ on $(R^+, \mathcal{B}_{R+})$ in the form

$$p(z|z_0,\sigma) = (\sigma z)^{-1}\ p_R(\sigma^{-1}\ln\ (\frac{z}{z_0})) = (\sigma z)^{-1}\ p_R(\ln(\frac{z}{z_0})^{1/\sigma}). \tag{5.3}$$

Corresponding IFD's are

$$h(z|z_0,\sigma) = \sigma^{-1}h_R(\ln(z/z_0)^{1/\sigma}). \tag{5.4}$$

p_R is usually called the parent density of the family $\mathcal{P}_{\mathcal{R}}$. h_R will be called the parent IFD of the family. It is apparent from (5.3), (5.4) that they can be taken as parents of the location and scale family $\mathcal{P}_{R+}$, too.

6 Weight function of a parametric distribution

Definition 3 *Let h be the influence function of a distribution $P_\theta \in \mathcal{P}_T$, defined on $(T, \mathcal{B}_T)$, continuous and strictly increasing on T. A real-valued function $g : T \to R$, given by*

$$g(u|\theta) = dh(u|\theta)/du \tag{6.1}$$

will be called the weight function of the distribution P_θ.

Motivation for this definition is as follows. By means of IFD's, a distance in the sample space T can be introduced by the formula

$$\rho(u_1,u_2|\theta) = |h(u_2|\theta) - h(u_1|\theta)| \qquad u_1,u_2 \in \Theta. \tag{6.2}$$

By Theorem 2, (6.2) is the distance of "ML values for location" of two points of the sample space. It appears to be a metric for a continuous, strictly increasing h. Expression (6.2) can be rewritten in such a case into $d\rho(u|\theta) = g(u|\theta)du$, where g is given by (6.1). The space (T, g) is apparently the one-dimensional Riemannian metric space. In Riemannian geometry (we refer to Kobayashi, Nomizu [6]), g represents the weight introduced in the space T.

From the direct differentiation of (5.1) and the use of (4.6) follows transformation relation for the weight function in the form

$$g(u|\theta) = J(u)g_R(\varphi^{-1}(u|\theta)). \tag{6.3}$$

Likewise in the case of densities on R^+, the term $w(u|\theta) = g_R(\varphi^{-1}(u|\theta))$ will be called the *proper weight function* of random variable U.

Consider now for the sake of simplicity a distribution without parameters with a density $p(u)$ and IFD $h(u)$, so that the weight function is given by $g(u) = dh(u)/du$.

Taking derivatives of (4.8), (4.13) with respect to x and z, respectively, we obtain weight functions on R and R^+, expressed by densities as

$$g_R(x) = \left(\frac{p_R'(x)}{p_R(x)}\right)^2 - \frac{p_R''(x)}{p_R(x)} \tag{6.4}$$

$$g(z) = -\frac{p'(z)}{p(z)} + z\left[\left(\frac{p'(z)}{p(z)}\right)^2 - \frac{p''(z)}{p(z)}\right]. \tag{6.5}$$

Example Proper densities $q(z)$ (4.5) and proper weight functions $w(z) = zg(z)$ of distributions defined on R^+ and quoted in Table 1 are given in Table 2.

TABLE 2 *Densities $p_R(x)$, weight functions $g_R(x)$ of distributions on R from Table 1 and proper densities $q(z)$, weight functions $g(z)$ and proper weight functions $w(z)$ of related distributions on R^+*

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

We suppose that proper weight function $w(u|\theta)$ could represent a relative weight (relative importance) of a point $u \in T$ of the sample space (of the observed value u) under the assumption of the distribution P_θ.

Let g_R be the weight function of the distribution P_R. Weight function of the location and scale model $\mathcal{P}_{R^+}$ is, using (5.4) and (6.1)

$$g(z|z_0, \sigma) = \sigma^{-1} dh_R(\ln(z/z_0)^{1/\sigma})/dz = \sigma^{-2}z^{-1}g_R(\ln(z/z_0)^{1/\sigma}). \tag{6.6}$$

7 Parametric form of a continuous random variable and Axiom 1 of the gnostical theory

Definition 4 *Let U be a random variable defined on $(T, \mathcal{B}_T)$ with distribution $P_\theta \in \mathcal{P}_T$ and let there exists a continuous and strictly increasing on T influence function h of the distribution P_θ. Let Ω be a random variable defined by the relation*

$$\Omega = h(U).$$

The expression

$$U = h^{-1}(\Omega), \tag{7.1}$$

will be called parametric form of random variable U.

Theorem 3 *Kovanic's Axiom 1* (1) *represents the parametric form of a random variable with lognormal distribution.*

Proof Density of the lognormal distribution is given by

$$p_l(z|z_0,\sigma) = \frac{1}{\sqrt{2\pi}\sigma z} e^{-\frac{1}{2}\ln^2[(z/z_0)^{1/\sigma}]}.$$

The corresponding IFD is

$$h_l(z|z_0,\sigma) = \frac{1}{p_l(z|z_0,\sigma)} \frac{d}{dz}(-zp_l(z|z_0,\sigma)) = \sigma^{-1}\ln(z/z_0)^{1/\sigma}.$$

Setting $\Omega = h_l(Z|z_0,\sigma)$, one obtains

$$Z = z_0 e^{s\Omega}, \tag{7.2}$$

where $s = \sigma^2$. The assertion then follows from the one-to-one correspondence of densities and IFD's (4.13), (4.15). □

Remark A general parametric form of (7.2) for the location and scale model on $(R^+, \mathcal{B}_{R^+})$ with parent IFD h_R is, using (5.4),

$$Z = z_0 e^{\sigma h_R^{-1}(\sigma\Omega)}. \tag{7.3}$$

(7.3) represents a generalized form of the "gnostical Axiom 1".

8 Probability densities implicitly assumed in the gnostical theory

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

$$h_q(z|z_0,s) = \sinh(2\Omega) = \frac{1}{2}[(z/z_0)^{2/s} - (z/z_0)^{-2/s}] \tag{8.1}$$

and providing sensitive gnostical estimates.

Theorem 4 *Probability densities corresponding to two types of Kovanic's irrelevances* (1.3), (8.1) *are*

$$p_1(z|z_0,s) = \frac{2\sqrt{2\pi}}{zs\Gamma^2(1/4)} \frac{1}{[(z/z_0)^{2/s} + (z/z_0)^{-2/s}]^{1/2}} \tag{8.2}$$

$$p_2(z|z_0,s) = \frac{1}{zsK_0(1/2)} e^{-\frac{1}{4}[(z/z_0)^{2/s}+(z/z_0)^{-2/s}]}, \tag{8.3}$$

respectively.

Proof Let

$$h_{R1}(u) = \mathrm{tgh}(2u), \qquad h_{R2}(u) = \sinh(2u) \tag{8.4}$$

be score functions of some distributions. The corresponding densities are given by (4.14), so that

$$p_1(u) = c_1^{-1} e^{-\int \mathrm{tgh}(2u)\, du} = c_1^{-1} \cosh^{-1/2}(2u) \tag{8.5}$$

$$p_2(u) = c_2^{-1} e^{-\int \sinh(2u)\, du} = c_2^{-1} e^{-\frac{1}{2}\cosh(2u)}. \tag{8.6}$$

By the use of integrals ([14]),

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

where Γ is the gamma function and K_α the modified Bessel function of the 3. kind, norming constants are

$$c_1 = \Gamma^2(1/4)/2\sqrt{2\pi}, \qquad c_2 = K_0(1/2).$$

By the substitution

$$u = \ln(z/z_0)^{1/s} \tag{8.7}$$

in (8.4) and using (5.4), one obtains influence functions of searched distributions in the form

$$h_1(z|z_0, s) = s^{-1} \mathrm{tgh}(\ln(z/z_0)^{2/s}) = -s^{-1} h_e(z|z_0, s) \tag{8.8}$$

$$h_2(z|z_0, s) = s^{-1} \sinh(\ln(z/z_0)^{2/s}) = s^{-1} h_q(z|z_0, s), \tag{8.9}$$

where $-h_e$ and h_q are gnostical irrelevances given by (1.3) and (8.1). Substituting (8.7) into (8.5) and (8.6) and using transformation relations (5.3), one obtains the searched densities in the form (8.2) and (8.3).

Since 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, and considering the one-to-one correspondence of IFD's and densities, assertion holds. □

Remark 1 "Gnostical probability densities" (8.2), (8.3) differ from that introduced by Axiom 1 due to an inconsistency in Kovanic's derivation of gnostical data characteristics (1.2), (1.3) from (1.1). In the step, in which the two components of random variable $e^{s\Omega} = \cosh(s\Omega) + \sinh(s\Omega)$ are considered to be independent (Kovanic's "dissimilarity laws"), is in fact redefined the original distribution.

Remark 2 Considering one of the probability models (8.2), (8.3), Kovanic's idea (**II**) is true. Indeed, given a model of a statistical experiment in the form of a parametric set $\mathcal{P}_T$, the data, realizations of i.i.d. random variables $U_1, \dots U_n$ 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 distribution, ("inference value" or, possibly, "relevance") $h(u_i|\theta)$, and the proper weight $w(u_i|\theta)$. They are in a "latent" form because of an unknown θ^0, similarly as with the likelihood. They can be approximately determined after an estimate $\hat{\theta}$ of the true value θ^0 is found. Thus, Kovanic's irrelevances (in the case of "estimating irrelevance" the function negative to it) appears to be special cases (in special models (8.2), (8.3)) of a basic probabilistic characteristic, surprisingly not known in a general form in probability theory untill introduced by Fabián.

Kovanic's idea (**III**) is then simply realized by defining the distance in the sample space by the relation (6.2). This is a Riemannian distance, being Euclidean only in cases of special distributions (for instance the normal distribution on R with IFD $h_R(x|x_0, \sigma) = \sigma^{-2}(x - x_0)$).

Theorem 5 *Square of the gnostical fidelity is, apart from the constant, the proper weight function in the probabilistic model given by the density family* (8.2).

Proof Weight functions of distributions with densities (8.5), (8.6) are, using (6.1) and (8.4)

$$g_1(u) = 2\cosh^{-2}(2u), \qquad g_2(u) = 2\cosh(2u). \tag{8.10}$$

After substitution (8.7) and by the use of (6.6),

$$g_1(z|z_0, s) = 2s^{-2}z^{-1}\cosh^{-2}(\ln(z/z_0)^{2/s}) = 2s^{-2}z^{-1}f^2(z|z_0, s) \tag{8.11}$$

$$g_2(z|z_0, s) = 2s^{-2}z^{-1}\cosh(\ln(z/z_0)^{2/s}) = 2s^{-2}z^{-1}f^{-1}(z|z_0, s), \tag{8.12}$$

where f is the fidelity (1.2). Thus, f^2 is, apart from the constant factor (Kovanic is aware of the relation (6.1)), the proper weight function $w_1(z|z_0, s) = zg_1(z|z_0, s)$ of the distribution (8.2) (and, similarly, $f^{-1}(z|z_0, s)$ the proper weight function of the distribution (8.3)). □

9 Conclusion

In this paper we explained Kovanic's "non-statistical" notions of irrelevance and fidelity of individual data. We did this in a rather unexpected fashion: by including their general equivalents into Kolmogorov probability theory.

We conjecture that a significance of the gnostical theory consists not in some of its special procedures, but in the fact, that it was the *first virtually probabilistic and statistical theory*, which was constructed on a sample space different from the whole real line. Since gnostical theory, likewise with the classical probability theory, does not itself realize the existence of the geometric term (4.6), the "bridge" between different sample spaces, gnostical theory naturally deals with influence functions of distribution on R^+ (as probability theory does likewise in cases of distributions on R) and with proper weight functions.

According to Theorem 2 of Section 5, the requirement (1.6) of zero average (ir)relevance of a data sample (generally: zero average "sample influence function of distribution") affords the ML estimate of the location parameter *without knowing the maximum likelihood principle* (e.g. without need of differentiation with respect to the location parameter). To this fact can be attributed the success of "gnostical estimator" (1.6) of the location parameter, as well as the source of difficulties with "gnostical" estimation of the scale parameter (which can be removed by the use of the IFD-moment method, see Fabián (1994b)).

It should be noted that we did not explain Kovanic's estimation procedures based on his "Axiom 2 of the gnostical theory". We suppose that, in probabilistic terms, the composition law (1.5) can be considered to be a "finite equivalent" of some limit theorem concerning sums of *weighted* i.i.d. random variables. "Qualitatively", (1.5) 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] Fabián, Z. (1988). Point estimation in case of small data sets. *Trans. 10th Prague Conf. on Information Theory*, held at Prague 1986, Academia, Prague, 305-312.

[2] Fabián, Z. (1993). Metric random variable. Techn. rep. ICS ASCR, V-552, Prague.

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

[4] Fabián, Z. (1994b). Generalized score function and its use. *Trans. 12th Prague Conf. on Information Theory*, Academia, Prague.

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

[6] Kobayashi, S., Nomizu, K. (1963). *Foundations of differential geometry.* Intersci Publishers, New York, London.

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

[8] Kovanic, P. (1984b). Gnostical theory of small samples of real data. *PCIT* **13(5)**, 303-319.

[9] Kovanic, P. (1984c). On relation between information and physics. *PCIT* **13(6)**, 383-399.

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

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

[12] Loève, M. (1977). *Probability theory.* 4th ed., Springer, Berlin, New York.

[13] Novovičová, J. (1989). M-estimators and gnostical estimators of location. *PCIT* **18(6)**, 397-407.

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

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

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

[17] Štěpán, J. (1987). *Probability theory* (in Czech). Academia, Praha.

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

[19] Vajda, I. (1987). Minimum-distance and gnostical estimators. *PCIT* **17(5)**, 253-266.

[20] Vajda, I. (1989). Comparison of asymptotic variances for several estimators of location. *PCIT* **18(2)**, 79-87.