KYBERNETIKA — VOLUME 33 (1997), NUMBER 3, PAGES 259 – 270

ON THE RELATION BETWEEN GNOSTICAL AND PROBABILITY THEORIES

Zdeněk Fabián

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 introduced in [10]. Gnostical model of uncertainty, claimed by its author to be independent of probabilistic concepts in [12 – 13], can be thus replaced by a special case of the classical probabilistic model.

1. 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 be a T-valued random variable with distribution P_T, distribution function F_T and density p_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 an R-valued random variable U_R:

$$h_R(x) = \frac{\mathrm{d}}{\mathrm{d}x}(-\log p_R(x)) = -\frac{p_R'(x)}{p_R(x)}. \tag{1}$$

It is known 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 (1) for T-valued random variable, where $T \neq R$, has been proposed in [4]. It has been assumed that the set Π_T is an image of the set Π_R under a given diffeomorphism $\varphi : R \to T$. Then, any U_T on $(T, \mathcal{B}_T)$ has a unique "prototype" U_R on $(R, \mathcal{B}_R)$ given by $U_R = \varphi^{-1}(U_T)$. Such U_T and U_R and their distributions we call *φ-related*. The relation between their distribution functions is obviously

$$F_T(u) = F_R(\varphi^{-1}(u)). \tag{2}$$

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 of the score function of its prototype under the mapping φ.

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

$$h_T(u) = h_R(\varphi^{-1}(u)), \tag{3}$$

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

An explicit form of the IFD is given by the following proposition.

Proposition 1. The IFD of random variable U_T specified in Definition 1 is given by

$$h_T(u) = \frac{1}{p_T(u)} \frac{\mathrm{d}}{\mathrm{d}u}(-L(u)\, p_T(u)), \tag{4}$$

where

$$L(u) = \left(\frac{\mathrm{d}(\varphi(x))}{\mathrm{d}x}\right)_{x=\varphi^{-1}(u)}. \tag{5}$$

Proof. Denote $v = \varphi^{-1}(u)$. According to (2), the density of U_T is

$$p_T(u) = \frac{\mathrm{d}F_T(u)}{\mathrm{d}u} = \frac{\mathrm{d}F_R(v)}{\mathrm{d}v}\frac{\mathrm{d}v}{\mathrm{d}u} = \frac{p_R(v)}{L(u)} \tag{6}$$

by the formula for the inverse function derivative. By (3) and (4)

$$h_T(u) = h_R(v) = \frac{1}{p_R(v)} \frac{\mathrm{d}}{\mathrm{d}v}(-p_R(v)) = \frac{1}{L(u)\, p_T(u)} \frac{\mathrm{d}}{\mathrm{d}u}(-L(u)\, p_T(u)) \cdot L(u). \qquad \square$$

The relation inverse to (4) is

$$p_T(u) = c^{-1} \exp\left(-\int L^{-1}(u)\,[h_T(u) + L'(u)]\,\mathrm{d}u\right) \tag{7}$$

(supposing that $c = \int_T p_T(u)\,\mathrm{d}u$ exists).

Let us consider the *halfline model*, where $T = R^+ = (0, \infty)$ and $\varphi : R \to T$ is the exponential function e^x. In this model $Z = U_{R^+}$ is related to its prototype $X = U_R$ by the formula

$$Z = \varphi(X) = e^X,$$

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

$$L(z) = z \tag{8}$$

and that the IFD of Z is given by

$$h(z) = -1 - zp'(z)/p(z). \tag{9}$$

The particular mapping φ proposed for $T = R^+$ has a statistical motivation. Namely, positive data are often logarithmically transformed and there are well-known "logarithmically related" pairs of distributions on R^+ and R (the lognormal and normal, the log-Cauchy and Cauchy etc.).

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

$$h_T(u|\theta) = \frac{1}{p_T(u|\theta)} \frac{\mathrm{d}}{\mathrm{d}u}(-L(u)\, p_T(u|\theta)), \qquad \theta \in \Theta, \tag{10}$$

where $L(u)$ is given by (5).

Consider now the *location and scale model,* where $T = R$ and $\varphi(x) = \sigma x + x_0$. Then it follows from Proposition 1 that $p_T(x|x_0, \sigma) = \sigma^{-1}\tilde{p}((x - x_0)/\sigma)$ where $\tilde{p} = p_R$ is the parent prototype density and $x_0 \in R$ and $\sigma \in R^+$ are location and scale parameters. The score function is, by (1), $h_R(x|x_0, \sigma) = \sigma^{-1}\tilde{h}((x - x_0)/\sigma)$ where $\tilde{h}$ is the "prototype" score function.

If $p_R(x|\theta) = p_R(x|x_0, \theta_2, \ldots, \theta_m)$ then we define the *transformed location parameter* $u_0 \in T$ of any φ-related distribution P_T on $(T, \mathcal{B}_T)$ by the formula $u_0 = \varphi(x_0)$. For example, in the halfline model the transformed location parameter is $z_0 = e^{x_0}$. Since in this case $L(z)$ is given by (8), it follows from (6) that the densities in the exponentially related transformed location and scale model are given by the formula

$$p_{R^+}(z|z_0, \sigma) = p(z|z_0, \sigma) = (z\sigma)^{-1}\tilde{p}(\sigma^{-1}(\ln z \; - \ln z_0)) = (z\sigma)^{-1}\tilde{p}(\ln(z/z_0)^{1/\sigma}). \tag{11}$$

According to (4), the corresponding IFDs are

$$h_{R^+}(z|z_0, \sigma) = h(z|z_0, \sigma) = \sigma^{-1}\tilde{h}(\ln\,(z/z_0)^{1/\sigma}). \tag{12}$$

2. PROPERTIES OF IFD s

We list properties of IFDs. Some of them were discussed in more details in [5] – [7].

i) *IFD represents an equivalent description of the distribution, which is often simpler than the density.*

Due to assumptions, relation (4) and its inverse (7) represent a one-to-one correspondence between the density and the IFD of a continuous probability distribution. The simplicity of IFDs 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	$\frac{1}{\sqrt{2\pi}}\, e^{-\frac{1}{2}x^2}$	$\ln z$	$\frac{1}{\sqrt{2\pi}\, z}\, e^{-\frac{1}{2}\ln^2 z}$
$e^x - 1$	$e^x e^{-e^x}$	$z - 1$	e^{-z}
$\mathrm{tgh}(x/2)$	$\frac{1}{4}\cosh^{-2}(x/2)$	$(z-1)/(z+1)$	$1/(z+1)^2$
$\sinh x$	$\frac{1}{2K_0(1)}\, e^{-\cosh x}$	$\frac{1}{2}(z - 1/z)$	$\frac{1}{2K_0(1)\, z}\, e^{-\frac{1}{2}(z+1/z)}$

In the first three rows of Table 1 are standardized forms of pairs of exponentially related distributions: normal and lognormal, double exponential and exponential, logistic and log-logistic. In the fourth row K_0 is the Bessel function of the third kind. The distribution with density p_R in the fourth row seems to be new – it is logarithmically related to the standardized form $p(z)$ of the Wald distribution (see [16]) in this row.

ii) *If the vector of parameters of a distribution contains the transformed location parameter, IFD is proportional to the likelihood score for this parameter.*

Recall that the likelihood scores are defined as

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

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

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

P r o o f. Let $P_{T_\theta} = \varphi(P_{R_{\theta'}})$ where $\theta' = (\varphi^{-1}(u_0), \alpha)$ and denote $v = \varphi^{-1}(u) - \varphi^{-1}(u_0)$. Analogically to (6), $p_T(u|\theta) = L^{-1}(u)\, p_R(v|\alpha)$. Then

$$\begin{aligned} r_1(u|\theta) &= \frac{1}{p_T(u|\theta)}\frac{\partial p_T(u|\theta)}{\partial u_0} = \frac{L(u)}{p_R(v|\alpha)}\frac{\partial(L^{-1}(u)p_R(v|\alpha))}{\partial v}\frac{\mathrm{d}v}{\mathrm{d}u_0} \\ &= -\frac{{p_R}'(v|\alpha)}{p_R(v|\alpha)}\, L^{-1}(u_0) = h_R(v|\alpha)L^{-1}(u_0) = L^{-1}(u_0)\, h_T(u|\theta). \end{aligned}$$ □

iii) *The IFD-moments sometimes better numerically characterize 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 kth IFD-moment of random variable U_T has been defined in [4] by the integral

$$M_k(\theta) = \int_T h_T^k(u|\theta)\, p_T(u|\theta)\, \mathrm{d}u. \tag{13}$$

It has been proved in [7] that under the considered conditions the IFD-moments exist, even for distributions with non-existing usual moments (Cauchy and log-logistic distributions, for instance).

Let $c_1 = \inf\{u : u \in T\}, c_2 = \sup\{u : u \in T\}$. By (4) and (6),

$$M_1 = \int_T h_T(u)\, p_T(u)\, \mathrm{d}u = -L(u)\, p_T(u)|_{-c_1}^{c_2} = -p_R(x)|_{-\infty}^{\infty} = 0. \qquad (14)$$

(14) obviously holds in the parametric case, too. All the other IFD-moments (13) are expressed by means of parameters only and not by functions of parameters, which is typical for the usual moments. This is important when estimates $\hat{\theta}$ of the true parameter θ^0 are defined as solutions of the equations

$$n^{-1} \sum_{i=1}^{n} h_T^k(u_i|\hat{\theta}) = M_k(\hat{\theta}), \qquad k = 1, \ldots, m, \qquad (15)$$

where $u_1, \ldots, u_n$ are observed values of independent, identically distributed (i.i.d.) random variables with distribution P_{θ^0}. In the halfline model the equations (15) take on according to (14) and (12) the form

$$n^{-1} \sum_{i=1}^{n} \tilde{h}(\ln\, (z_i/\hat{z}_0)^{1/\hat{\sigma}}) = 0, \qquad (16)$$

$$n^{-1} \sum_{i=1}^{n} \tilde{h}^2(\ln\, (z_i/\hat{z}_0)^{1/\hat{\sigma}}) = \hat{\sigma}^2 M_2(\hat{z}_0, \hat{\sigma}). \qquad (17)$$

According to Proposition 2, the first moment equation (16) is identical with the maximum likelihood equation for the transformed location parameter. It has been shown in [6] that if estimates (15) exist (e. g. when $\tilde{h}$ is monotonous) then they are consistent and asymptotically normal. Moreover, in cases of distributions with bounded IFDs, the asymptotic variances of estimates (15) are near to the Cramér–Rao bound. Simultaneously, IFD-moment estimates of both location and scale parameters are robust, whereas the ML estimates of the scale parameter are known to be sensitive to the outliers.

iv) *The second IFD-moment is proportional to the Fisher information of a distribution.*

The Fisher information is usually defined and interpreted in parametric models. The non-parametric Fisher information (the Fisher information of the distribution) is defined as mean value of the score function (e. g. [1, p. 494]). An alternative definition of the Fisher information of the distribution has been proposed in [5], namely

$$M_2 = \int_T h_T^2(u)\; p_T(u)\, \mathrm{d}u.$$

M_2 is defined for distributions on arbitrary $(T, \mathcal{B}_T)$ even in parametric models, where $M_2(\theta) = \int_T h_T^2(u|\theta)\; p_T(u|\theta)\, \mathrm{d}u$ is finite for all Cramér–Rao regular distributions. The advantage of the definition using IFD is that, according to Proposition 2, $M_2(\theta)$ in the transformed location model is proportional to the classical Fisher information $FI(u_0)$, $M_2(u_0) = L^{-1}(u_0)\, FI(u_0)$ even when $T \neq R$.

3. DERIVATIVE OF THE IFD

By means of the IFD, a reasonable distance of points $u_1, u_2 \in T$ can be introduced by the formula

$$\rho(u_1, u_2|\theta) = |h_T(u_2|\theta) - h_T(u_1|\theta)| = \int_{u_1}^{u_2} g_T(u|\theta)\, \mathrm{d}u \tag{18}$$

where

$$g_T(u|\theta) = \mathrm{d}h_T(u|\theta)/\mathrm{d}u. \tag{19}$$

By Proposition 2, (18) is proportional to a distance introduced in the sample space by the likelihood function of transformed location parameter. If h_T is continuous and strictly increasing, (18) is a metric. The space (T, ρ) is in such a case a one-dimensional Riemannian metric space.

Let g_R be a derivative of the score function of a distribution $P_R \in \Pi_R$. It follows from the direct differentiation of (10) and Proposition 1 that the derivative of the IFD of the $\varphi-$related distribution on $(T, \mathcal{B}_T)$ is given by

$$g_T(u|\theta) = L^{-1}(u)\, g_R(\varphi^{-1}(u|\theta)). \tag{20}$$

In the spirit of the Riemannian geometry, the term

$$w_T(u|\theta) = g_R(\varphi^{-1}(u|\theta)) \tag{21}$$

may be called a *weight function* of the distribution P_{T_θ} (WFD). It represents a relative importance of an observed point $u \in T$ under the assumption that the distribution is P_{T_θ}.

Consider for simplicity a distribution without parameters, with density $p_T(u)$ and IFD $h_T(u)$, so that $g_T(u) = \mathrm{d}h_T(u)/\mathrm{d}u$. Taking derivatives of (1) and (9), (20) are on $(R, \mathcal{B}_R)$ and $(R^+, \mathcal{B}_{R^+})$ expressed by densities as

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

Weight functions of distributions from Table 1 are given in Table 2.

Table 2. Weight functions g_R of distributions P_R and $w(z) = zg(z)$ of the exponentially related distributions.

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

Finally, using (12) and (20), it holds in the case of the transformed location and scale model on $(R^+, \mathcal{B}_{R^+})$

$$g_{R^+}(z|z_0,\sigma) = g(z|z_0,\sigma) = \sigma^{-1}\mathrm{d}\tilde{h}(\ln(z/z_0)^{1/\sigma})/\mathrm{d}z = \sigma^{-2}z^{-1}\tilde{g}(\ln(z/z_0)^{1/\sigma}) \quad (22)$$

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

4. GNOSTICAL THEORY

A nonstandard theory of data processing 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 was proposed by the author as non-probabilistic.

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

$$z = z_0 e^{s\Omega} \quad (23)$$

where $z_0 \in R^+$ is an "ideal value" of z and $\Omega \in R$ the uncertainty, scaled (in [13]) by a parameter $s \in R^+$. Since (23) seems to be a general parametric model of positive data items and any real measured data are in fact positive, Kovanic considered that (23) is a universal mathematical model of data "suffering from uncertainty". Based on this model, he derived two individual "gnostical" data characteristics that depend on the uncertainty. They are "fidelity", given by the expression

$$f(z|z_0,s) = \cosh^{-1}(2\Omega) = 2/\left[(z/z_0)^{2/s} + (z/z_0)^{-2/s}\right], \quad (24)$$

and "irrelevance", given by

$$h_e(z|z_0,s) = -\mathrm{tgh}(2\Omega) = -\frac{(z/z_0)^{2/s} - (z/z_0)^{-2/s}}{(z/z_0)^{2/s} + (z/z_0)^{-2/s}}. \quad (25)$$

These are the two basic gnostical characteristics of one data item when the model (23) is known, mutually related by

$$h_e^2(z|z_0,s) = 1 - f^2(z|z_0,s).$$

Having a sample $\boldsymbol{Z}_n = (z_1,\ldots,z_n)$ of data from one source (23), each data item z_i can be characterized, after Kovanic, by its fidelity and irrelevance. They are in a latent form because of the unknown parameters z_0, s which can, however, be estimated from the data sample $\boldsymbol{Z}_n$. The simplest gnostical estimate of the ideal value z_0 is obtained by Kovanic's requirement of zero average irrelevance of the sample $\boldsymbol{Z}_n$. This gives the estimation equation

$$n^{-1}\sum_{i=1}^{n} h_e(z_i|\hat{z}_0,\hat{s}_a) = 0, \quad (26)$$

where $\hat{s}_a$ is a prior estimate of the scale parameter s. The function h_e is bounded, $|h_e(z|z_0)), s)| \leq 1$. A consequence of this fact is the insensitivity of estimates (26) to outlying values in data, without introducing any of the robustifying functions of robust statistics. The fact that the gnostical estimator (26) 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 [18]. The gnostical estimator, giving quite realistic estimates, was found in [14] to achieve 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

$$C_e(k) = \frac{1}{n-k} \sum_{i=1}^{n-k} h_e(z_i|z_0, s)\, h_e(z_{i+k}|z_0, s).$$

In some of procedures Kovanic uses the square of fidelity as the weight of data. 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 $\boldsymbol{Z}_n$ is given by

$$h_e(z_c|z_0, s) = \sum_{i=1}^{n} h_e(z_i|z_0, s)/w_e, \tag{27}$$

where

$$w_e = \left(\left[\sum_{i=1}^{n} f(z_j|z_0, s) \right]^2 + \left[\sum_{i=1}^{n} h_e(z_j|z_0, s) \right]^2 \right)^{1/2},$$

i.e. that the irrelevance of the composite event is the weighted sum of individual irrelevances.

Kovanic argues that the "gnostical data processing" principally differs from the data processing folllowing the principles of mathematical statistics ([13], p. 657). 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], p. 658).

The first statistical light was thrown on this assertions in [3]. The author of the present paper noticed that the square of fidelity (24) 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 the α-estimators introduced by Vajda [19], for the log-logistic family. Based on this result, Vajda [20], [21] and Novovičová [15] were able to establish asymptotic statistical properties of gnostical estimators. They proved that the gnostical estimators are the usual statistical M-estimators, strongly consistent and asymptotically normal, and they derived the corresponding asymptotic variances.

The success of the estimator (26) applied to the Stiegler data sets can be explained as follows. The influence function of the robust estimator (26) is, contrary to the usual robust estimators, non-symmetrical. This fits well the clear non-symmetry of the 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 (26) belongs to the class of statistical M-estimators, although the maximum likelihood principle is not postulated in gnostical theory? In the next section we try to answer these questions.

5. STATISTICAL MEANING OF THE GNOSTICAL IRRELEVANCE AND FIDELITY

In the previous section we mentioned only one of the 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

$$h_q(z|z_0,s) = \sinh(2\Omega) = \frac{1}{2}\left[(z/z_0)^{2/s} - (z/z_0)^{-2/s}\right]. \tag{28}$$

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

Theorem 1. Probability densities corresponding to two types of Kovanic's irrelevances (25), (28) are

$$p_1(z|z_0,s) = \frac{\sqrt{2\pi}}{zs\Gamma^2(1/4)}\,\frac{1}{[(z/z_0)^{2/s} + (z/z_0)^{-2/s}]^{1/2}} \tag{29}$$

$$p_2(z|z_0,s) = \frac{1}{zsK_0(1/2)}\,e^{-\frac{1}{4}[(z/z_0)^{2/s}+(z/z_0)^{-2/s}]}, \tag{30}$$

respectively.

P r o o f. Let

$$h_{R1}(u) = \mathrm{tgh}(2u), \qquad h_{R2}(u) = \sinh(2u) \tag{31}$$

be the score functions of distributions on $(R, \mathcal{B}_R)$ (they are modifications of score functions of distributions in last two rows in Table 1). The corresponding densities are, according to (1),

$$p_{R1}(x) = c_1^{-1}e^{-\int \mathrm{tgh}(2x)\,\mathrm{d}x} = c_1^{-1}\cosh^{-1/2}(2x) \tag{32}$$

$$p_{R2}(u) = c_2^{-1}e^{-\int \sinh(2x)\,\mathrm{d}x} = c_2^{-1}e^{-\frac{1}{2}\cosh(2x)}. \tag{33}$$

Norming constants are given by integrals (see e. g., [17])

$$\int_{-\infty}^{\infty} \cosh^{-\nu} ax\,\mathrm{d}x = \frac{2^\nu\Gamma^2(\nu/2)}{a\Gamma(\nu)}, \qquad \int_{-\infty}^{\infty} e^{-\nu\cosh ax}\mathrm{d}x = 2a^{-1}K_0(\nu)$$

where Γ is the gamma function. According to (12), the IFDs of the exponentially related distributions on R^+ are given by the substitution

$$u = \ln(z/z_0)^{1/s} \tag{34}$$

into (31). We obtain

$$h_1(z|z_0, s) = s^{-1}\mathrm{tgh}(\ln(z/z_0)^{2/s}) = -s^{-1}h_e(z|z_0, s) \tag{35}$$

$$h_2(z|z_0, s) = s^{-1}\sinh(\ln(z/z_0)^{2/s}) = s^{-1}h_q(z|z_0, s), \tag{36}$$

where $-h_e$ and h_q are gnostical irrelevances (25) and (28). 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 (e. g. equation (26)). By (11), one obtains the densities corresponding to IFDs (35) and (36) from the prototype densities (32) and (33) after the substitution (34) and division by z, which gives (29) and (30). □

Theorem 2. Square of the gnostical fidelity is, apart from the constant, the weight function of the family (29).

P r o o f. Weight functions of distributions with densities (32), (33) are, using (19) and (31),

$$g_1(u) = 2\cosh^{-2}(2u), \qquad g_2(u) = 2\cosh(2u). \tag{37}$$

After substitution (34) and by the use of (21),

$$g_1(z|z_0, s) = 2s^{-2}z^{-1}\cosh^{-2}(\ln(z/z_0)^{2/s}) = 2s^{-2}z^{-1}f^2(z|z_0, s) \tag{38}$$

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

where f is the fidelity (24). Apart from the factor $2s^{-2}$, $f^2(z|z_0, s)$ is the weight function (21) of the distribution (29) (and, similarly, $f^{-1}(z|z_0, s)$ is the weight function of the distribution (30)). □

6. CONCLUSIONS

Given a model of a statistical experiment in the form of a parametric set $\mathcal{P}_T$, the observed values $u_1, \ldots, u_n$, 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. We propose a model which prescribes for each data item u_i some a priori data characteristics: the value of the IFD, $h_T(u_i|\theta^0)$, and the value of the WFD, $w_T(u_i|\theta^0)$. They are, similarly as the likelihood, in a "latent form" because of the unknown true parameter value θ^0. However, they can be approximately specified by using an appropriate estimate $\hat{\theta}$ of θ^0.

With the help of this model, theorems in the previous section give a possible statistical explanation of gnostic characteristics of data. The "ideal value" z_0 can be understood as the transformed location parameter, the "scale" s as the usual scale parameter and the "irrelevance" and "fidelity" as the IFD and the square root of the WFD of distributions (29) and (30). We thus give an explanation of Kovanic's "non-statistical" notions of irrelevance and fidelity of individual data in a rather unexpected way by including their general equivalents into the probability theory.

The Kovanic's heuristic estimate given by (26) appears to be the first IFD-moment estimate in the special model (29). By iii) of Section 2, (16) yields the maximum likelihood estimate of the location parameter without the need to apply directly the maximum likelihood principle (e. g. without the need of differentiation with respect to the location parameter). Good performances of the gnostical estimator of the location parameter can be attributed to this fact, and to the boundedness of the influence function of distribution (29)). The difficulties with the gnostical estimation of scale parameter (which are not mentioned in this paper) could be circumvented by the use of the second IFD-moment estimation equation (17).

It should be noted that we did not explain Kovanic's estimation procedures based on his "data composition law". We suppose that, in the probabilistic terms, the composition law (27) can be considered as a "finite equivalent" of a limit theorem concerning sums of i.i.d. random variables, weighted in a special way. "Qualitatively", (27) asserts that the weighted sum of i.i.d. random variables is distributed according to the original probability law. But this problem remains to be open.

ACKNOWLEDGEMENTS

The author thanks the referees for their valuable comments and suggestions. The research was partly supported by Grant Agency of the Academy of Sciences of the Czech republic under Grant No. A 1075709.

(Received December 29, 1995.)

REFERENCES

[1] S. Amari: Differential–Geometrical Methods in Statistics (Lecture Notes in Statistics *28*). Springer–Verlag, Berlin – Heidelberg – New York 1985.

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

[3] Z. Fabián: Point estimation in case of small data sets. In: Trans. 10th Prague Conf. on Inform. Theory, Statist. Dec. Functions, Random Processes, Academia, Prague 1988, pp. 305–312.

[4] Z. Fabián: Generalized score function and its use. In: Trans. 12th Prague Conf. on Inform. Theory, Statist. Dec. Functions, Random Processes, ÚTIA AV ČR, Prague 1994.

[5] Z. Fabián: Information and entropy of continuous random variables. IEEE Trans. Inform. Theory *43* (1997), 3.

[6] Z. Fabián: Geometric Moments. Techn. Report No. V–699, ICS AS CR, Prague 1996.

[7] Z. Fabián: Geometric moments (in Czech). In: Trans. ROBUST'96, JČMF, Prague 1997.

[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. Interscience Publishers, New York – London 1963.
[10] P. Kovanic: Gnostical theory of individual data. Problems Control Inform. Theory *13* (1984), 4, 259–274.
[11] P. Kovanic: Gnostical theory of small samples of real data. Problems Control Inform. Theory *13* (1984), 5, 303–319.
[12] P. Kovanic: On relation between information and physics. Problems Control Inform. Theory *13* (1984), 6, 383–399.
[13] P. Kovanic: A new theoretical and algorithmical tool for estimation, identification and control. Automatica *22* (1986), 6, 657–674.
[14] P. Kovanic and J. Novovičová: Comparizon of statistical and gnostical estimates of parameter of location on real data (in Czech). In: Proc. of ROBUST, JČMF, Prague 1986.
[15] J. Novovičová: M-estimators and gnostical estimators of location. Problems Control Inform. Theory *18* (1989), 6, 397–407.
[16] G. P. Patil, M. T. Boswell and M. V. Ratnaparkhi: Dictionary and classified bibliography of statistical distributions in scientific work. In: Internat. Co-operative Publ. House, Maryland 1984.
[17] A. P. Prudnikov, J. A. Brychkov and O. I. Marichev: Integrals and Series (in Russian). Nauka, Moskva 1981.
[18] S. M. Stiegler: Do robust estimators work with real data? Ann. Statist. *6* (1977), 1055–1098.
[19] I. Vajda: Efficiency and robustness control via distorted maximum likelihood estimation. Kybernetika *22* (1986), 1, 47–67.
[20] I. Vajda: Minimum–distance and gnostical estimators. Problems Control Inform. Theory *17* (1987), 5, 253–266.
[21] I. Vajda: Comparison of asymptotic variances for several estimators of location. Problems Control Inform. Theory *18* (1989), 2, 79–87.

RNDr. Zdeněk Fabián, CSc., Ústav informatiky a výpočetní techniky AV ČR (Institute of Computer Science – Academy of Sciences of the Czech Republic), Pod vodárenskou věží 2, 182 07 Praha 8. Czech Republic.