KYBERNETIKA — VOLUME 16 (1980), NUMBER 2

Analysis of a Measurement Information

IGOR VAJDA, KAREL ECKSCHLAGER

In certain papers, an information in a measurement x concerning an unknown parameter with (subjective or objective) prior distribution p_0 has been suggested, based on a suitably specified concept of posterior distribution p_x. We analyse this information in several respects. First, its logarithmical nature is characterized axiomatically. Then it is shown that the p_x should be defined by the well-known Bayes formula or, alternatively, by using asymptotic efficiency and normality of estimators involved, depending on whether the p_0 is rather objective or subjective respectively. It is further shown that, in the first case, the resulting measurement information is in an average sense equal to a Shannon information while, in the second case, it is closely related to a Fisher information of the underlying statistical measurement model. All results are illustrated by examples of discrete as well as continuous-type measurements.

1. INTRODUCTION

In recent years many authors underlined the need for a quantitative definition of information in situations described by mathematical models partly or completely different from the classical "pair of random variables" model of Shannon information theory. One such concept of information has been quite systematically applied and investigated by the authors dealing with statistical and theoretical aspects of measurements in analytical chemistry (see [1, 2] and further references given there) with applicability, however, far beyond this particular field. Let us briefly describe it.

Suppose that one has to specify the value of a real variable θ from an interval Θ on the real line E_1 and that a prior knowledge concerning θ is expressed by a probability P_0 with density $p_0(\theta)$ on Θ. Suppose now that a measurement (experiment, observation) has been carried out resulting into a posterior probability P with density $p(\theta)$ on Θ. Then the information content of the result is recommended to be measured either by

$$(1.1)\quad I(p_0, p) = H(p_0) - H(p) = -\int_\Theta p_0(\theta) \ln p_0(\theta)\, d\theta + \int_\Theta p(\theta) \ln p(\theta)\, d\theta$$

(a difference between prior and posterior differential entropy [3]) or by

$$\tilde{I}(p_0, p) = \int_{\Theta} p(\theta) \ln \frac{p(\theta)}{p_0(\theta)}\, d\theta \tag{1.2}$$

(informational divergence between prior and posterior probabilitiy [4]). It is clear that, though both these functionals originate in the Shannon information theory, neither of them can be interpreted as the Shannon information itself. The reason is that the mathematical model of measurement, described by the pair $\{p_0, p\}$ only, is too superficial and it does not exhibit explicitely the probability-product-space structure required by the Shannon definition.

Let us introduce into our consideration an observation channel $(\Theta, p(\cdot|\cdot), E_n)$ with input alphabet Θ, output alphabet E_n and channel transition probability function $p(x \mid \theta)$. This function is supposed to be a probability density on E_n, measurable with respect to the Borel σ-algebra $\mathscr{B}_n$ for every $\theta \in \Theta$, and measurable with respect to $\Theta \cap \mathscr{B}_1$ for every $x = (x_1, \ldots, x_n) \in E_n$. Define, moreover, a statistic $T : X \to \Theta$ with $T(x)$ used to estimate (or to measure) the true but unknown value $\theta \in \Theta$. The observation channel and statistic represent basic components of any statistical decision concerning the unknown parameter θ. They already allow us to define the Shannon information at the output of the observation channel concerning θ (a priori described by p_0) by

$$I = H(p_0) - \int_{E_n} H(p(\cdot \mid x))\, p(x)\, dx = \tag{1.3}$$

$$= \int_{\Theta} \int_{E_n} p_0(\theta)\, p(x \mid \theta) \ln \frac{p(x \mid \theta)}{p(x)}\, d\theta\, dx\,,$$

where

$$p(\theta \mid x) = \frac{p(x \mid \theta)\, p(\theta)}{p(x)} \quad \text{and} \quad p(x) = \int_{\Theta} p(x \mid \theta)\, p(\theta)\, d\theta \tag{1.4}$$

is posterior probability density of θ provided x was observed or unconditional (marginal) probability density of the observation channel output respectively. We could also define the Shannon information in the outcome of all the measurement process $T(x)$ concerning the unknown θ by

$$I(T) = H(p_0) - \int_{E_n} H(p*(\cdot \mid T(x)))\, p(x)\, dx\,,$$

where $p*(\theta \mid \theta')$ is a posterior probability density of θ provided $T(x) = \theta' \in \Theta$ is estimating θ. It is known that $I(T) \leqq I$ with equality if T is sufficient statistics for the observation channel. Since we shall restrict our attention to sufficient estimators only, we shall consider Shannon information (1.3) only.

The above stated statistical description of the measurement process provides at the same time the necessary basis for an explicit specification of the posterior probability p on Θ.

(i) First, we can define $p(\theta) = p(\theta \mid x)$ depending on the empirical evidence $x \in E_n$ observed. Then $I(p_0, p(\cdot \mid x))$ or $\bar{I}(p_0, p(\cdot \mid x))$ measure the information contained in the measurement x concerning the unknown parameter θ.

(ii) Second possibility can be applied if there exists an observation channel $(\Theta, q(\cdot|\cdot), E_1)$ such that $p(x_1, \ldots, x_n \mid \theta) = q(x_1 \mid \theta) \ldots q(x_n \mid \theta)$ for every $(x_1, \ldots x_n \in E_n$ and $\theta \in \Theta$. In this case the sample $x = (x_1, \ldots, x_n) \in E_n$ can be interpreted as a realization of a random vector $\xi = (\xi_1, \ldots, \xi_n)$ with components independent and identically distributed according to densities $q(\cdot \mid \theta)$. For practically each estimator T it is then true that (see Sec. 5.5 in [5]) $\sqrt{(n)}\,(T(\xi_1, \ldots, \xi_n) - \theta) \approx No(0, (n\, I(\theta))^{-1})$ provided the observation channel $\{\Theta, q(\cdot|\cdot), E_1\}$ is regular, where $I(\theta) = \int (q'(x \mid \theta)^2 / q(x \mid \theta))\,dx$ is the Fisher information at the output of $\{\Theta, q(\cdot|\cdot), E_1\}$ concerning the value θ. This fact also extends with small modifications to non-regular situations as well (see [6, 7]). Consequently, $T(\xi_1, \ldots, \xi_n) \approx No(\theta, (n\, I(\theta))^{-1})$. Thus, as soon as an estimator T is a part of the measurement process, $No(\theta, (n\, I(\theta))^{-1})$ becomes a realistic alternative for posterior probability corresponding to a measurement. Since the true parameter $\theta \in \Theta$ is unknown, we can use the so called estimator-generated posterior probability

$$(1.5)\quad p(\theta) = No(T(x), (n\, I(T(x)))^{-1}) = \sqrt{\frac{n\, I(T(x))}{2\pi}} \exp\left(-\frac{(\theta - T(x))^2\, n\, I(T(x))}{2}\right)$$

(truncated outisde Θ if $\Theta \neq E_1$).

In general, posterior probability defined by this manner differs from $p(\theta \mid x)$ as defined in (i). However, in the general context specified at the beginning of this section, the estimator-generated posterior probability seems to be justified by the argument given above.

If the prior density p_0 results from a prior measurement, not included in our model, then it follows from (ii) that is is usually $No(\mu_0, \sigma_0^2)$ with some fixed $\mu_0 \in E_1$, $\sigma_0^2 > 0$. If p_0 has to be stated without any prior empirical evidence, it is usually uniform on Θ (in symbols, $p_0 = U(\Theta)$), which can practically always be a finite interfal (θ_0, θ_1). This is why we shall pay a particular attention to $p_0 = No(\mu_0, \sigma_0^2)$ and $p_0 = U(\theta_0, \theta_1)$.

To give a simple example, put $q(x \mid \theta) = \chi_{\{0,1\}}(x)\, \theta^x (1 - \theta)^{1-x}$, i.e.

$$p(x_1, \ldots, x_n \mid \theta) = \varkappa_{\{0,1\}}(x_1, \ldots, x_n)\, \theta^{\Sigma x_i} (1 - \theta)^{n - \Sigma x_i} \quad \text{for} \quad \theta \in (0, 1) = \Theta\,.$$

Here $I(\theta) = 1/[\theta(1 - \theta)]$. If $p_0 = U(0, 1)$, then $p(\theta \mid x_1, \ldots, x_n) = Beta\,(\Sigma x_i + 1, n - \Sigma x_i + 1)$ so that

$$\frac{\Sigma x_i + 1}{n + 2} \quad \text{and} \quad \frac{(\Sigma x_i + 1)(n - \Sigma x_i + 1)}{(n + 3)(n + 2)^2}$$

are the posterior expectation and variance of the unknown parameter. On the other hand, if we estimate the unknown parameter by means of the Bayes estimator

$$T(x) = \frac{\Sigma x_i + 1}{n + 2}$$

(corresponding to the mean square error loss function), then the estimator-generated posterior probability (1.5) is

$$No\left(\frac{\Sigma x_i + 1}{n + 2}, \frac{(\Sigma x_i + 1)(n - \Sigma x_i + 1)}{n(n + 2)^2}\right)$$

truncated outside $\Theta = (0, 1)$. This probability obviously differs from the posterior probability *Beta* $(\Sigma x_i + 1, n - \Sigma x_i + 1)$.

Remark at the end of this introductory section that, while Shannon information describes an average information at the output of the observation channel, the measurement information mentioned at the beginning of this section aims at specification of an information in concrete individual observations $x \in E_n$. This feature differs the theory of measurement information, as developed in [1, 2] and papers cited there, from the classical Shannon theory.

Let us also remark, that a problem of information in a measurement has been introduced many years ago by Perez [8], but his theory goes in somewhat different direction since the object of estimation is not the parameter θ itself, but the prior density $p_0(\theta)$. Nevertheless it may be stimulating to join the theory of measurement information with the theory developed in [8].

The aim of the present paper is (i) to extend the motivation of the definitions (1.1) and (1.2), (ii) to compare these two definitions mutually and with the Shannon information, and (iii) to introduce in more detail the idea of the estimator-generated posterior probability and its impact to the definitions (1.1) and (1.2).

2. INTUITIVE INFORMATION AND ITS QUANTIFICATION

If we want to characterize mathematically the information concerning an unknown parameter θ from an abstract set Θ gained by carrying out a measurement (experiment, observation) of θ then, according to what was said above, one possibility is to characterize the measurement by a probability P defined in a suitably specified measure space $(\Theta, \mathscr{T})$. Suppose that the situation in which the measurement is to be carried out is characterized by a prior probability P_0 on $(\Theta, \mathscr{T})$. Thus, the information is to be related to pairs $\{P_0, P\}$ of probabilities on an abstract measure space $(\Theta, \mathscr{T})$. As examples of the pair let us consider $(\Theta, \mathscr{T}) = (E_1, \mathscr{B}_1)$ and $\{P_0, P\} = \{No(\mu_0, \sigma_0^2), No(\mu, \sigma^2)\}$ or $(\Theta, \mathscr{T}) = (\{0, 1, \ldots, 100\}$ = percentage levels, $\mathscr{T}$ = all subsets of $\{0, 1, \ldots, 100\})$, and $\{P_0, P\} = \{Bi(\theta_0, 100), Bi(\theta, 100)\}$ ($Bi(\theta, m)$ stands for binomial with parameters $\theta \in (0, 1)$, $m = 1, 2, \ldots$).

Let now $\mathscr{C}$ be a non-empty class of measurable spaces $(\Theta, \mathscr{T})$ and $\mathscr{P}(\mathscr{C})$ the class of all pairs $\{P_0, P\}$ on elements $(\Theta, \mathscr{T})$ of the $\mathscr{C}$. We introduce first the concept of intuitive information as a non-empty subset $\mathscr{I} \subset \mathscr{P}(\mathscr{C}) \times \mathscr{P}(\mathscr{C})$. Thus, the intuitive information we define as a partial ordering $\gtrsim$ in $\mathscr{P}(\mathscr{C})$, where $(P_0, P) \gtrsim (P'_0, P')$ is interpreted in such a way, that the information of a measurement P in a situation P_0 is at least as large as the information of a measurement P' in a situation P'_0.

Let $\mathscr{P}_{\mathscr{I}}(\mathscr{C})$ be a projection of $\mathscr{I}$ to $\mathscr{P}(\mathscr{C})$, i.e., let $\mathscr{P}_{\mathscr{I}}(\mathscr{C}) = \{(P_0, P) \in \mathscr{P}(\mathscr{C})$: either $(P'_0, P') \gtrsim (P_0, P)$ or $(P''_0, P'') \lesssim (P_0, P)$ for some $P'_0, P', P''_0, P''\}$. A real-valued function $I(P_0, P)$ defined on a subset $\mathscr{P}_I(\mathscr{C})$, $\mathscr{P}_{\mathscr{I}}(\mathscr{C}) \subset \mathscr{P}_I(\mathscr{C}) \subset \mathscr{P}(\mathscr{C})$, possessing the property

$$I(P_0, P) \geqq I(P'_0, P') \quad \text{iff} \quad (P_0, P) \gtrsim (P'_0, P') \tag{2.1}$$

is a quantificator of the intuitive information and we call it simply an *information* (of a measurement P in a situation P_0) consistent with the intuitive information.

Suppose for example, that $\mathscr{C}$ contains the measurable space $(E_1, \mathscr{B}_1)$ considered in the example above, and that $\mathscr{P}(\mathscr{C})$ consists of all pairs $\{P_0, P\}$ on $(E_1, \mathscr{B}_1)$. Let us define $\{No(\mu_0, \sigma_0^2), No(\mu, \sigma^2)\} \gtrsim \{No(\mu'_0, \sigma_0'^2), No(\mu', \sigma'^2)\}$ iff

$$\frac{\sigma_0}{\sigma} \geqq \frac{\sigma'_0}{\sigma'}\,. \tag{2.2}$$

This relation defines an intuitive information $\mathscr{I}$ with

$$\mathscr{P}_{\mathscr{I}}(\mathscr{C}) = \{(No(\mu_0, \sigma_0^2), No(\mu, \sigma^2)) : \mu_0, \mu \in E_1, \sigma_0^2, \sigma^2 > 0\}$$

and

$$I(No(\mu_0, \sigma_0^2), No(\mu, \sigma^2)) = \frac{\sigma_0}{\sigma} \quad \text{or} \quad \ln \frac{\sigma_0}{\sigma} \tag{2.3}$$

or

$$I(P_0, P) = \ln \left(\frac{\text{dispersion of } P_0}{\text{dispersion of } P}\right) \tag{2.4}$$

are examples of possible information with different domains $\mathscr{P}_I(\mathscr{C}) \supset \mathscr{P}_{\mathscr{I}}(\mathscr{C})$ but all consistent with $\mathscr{I}$.

Since the intuitive information can be set up much easier than a quantitative information measure, it may substantially help in selecting an adequate information measure $I(P_0, P)$ in a given concrete situation.

3. ADDITIVE INFORMATION

We start with some technical remarks. If we want to generalize (1.1) or (1.2) to abstract Θ (more precisely, to abstract measurable spaces $(\Theta, \mathscr{T})$), we must refer to some σ-finite measures μ_0, μ dominating P_0, P on $\mathscr{T}$. Indeed, if we consider the Radon-Nikodym densities $p_0 = dP_0/d\mu_0$, $p = dP/d\mu$, then we can consider differences

$$(3.1) \qquad I(P_0, P \mid \mu_0, \mu) = \int_\Theta p(\theta) \ln p(\theta) \, d\mu(\theta) - \int_\Theta p_0(\theta) \ln p_0(\theta) \, d\mu_0(\theta) ,$$

or we can consider quantities

$$(3.2) \qquad \tilde{I}(P_0, P \mid \mu) = \frac{1}{P(\Theta)} \int_\Theta p(\theta) \ln \frac{p(\theta)}{p_0(\theta)} \, d\mu(\theta) ,$$

where it is supposed that μ is dominating both P_0 and P and $p_0 = dP_0/d\mu$, $p = dP/d\mu$. The notation employed in (3.1), (3.2) we use throughout this paper. In this section we shall suppose, moreover, that the class $\mathscr{C}$ of measurable spaces is closed with respect to Cartesian product $(\Theta, \mathscr{T}) \times (\Theta', \mathscr{T}') = (\Theta \times \Theta', \mathscr{T} \times \mathscr{T}')$ and that $\mathscr{C}$ contains at least one pair $(\Theta, \mathscr{T}), (\Theta', \mathscr{T}')$ with atoms $\emptyset \neq A \in \mathscr{T}$, $\emptyset \neq A' \in \mathscr{T}'$ (i.e. with subsets such that $A \cap \mathscr{T} = \{\emptyset, A\}$, $A' \cap \mathscr{T}' = \{\emptyset, A'\}$). Since truncation may lead to probabilities which are not necessarity normed to 1, it will be useful for us to refer also to the class $\mathscr{P}^*(\mathscr{C})$ of all pairs (P_0, P) of finite measures on elements $(\Theta, \mathscr{T})$ of $\mathscr{C}$. This also explains the role of the denominator $P(\Theta)$ in (3.2) and (3.4).

Among of many functions $I(P_0, P)$ which may be consistent with various intuitive information on various classes $\mathscr{P}(\mathscr{C})$ we shall concentrate our attention on information of entropy-difference-type

$$(3.3) \qquad I_f(P_0, P \mid \mu_0, \mu) = \int_\Theta f(p(\theta)) \, d\mu(\theta) - \int_\Theta f(p_0(\theta)) \, d\mu_0(\theta) ,$$

which is an obvious generalisation to (3.1), and on information of informational-divergence-type

$$(3.4) \qquad \tilde{I}_f(P_0, P \mid \mu) = \frac{1}{P(\Theta)} \int_\Theta f(p(\theta), p_0(\theta)) \, d\mu(\theta)$$

generalizing (3.2).

Our last general remark concerns independent measurements. If two measurements characterized by pairs (P_0, P), (P'_0, P') on measurable spaces $(\Theta, \mathscr{T})$, $(\Theta', \mathscr{T}')$ from $\mathscr{C}$ are carried out independently, they can be obviously viewed as one measurement $(P_0 \times P'_0, P \times P') \in \mathscr{P}(\mathscr{C})$ on $(\Theta \times \Theta', \mathscr{T} \times \mathscr{T}')$ from $\mathscr{C}$. In other words, here prior knowledges P_0, P'_0 in both measurements do not mutually interfere as do not interfere the measurements themselves. In this situation, in accordance with Shannon or Fisher information theory, it is natural to require the additivity property of any meaningful information:

$$(3.5) \qquad I(P_0 \times P'_0, P \times P') = I(P_0, P) + I(P'_0, P').$$

As we shall see, this property reduces the information (3.3) to the information (3.1) and (3.4) to (3.2).

Theorem 3.1. If $f : [0, \infty) \to E_1$ is continuous with $f(0) = 0$ and if it holds

$$(3.6) \qquad I_f(P_0 \times P'_0, P \times P' \mid \mu_0 \times \mu'_0, \mu \times \mu') = \\ = I_f(P_0, P \mid \mu_0, \mu) + I_f(P'_0, P' \mid \mu'_0, \mu')$$

on $\mathscr{P}(\mathscr{C})$, then there exist constants $c, c^* \in E_1$ such that $f(u) = c \ln u + c^* u$ for every $u \in [0, \infty)$ so that $I_f(P_0, P \mid \mu_0, \mu) = I(P_0, P \mid \mu'_0, \mu)$ on $\mathscr{P}(\mathscr{C})$.

Proof. Let $u, v > 0$ be arbitrary and let us consider μ_0, μ, μ'_0 and μ' such, that

$$\mu_0(A) = 1/u\,, \quad \mu'_0(A') = 1/v\,, \quad \mu(A) = \mu'(A') = 1\,,$$

where A, A' are atoms of $(\Theta, \mathscr{T})$, $(\Theta', \mathscr{T}')$ so that $A \times A'$ is an atom of $(\Theta \times \Theta', \mathscr{T} \times \mathscr{T}')$. Let us further consider densities

$$p_0(\theta) = \begin{cases} u & \text{on } A \\ 0 & \text{on } A^c \end{cases} \qquad p'_0(\theta) = \begin{cases} v & \text{on } A' \\ 0 & \text{on } A'^c \end{cases}$$

$$p(\theta) = \begin{cases} 1 & \text{on } A \\ 0 & \text{on } A^c \end{cases} \qquad p'(\theta) = \begin{cases} 1 & \text{on } A' \\ 0 & \text{on } A'^c\,. \end{cases}$$

Then we obtain from the definition (3.3)

$$I_f(P_0 \times P'_0\, P, \times P' \mid \mu_0 \times \mu'_0, \mu \times \mu') = \frac{f(uv)}{uv} - f(1)\,,$$

$$I_f(P_0, P \mid \mu_0, \mu) = \frac{f(u)}{u} - f(1)\,,$$

$$I_f(P'_0, P' \mid \mu'_0, \mu) = \frac{f(v)}{v} - f(1)\,.$$

Let us now consider $g(u) = f(u)/u - f(1)$. It follows from what we found here and from (3.6) that $g(uv) = g(u) + g(v)$. Since g is obviously continuous, it follows from the well-known lemma of Cauchy that $g(u) = conts. \ln u$. The remainder is clear. □

Theorem 3.2. If $f : [0, \infty)^2 \to E_1$ is continuous with $f(0, 0) = 0$ and if it holds

$$(3.7) \qquad \tilde{I}_f(P_0 \times P'_0, P \times P' \mid \mu_0 \times \mu'_0) = \tilde{I}_f(P_0, P \mid \mu_0) + \tilde{I}_f(P'_0, P' \mid \mu'_0)$$

and

(3.8) $$\tilde{I}_f(P_0, P_0 \mid \mu_0) = 0$$

on $\mathscr{P}^*(\mathscr{C})$, then there exists $c \in E_1$ such that $f(u, v) = cu \ln u/v$ for every $u, v \in [0, \infty)$, so that $\tilde{I}_f(P_0, P \mid \mu) = \tilde{I}(P_0, P \mid \mu)$ on $\mathscr{P}^*(\mathscr{C})$.

Proof. Let A, A' be atoms as in the preceding proof and let u, v, P_0, P'_0 and μ_0, μ'_0 be also the same as there. Let $u^*, v^* > 0$ be arbitrary and define

$$p(\theta) = \begin{cases} u^* & \text{on } A \\ 0 & \text{on } A^c \end{cases} \qquad p'(\theta) = \begin{cases} v^* & \text{on } A' \\ 0 & \text{on } A'^c . \end{cases}$$

Then, by (3.4),

$$\tilde{I}_f(P_0, P'_0, P \times P' \mid \mu \times \mu'_0) = \frac{f(u^*v^*, uv)}{u^*v^*},$$

$$\tilde{I}_f(P_0, P \mid \mu_0) = \frac{f(u^*, u)}{u^*},$$

$$\tilde{I}_f(P'_0, P' \mid \mu'_0) = \frac{f(v^*, v}{v^*}.$$

Therefore it follows from (3.7) that

$$\frac{f(u^*v^*, uv)}{u^*v^*} = \frac{f(u^*, u)}{u^*} + \frac{f(v^*, v)}{v^*}$$

or

(3.9) $$g(uv, u^*v^*) = g(u, u^*) + g(v, v^*),$$

where

$$g(u, u^*) = \frac{f(u^*, u)}{u^*} \quad \text{for every} \quad u, u^* > 0 .$$

By (3.9) it holds

(3.10) $$g(u, u^*) = g(u \,.\, 1, 1 \,.\, u^*) = g(u, 1) + g(1, u)$$

and

$$g(uv, 1) = g(u, 1) + g(v, 1),$$

$$g(1, u^*v^*) = g(1, u^*) + g(1, v^*).$$

Since $g(u, 2)$ as well as $g(1, u^*)$ are continuous, by the lemma of Cauchy we obtain that there exist $c, c^* \in E_1$ such that

$$g(u, 1) = c^* \ln u, \quad g(1, u^*) = c \ln u^* .$$

Hence, by (3.10), $g(u, u^*) = c^* \ln u + c \ln u^*$ so that $f(u^*, u) = c^* u^* \ln u + cu^* \ln u^*$ or $f(u, v) = c^* u \ln v + cu \ln u$ and, finally,

$$f(u, v) = cu \ln \frac{u}{v} + 2c^* u \ln v . \tag{3.11}$$

Since

$$\frac{f(u, u)}{u} = \tilde{I}_f(P_0, P_0 \mid \mu_0) ,$$

we obtain from (3.8) and (3.11) $2c^* \ln u = 0$ for every $u > 0$ so that $c^* = 0$ and theorem is proved.

It is a well-known fact of information theory (see [4, 11]) that $\tilde{I}(P_0, P \mid \mu)$ is constant when the dominating measure μ is varied. Consequently, in what follows we shall drop the symbol μ and, for probabilities P_0, P, we shall write simply

$$\tilde{I}(P_0, P) = \int_\Theta p(\theta) \ln \frac{p(\theta)}{p_0(\theta)} \, d\mu(\theta) . \tag{3.12}$$

The information (3.1) unfortunatelly depends on both μ_0, μ. It is convenient, however, to consider the entropy difference with respect to the same dominating measure, so that

$$I(P_0, P \mid \mu) = \int_\Theta p(\theta) \ln p(\theta) \, d\mu(\theta) - \int_\Theta p_0(\theta) \ln p_0(\theta) \, d\mu(\theta) \tag{3.13}$$

with $p = dP/d\mu$, $p_0 = dP/d\mu$ is the most convenient variant of (3.1). But, as we shall see from (4.2), (4.3) below, even (3.13) depends on the dominating measure μ.

4 SIMILARITIES AND DIFFERENCES BETWEEN $I(P_0, P \mid \mu)$ AND $\tilde{I}(P_0, P)$.

It is easy to see from (3 12) and (3 13) that

$$I(P_0, P \mid \mu) = \tilde{I}(P_0, P) + \int_\Theta (p(\theta) - p_0(\theta)) \ln p_0(\theta) \, d\mu(\theta) . \tag{4.1}$$

Here $\bar{I}(P_0, P)$ is invariant to a modification of the dominating measure μ as well as the elements $p(\theta)\,\mathrm{d}\mu(\theta) = \mathrm{d}P(\theta)$, $p_0(\theta)\,\mathrm{d}\mu(\theta) = \mathrm{d}P_0(\theta)$ Thus if we pass to a new dominating measure ν, we get

$$I(P_0, P \mid \nu) = I(P_0, P \mid \mu) + \Delta(P_0, P \mid \mu, \nu)\,, \tag{4.2}$$

where

$$\Delta(P_0, P \mid \mu, \nu) = \begin{cases} \mathsf{E}_P \ln \dfrac{\mathrm{d}\nu}{\mathrm{d}\mu} - \mathsf{E}_{P_0} \ln \dfrac{\mathrm{d}\nu}{\mathrm{d}\mu} & \text{if} \quad \nu \ll \mu \\ \mathsf{E}_{P_0} \ln \dfrac{\mathrm{d}\mu}{\mathrm{d}\nu} - \mathsf{E}_P \ln \dfrac{\mathrm{d}\mu}{\mathrm{d}\nu} & \text{if} \quad \mu \ll \nu \end{cases} \tag{4.3}$$

(measures μ, ν satisfying neither of the two relations we exclude from our considerations). From a generalized Cramér-Rao inequality (15) in [10] we easily obtain the following

Theorem 4.1. It holds

$$|\Delta(P_0, P \mid \mu, \nu)| \leqq \varrho_{P_0}(\mu, \nu)\,\sqrt{\chi^2(P_0, P)}\,, \tag{4.4}$$

where

$$\varrho^2_{P_0}(\mu, \nu) = \begin{cases} \mathsf{E}_{P_0} \ln^2 \dfrac{\mathrm{d}\nu}{\mathrm{d}\mu} & \text{if} \quad \nu \ll \mu\,, \\ \mathsf{E}_{P_0} \ln^2 \dfrac{\mathrm{d}\mu}{\mathrm{d}\nu} & \text{if} \quad \mu \ll \nu\,, \end{cases}$$

and

$$\chi^2(P_0, P) \doteq \int_\Theta \frac{(p(\theta) - p_0(\theta))^2}{p_0(\theta)}\,\mathrm{d}\mu(\theta)$$

is the so-called χ^2-divergence of probabilities P, P_0 (which is again independent on the dominating μ).

We see from (4.4) that the difference $|I(P_0, P \mid \mu) - I(P_0, P \mid \nu)|$ is small if the pseudo-distance $\varrho_{P_0}(\mu, \nu)$ between μ and ν is small.

Since (4.1) can be rewritten in the form

$$I(P_0, P \mid \mu) - \bar{I}(P_0, P) = \mathsf{E}_P \ln \frac{\mathrm{d}P_0}{\mathrm{d}\mu} - \mathsf{E}_{P_0} \ln \frac{\mathrm{d}P_0}{\mathrm{d}\mu}\,, \tag{4.5}$$

inequality (15) of [10] can again be applied and we obtain

$$\left|I(P_0, P \mid \mu) - \tilde{I}(P_0, P)\right| \leqq \varrho_{P_0}(P_0, \mu) \sqrt{\chi^2(P_0, P)} ,$$

where

$$\varrho^2_{P_k}(P_0, \mu) = \mathsf{E}_{P_0} \ln^2 \frac{\mathrm{d}P_0}{\mathrm{d}\mu} .$$

This result indicates that the difference between $I(P_0, P \mid \mu)$ and $\tilde{I}(P_0, P)$ can be made arbitrary small by selecting

$$\mu = wP + (1 - w) P_0 \gg P_0, P \quad \text{for} \quad w \in (0, 1), \tag{4.6}$$

with w small enough. Since the dominating measures (probabilities) formed by mixing prior and posterior probabilities seem to be interesting, we shall state this result in more details.

Theorem 4.2. For every μ defined by (4.6) it holds

$$0 \leqq \tilde{I}(P_0, P) - I(P_0, P \mid \mu) \leqq w^2 \chi^2(P_0, P), \tag{4.7}$$

where the left-hand inequality is strict unless $P_0 = P$. Consequently, if $\chi^2(P_0, P) < \infty$, we have

$$\lim_{w \to 0} \tilde{I}(P_0, P) - I(P_0, P \mid \mu) = 0 . \tag{4.8}$$

If $P \not\ll P_0$, then $\tilde{I}(P_0, P) = \infty$, and if $P \ll P_0$, then

$$\tilde{I}(P_0, P) = I(P_0, P \mid \mu) \quad \text{with} \quad \mu = P_0. \tag{4.9}$$

Proof. Let ν be arbitrary σ-finite measure dominating both P_0, P and $p_0 = \mathrm{d}P_0/\mathrm{d}\nu,$, $p = \mathrm{d}P/\mathrm{d}\nu$. Then for $\mu = wP + (1 - w) P_0$ we can write

$$I(P_0, P \mid \mu) = \int_\Theta p \ln \frac{p}{wp + (1 - w) p_0} \mathrm{d}\nu + \int_\Theta p_0 \ln \frac{p_0}{wp + (1 - w) p_0} \mathrm{d}\nu =$$

$$= \int_\Theta p \ln \frac{p}{p_0} \mathrm{d}\nu + \int_\Theta p_0 \ln \left(1 + w \frac{p - p_0}{p_0}\right) \mathrm{d}\nu - \int_\Theta p \ln \left(1 + w \frac{p - p_0}{p_0}\right) \mathrm{d}\nu =$$

$$= \tilde{I}(P_0, P) - \mathsf{E}_{P_0} f(\xi) ,$$

where

$$f(u) = u \ln (1 + wu) \quad \text{for} \quad u \in [-1, \infty) ,$$

$$\xi(\theta) = \frac{p(\theta) - p_0(\theta)}{p_0(\theta)} \in [1, \infty) .$$

Since $f(u) \leqq wu^2$ and $\mathsf{E}_{P_0} \xi^2 = \chi^2(P_0, P)$, it holds

$$\mathsf{E}_{P_0} f(\xi) \leqq w\,\chi^2(P_0, P)$$

so that the right-hand inequality in (4.7) holds. Since, further, $f(u) \geqq wu^2/(1 + wu) \geqq$ $\geqq wu^2 > 0$ for $u \in [-1, 0]$, it holds

$$\mathsf{E}_{P_0} f(\xi) \geqq \int_\Theta f(\xi)\, p_0 \,\mathrm{d}\nu = P_0(\Theta_0)\, \mathsf{E}_{P_0(\cdot|\Theta_0)} f(\xi)\,,$$

where $\Theta_0 = \{\theta \in \Theta : \xi(\theta) > 0\}$. If $P_0(\Theta_0) = 0$, then $P_0 = P$ and (4.7) is proved. In the opposite case we can define the conditional probability $P_0(\cdot \mid \Theta_0)$ on Θ. Since $f(u)$ is convex in the domain $u \in [0, \infty)$ and $P_0(\xi > 0 \mid \Theta_0) = 1$, we can apply the Jensen inequality to obtain

$$\mathsf{E}_{P_0(\cdot|\Theta_0)} f(\xi) \geqq f(\mathsf{E}_{P_0(\cdot|\Theta_0)}\, \xi) = f\left(\frac{P(\Theta) - P_0(\Theta)}{P_0(\Theta)}\right).$$

Since $(P(\Theta) - P_0(\Theta))/P_0(\Theta) > 0$ and $f(u) > 0$ for $u > 0$, (4.7) is proved. The fact that $\tilde{I}(P_0, P) < \infty$ only if $P \ll P_0$ has been proved in [11]. In this case

$$I(P_0, P \mid P_0) = \int_\Theta p \ln p \,\mathrm{d}P_0 - \int_\Theta 1 \ln 1 \,\mathrm{d}P_0 \quad \text{for} \quad p = \frac{\mathrm{d}P}{\mathrm{d}P_0}$$

so that (4.9) obviously holds.

Remark. Theorem 4.2 yields a result, which is of certain interest itself. The quantity

$$H(P_0 \mid \mu) = -\mathsf{E}_{P_0} \ln \frac{\mathrm{d}P_0}{\mathrm{d}\mu} = -\int_\Theta p_0 \ln p_0 \,\mathrm{d}\mu \quad \text{for} \quad p_0 = \frac{\mathrm{d}P_0}{\mathrm{d}\mu} \tag{4.10}$$

is a generalized *Shannon entropy* of probability P_0 relative to a dominating measure μ (see Perez [11]; $H(P_0 \mid \mu)$ reduces to the well-known Shannon entropy if Θ is discrete and μ is a counting measure, while it reduces to the Shannon differential entropy if $\Theta \subset E_1$ and μ is the Lebesgue measure). On the other hand

$$H(P_0, P \mid \mu) = -\mathsf{E}_P \ln \frac{\mathrm{d}P_0}{\mathrm{d}\mu} = -\int_\Theta p \ln p_0 \,\mathrm{d}\mu \tag{4.11}$$

for $p = \mathrm{d}P/\mathrm{d}\mu$ is a generalized *Bongard entropy* [12] (or *Kerridge inaccuracy* [13]) of probabilities P_0, P relative to a dominating measure μ (it reduces to the Bongard entropy or Kerridge inaccuracy if Θ is discrete and μ counting). Now it follows from (4.5) and (4.7), that the generalized Bongard entropy is always greater than or equal to generalized Shannon entropy and that the two are equal iff $P_0 = P$, i.e.

$$(4.12) \qquad -\int_\Theta p \ln p_0 \, d\mu \geqq -\int_\Theta p_0 \ln p_0 \, d\mu$$

with equality iff $p = p_0$ a.e. $[\mu]$.

Now we turn our attention to cases where $\mu(\Theta_0) < \infty$ for μ dominating P_0, P and some measurable subset $\Theta_0 \subset \Theta$ and where the prior probability P_0 is uniform on Θ_0, i.e.

$$p_0 = \frac{dP_0}{d\mu} = \begin{cases} \mu(\Theta_0)^{-1} & \text{on } \Theta_0 \\ 0 & \text{on } \Theta - \Theta_0 = \Theta\,. \end{cases}$$

In this case (for $p = dP/d\mu$)

$$(4.13) \qquad I(P_0, P \mid \mu) = \int_\Theta p \ln p \, d\mu + \ln \mu(\Theta_0)\,,$$

$$\tilde{I}(P_0, P) = \int_{\Theta_0} p \ln \frac{p}{1/\mu(\Theta_0)} \, d\mu + \int_{\Theta_0} p \ln \frac{p}{0} \, d\mu =$$

$$= I(P_0, P \mid \mu) + \begin{cases} \infty & \text{if } P(\Theta_0^c) > 0 \\ 0 & \text{if } P(\Theta_0^c) = 0\,. \end{cases}$$

If $P(\Theta_0^c) > 0$, then it is possible to replace P by a "truncated" version

$$P' = P(\cdot \mid \Theta_0) \quad \text{with} \quad p' = \frac{dP'}{d\mu} = \begin{cases} \dfrac{p(\theta)}{P(\Theta_0)} & \text{for } \theta \in \Theta_0 \\ 0 & \text{for } \theta \in \Theta_0^c\,. \end{cases}$$

Then

$$(4.14) \qquad I(P_0, P' \mid \mu) = \int_{\Theta_0} \frac{p}{P(\Theta_0)} \ln \frac{p}{P(\Theta_0)} \, d\mu + \ln \mu(\Theta_0) =$$

$$= \frac{1}{P(\Theta_0)} \int_{\Theta_0} p \ln p \, d\mu + \ln \frac{\mu(\Theta_0)}{P(\Theta_0)} = \tilde{I}(P_0, P')\,.$$

Thus we have proved the following

Theorem 4.3. If the prior probability P_0 is uniform on a subset $\Theta_0 \subset \Theta$ with respect to a fixed measure μ dominating P, then $I(P_0, P \mid \mu)$ is given by (4.13) and $\tilde{I}(P_0, P)$ is infinite unless $P(\Theta_0) = 1$ in which case it equals $I(P_0, P \mid \mu)$. If $P(\Theta_0) < 1$, then $I(P_0, P(\cdot \mid \Theta_0) \mid \mu) = \tilde{I}(P_0, P(\cdot \mid \Theta_0))$ are given by (4.14) and

$$\lim_{P(\Theta_0) \to 1} \tilde{I}(P_0, P(\cdot \mid \Theta_0)) = I(P_0, P \mid \mu)\,.$$

Thus, in the case the prior distribution P_0 is uniform on $\Theta_0 \subset \Theta$ we can say that $\bar{I}(P_0, P)$ and $I(P_0, P \mid \mu)$ are either equal (if $P(\Theta_0) = 1$), or mutually near, provided $P(\Theta_0)$ is near to 1. Theoretical models where $P(\Theta_0)$ is neither equal nor near to 1 are hardly of any practical importance. Thus we can conclude that in models with uniform prior distributions there is no or a negligible difference between the methods of quantification of information given by (3.12) and (3.13). It is true more, namely, that if the two methods are applied many times in any model then, in average, there is no difference between the information measures which both converge to a Shannon information. In the next section we specify a statistical model of measurements, which allows us to describe this result exactly.

5. STATISTICAL MODEL OF MEASUREMENT

In the following section we continue the analysis of the two concepts of information given by (3.12) and (3.13), but the initial simple stochastic model (P_0, P) we replace by a statistical model of measurement $(P_0, \{P_{\theta,\sigma} : \theta \in \Theta, \sigma \in \Xi\}, T)$. Here $(\Theta, \mathscr{T})$, $(\Xi, \mathscr{S})$ are abstract measurable spaces, $P_{\theta,\sigma}$ probabilities on an abstract measurable space $(X, \mathscr{X})$ $(P_{\theta,\sigma}(A)$ supposed to be $\mathscr{T} \times \mathscr{S}$-measurable from each $A \in \mathscr{X})$ and T a measurable mapping from $(X, \mathscr{X})$ into $(\Theta, \mathscr{T})$. Here $(X, \mathscr{X})$ is interpreted as an observation (sample) space, $(\Theta, \mathscr{T})$ as a parametric space, P_0 as a prior probability on $(\Theta, \mathscr{T})$, $P_{\theta,\sigma}$ as a sample probability depending on the unknown parameter $\theta \in \Theta$ and on a nuisance parameter $\sigma \in \Xi$ and T as a fixed estimator of θ. Finally, we suppose that the system of probabilities $\{P_{\theta,\sigma} : \theta \in \Theta, \sigma \in \Xi\}$ is uniformly dominated by a σ-finite measure ν on $(X, \mathscr{X})$ and P_0 by a σ-finite measure μ on $(\Theta, \mathscr{T})$. Measurement of θ is supposed to be given by the value $\hat{\theta} = T(x) \in \Theta$ depending on an empirical evidence $x \in X$, which is supposed to be realisation of a random sample $\xi \equiv \equiv (X, \mathscr{X}, P_{\theta,\sigma})$ the distribution $P_{\theta,\sigma}$ of which depends on both the unknown true value of θ and on a nuisance parameter σ. Obviously, the nuisance parameter statistically disappears as soon as Ξ is a one-point set.

In every statistical model of measurement $(P_0, \{P_{\theta,\sigma} : \theta \in \Theta, \sigma \in \Xi\}, T)$ we specify the pair (P_0, P) by two possible ways:

(i) $P = P_{x,\sigma}$, where $P_{x,\sigma}$ is the *posterior probability* on $(\Theta, \mathscr{T})$ corresponding to a nuisance parameter $\sigma \in \Xi$ and a concrete sample $x \in X$. The posterior probability is defined by the well-known Bayes formula applied for a fixed $\sigma \in \Xi$ to the prior probability P_0 on $(\Theta, \mathscr{T})$ and to conditional probabilities $\{P_{\theta,\sigma} : \theta \in \Theta\}$, i.e.

$$\frac{\mathrm{d}P_{x,\sigma}}{\mathrm{d}\mu} = p_\sigma(\theta \mid x) = \frac{p_0(\theta)\, p_\sigma(x \mid \theta)}{p_\sigma(x)}, \tag{5.1}$$

where

(5.2) $$p_\sigma(x) = \int_\Theta p_o(\theta)\, p_\sigma(x \mid \theta)\, \mathrm{d}\mu(\theta)$$

and

(5.3) $$p_o = \frac{\mathrm{d}P_0}{\mathrm{d}\mu}\,, \qquad p_\sigma(x \mid \theta) = \frac{\mathrm{d}P_{\theta,\sigma}}{\mathrm{d}\nu}\,.$$

If Ξ is a one-point set, then index σ can to omitted in the formulas above. In the opposite case, when σ is not known, we can specify $P = P_{x,S(x)}$, where

(5.4) $$S : X \to \Xi$$

is a suitable (e.g. unbiased minimum variance when $\Xi \subset E_s$) estimator of σ based on the empirical evidence $x \in X$. In this paper, however, we do not analyse this variant of the measurement model, and we restrict ourselves to $P = P_{x,\sigma}$ with fixed (and possibly unknown) $\sigma \in \Xi$.

(ii) Let $\Pi_{\theta,\sigma}$ denotes probability distribution of the random variable $T(\xi)$ on its sample space $(\Theta, \mathscr{T})$ i.e. let $\Pi_{\theta,\sigma} = P_{\theta,\sigma}T^{-1}$ for $\theta \in \Theta$, $\sigma \in \Xi$. Since estimates $T(x)$, $x \in X$, respresenting a final measurement inference concerning unknown $\theta \in \Theta$ are submitted to random errors (it is not true, that $P_{\theta,\sigma}(T(\xi) = \theta) = 1$), the information in the measurement $(P_0, \{P_{\theta,\sigma} : \theta \in \Theta, \sigma \in \Xi\}, T)$ may be based on the class $\{\Pi_{\theta,\sigma} : : \theta \in \Theta, \sigma \in \Xi\}$ of distributions of $T(\xi)$ as well. A measurement is more informative if $\Pi_{\theta,\sigma}$ are concentrated in close neighbourhoods of θ (uniformly with respect to all θ, σ) than a measurement where $\Pi_{\theta,\sigma}$ are almost uniform on Θ for all θ, σ.

Since individual outcomes $x \in X$ allow to approximate the unknown pair θ, σ by $T(x)$, $S(x)$ (see (5.4)), in regular measurement models we can furnish each $x \in X$ with a posterior probability $\Pi_{T(x),S(x)}$. This is why the second way we adopt for specification of the posterior probability P in the pair (P_0, P) is $P = \Pi_{T(x),S(x)}$. We call this probability (T, S)-*generated posterior probability*.

Remark. Probability P_0 in the pair (P_0, P) considered in Sec. 4 is frequently rather a subjective probability than a probability governing realizations of the unknown θ in repeated realizations (e.g. P_0 is considered to be uniform if no prior empirical evidence conserning θ has been collected while, in fact, θ may actually be distributed by a highly selective probability P_0'). If this happens, then the application of the P_0 in the Bayes formula (5.1) is not justified and, consequently, the probablities $P = P_{x,\sigma}$ or $P_{x,S(x)}$ considered in (i) are not realistic. The method (i) is thus applicable only if P_0 is based on a prior measurement (in this case the measurement we consider is a second stage of a two-stage measurement) and P_0 well approximates the distribution of θ in independently repeated measurements. If these conditions are not satisfied, (ii) should be prefered to (i). In fact, (ii) has already been applied in simple concrete situations as a standard method of specification of P but, however, without attempts to formalize and generalize it (see [1] and references given there).

Such generalization and formalization has been developed in mathematical statistics in a connection with maximum likelihood estimators T. If the statistical model is free of nuisance parameters then the well-known asymptotic normality yields approximation $\Pi_\theta \doteq No(\theta, 1/I(\theta))$ for all $\theta \in \Theta$, where $I(\theta)$ is the Fisher information of the model. In this situation, each sample $x \in X$ is traditionally used to approximate the unknown expectation $\mathsf{E}_\theta T \doteq \theta$ by $T(x)$ and the unknown variance $\mathsf{D}_\theta T \doteq 1/I(\theta)$ by $1/I(T(x))$ (see e.g. [16]). In this sense we thus can say that the unknown Π_θ is approximated by $\Pi_{T(x)}$. In the present paper we extend this idea beyond the scope of maximum likelihood estimators as well as beyond the scope of extremely large sample size. For a deeper analysis we refer to Sec. 7 below.

6. $I(P_0, P|\mu), \tilde{I}(P_0, P)$ AND SHANNON INFORMATION

Consider the statistical measurement model $(P_0, \{P_{\theta,\sigma} : \theta \in \Theta, \sigma \in \Xi\}, T)$ of the preceding section with P_0 actually governing realizations of θ and, according to (i), select $P = P_{x,\sigma}$. Suppose now that the information $I(P_0, P_{x,\sigma}|\mu)$ or $\tilde{I}(P_0, P_{x,\sigma})$ is repeatedly used N times in mutually independent measurements described by the same statistical models $(P_0, \{P_{\theta,\sigma} : \theta \in \Theta, \sigma \in \Xi\}, T)$. Denote by $\theta_1, \ldots, \theta_N$ parameters from Θ and by $x_1, \ldots, x_N$ observations from X realized in these N measurements. Since $\theta_1, \ldots, \theta_N$ are supposed to be realizations of independent random variables with common distribution P_0, $x_1, \ldots, x_N$ are realizations of independent random variables with common density $p_\sigma(x)$ given in (5.2). Now we can formulate the main result of this section.

Theorem 6.1. With probability 1 it holds

$$\lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} I(P_0, P_{x_i,\sigma} \mid \mu) = \lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} \tilde{I}(P_0, P_{x_i,\sigma}) = I(\sigma)$$

where

$$I(\sigma) = \int_\Theta \int_X p_0(\theta)\, p_\sigma(x \mid \theta) \ln \frac{p_\sigma(x \mid \theta)}{p_\sigma(x)}\, d\mu(\theta)\, d\nu(x) \tag{6.2}$$

is the Shannon information in an observation from X concerning the unknown parameter.

Proof. By the strong law of large numbers it will suffice to prove, that

$$\int_X \tilde{I}(P_0, P_{x,\sigma})\, p_\sigma(x)\, d\nu(x) = I(\sigma) \tag{6.2}$$

$$\int_X I(P_0, P_{x,\sigma} \mid \mu)\, p_\sigma(x)\, d\nu(x) = I(\sigma)\,. \tag{6.3}$$

By (3.12) and (5.1)

$$\tilde{I}(P_0, P_{x,\sigma}) = \int_\Theta \frac{p_0(\theta)\, p_\sigma(x \mid \theta)}{p_\sigma(\theta)} \ln \frac{p_\sigma(x \mid \theta)}{p_\sigma(x)} \, d\mu(\theta)$$

so that (6.2) follows from (6.1). By (3.13) and (5.1)

$$I(P_0, P_{x,\sigma} \mid \mu) = \int_\Theta \frac{p_0(\theta)\, p_\sigma(x \mid \theta)}{p_\sigma(x)} \ln \frac{p_0(\theta)\, p_\sigma(x \mid \theta)}{p_0(x)} \, d\mu(\theta) -$$

$$- \int_\Theta p_0(\theta) \ln p_0(q\theta) \, d\mu(\theta) =$$

$$= \tilde{I}(P_0, P_{x,\sigma}) + \int_\Theta \frac{p_0(\theta)\, p_\sigma(x \mid \theta)}{p_\sigma(x)} \ln p_0(\theta) \, d\mu(\theta) -$$

$$- \int_\Theta p_0(\theta) \ln p_0(\theta)\, \mu(\theta) .$$

Since

$$\int_X \left[\int_\Theta \frac{p_0(\theta)\, p_\sigma(x \mid \theta)}{p_\sigma(x)} \ln p_0(\theta) \right] p_0(\theta) \, d\mu(\theta) \, d\nu(\theta) = \int_\Theta p_0(\theta) \ln p_0(\theta) \, d\mu(\theta) ,$$

(6.3) holds as well, and the theorem is proved.

Thus, in an average, there is no difference between information (3.12) and (3.13), if they are applied repeatedly many times to the same measurement, provided a realistic distribution of θ is used as the prior probability P_0 and the posterior probability P is specified by (i) in Sec. 5.

7. $I(P_0, P \mid \mu)$, $\tilde{I}(P_0, P)$ AND FISHER INFORMATION

Suppose that in a statistical model of measurement $(P_0, \{P_{\theta,\sigma} : \theta \in \Theta, \sigma \in \Xi\}, T)$ P is specified as a (T, S)-generated posterior probability $\Pi_{T(x),S(x)}$ for $x \in X$. In a completely abstract case one cannot say much about the information $I(P_0, \Pi_{(Tx),S(x)} \mid \mu)$, $\tilde{I}(P_0, \Pi_{T(x),S(x)})$ defined by (3.12), (3.13). Suppose therefore that $\Theta = E_1$, Ξ is an interval of E_1 (the results below can easily be extended by truncation to Θ being an interval of E_1 and they can also be extended to the more general cases $\Theta \subset E_r$, $\Xi \subset E_s$). Suppose further that μ is the Lebesgue measure in E_1, $(X, \mathscr{X}) = (E_n, \mathscr{B}_n)$, ν be the Lebesgue measure on $(E_n, \mathscr{B}_n)$ and

$$p_\sigma(x \mid \theta) = q_\sigma(x_1 \mid \theta) \dots q_\sigma(x_n \mid \theta) \tag{7.1}$$

for every $x = (x_1, \dots, x_n) \in E_n$ and $\theta \in E_1$, $\sigma \in \Xi$.

Thus the statistical model of measurement is now specified by $(p_0, \{q_\sigma(\cdot \mid \theta) : : \theta \in E_1, \sigma \in \Xi\}, T_n)$, where p_0, $q_\sigma(\cdot \mid \theta)$ are Lebesgue measurable probability densities on E_1, $T_n : E_n \to E_1$ an arbitrary estimator of θ on the basis of samples $x = (x_1, \ldots, x_n)$ which are realizations of random vectors $\xi = (\xi_1, \ldots, \xi_n)$ with sample probability density (7.1). While we suppose that p_0 is arbitrary, we restrict ourselves to such models, for which the class of densities $\{q_\sigma(\cdot \mid \theta) : \theta \in E_1, \sigma \in \Xi\}$ is regular in the sense specified in Chap. 5 of Rao [14] or XV. 4. of Anděl [15], so that the Fisher information

$$(7.2) \qquad I(\theta, \sigma) = \int_{E_1} \frac{\left(\frac{\partial}{\partial \theta} q_\sigma(x \mid \theta)\right)^2}{q_\sigma(x \mid \theta)} \, \mathrm{d}x$$

in a sample from E_1 concerning the value θ for the given nuissance parameter value σ exists and is positive for every $\theta \in E_1$, $\sigma \in \Xi$. The final assumption concerning our measurement model is that

$$(7.3) \qquad \lim_{n \to \infty} \sqrt{(n)}\,(T_n(\xi_1, \ldots, \xi_n) - \theta) = No(0, 1/I(\theta, \sigma)) \quad \text{for each} \quad \theta \in E_1, \sigma \in \Xi$$

in the sense of convergence of distributions. This allows to write for large n

$$(7.4) \qquad \Pi_{\theta,\sigma} \doteq No(\theta, 1/n\, I(\theta, \sigma)) \quad \text{for each} \quad \theta \in E_1, \sigma \in \Xi\,.$$

Conditions under which (7.4) holds under the assumptions given above are generally mild, independently of whether T_n is a maximum likelihood estimator, or Bayes estimator

$$(7.5) \qquad T_n(x) = \int_{E_1} \theta \, p_\sigma(\theta \mid x) \, \mathrm{d}\theta$$

or a minimum variance unbiased estimator respectively. They are given e.g. in § 5.5 of Zacks [5]. We shall say that a statistical measurement model $(P_0, \{p_\sigma(\cdot \mid \theta) : : \theta \in E_1, \sigma \in \Xi\}, T_n)$ is strongly regular, if all the conditions stated above are satisfied.

Let now $S_n : E_n \to \Xi$ be a reasonable estimator of σ, e.g. minimum variance unbiased estimator in a strongly regular model. In accordance with (7.4) we can write for $\Pi_{T_n(x),S_n(x)}$ in (ii) of Sec. 5

$$(7.6) \qquad \Pi_{T_n(x),S_n(x)} \doteq No(T_n(x), 1/J_n(x))\,,$$

where

$$(7.7) \qquad J_n(x) = n\, I(T_n(x), S_n(x))\,.$$

Note that if we can find out a minimum variance unbiased estimator $\hat{I}_n : : E_n \to [0, \infty)$ of the parametric function $I(\theta, \sigma)$ or $1/I(\theta, \sigma)$ then (7.7) can be replaced by

$$(7.8)\qquad J_n(x) = n\,\hat{I}_n(x) \quad \text{or} \quad n/\hat{I}_n(x) \quad \text{respectively}.$$

We investigate this possibility only in an example below.

On the basis of (7.6) we proceed now with investigation of the following measurement information (see (3.12), (3.13)):

$$(7.9)\qquad I(P_0, No(T_n(x), 1/J_n(x))) = \int_{E_1} No(T_n(x), 1/J_n(x)) \ln No(T_n(x), 1/J_n(x))\, d\theta -$$

$$- \int_{E_1} p_0 \ln p_0 \, d\theta \doteq I(P_0, \Pi_{T_n(x), S_n(x)}),$$

$$(7.10)\qquad \bar{I}(P_0, No(T_n(x), 1/J_n(x))) = \int_{E_1} No(T_n(x), 1/J_n(x)) \ln \frac{No(T_n(x), 1/J_n(x))}{p_0}\, d\theta =$$

$$\doteq \bar{I}(P_0, \Pi_{T_n(x), S_n(x)}).$$

In analysis of this information we shall obviously need the following two lemmas:

Lemma 7.1. For every $\mu \in E_1$, $\sigma > 0$ it holds

$$(7.11)\qquad \int_{E_1} No(\mu, \sigma^2) \ln No(\mu, \sigma^2)\, d\theta = -\tfrac{1}{2} \ln (2\pi e \sigma^2),$$

$$(7.12)\qquad \int_{E_1} No(\mu, \sigma^2) \ln \frac{No(\mu, \sigma^2)}{No(\mu_0, \sigma_0^2)}\, d\theta = \frac{1}{2}\left(\frac{\mu - \mu_0}{\sigma_0}\right)^2 + \frac{1}{2}\left[\left(\frac{\sigma}{\sigma_0}\right)^2 - 1\right] + \ln \frac{\sigma_0}{\sigma} \geqq$$

$$\geqq \frac{1}{2}\left(\frac{\mu - \mu_0}{\sigma_0}\right)^2.$$

Proof. Simple integration.

Lemma 7.2. For every $\mu_0, \mu, \theta_0, \theta_1 \in E_1$, $\theta_0 < \theta_1$, and $\sigma_0, \sigma > 0$ it holds

$$(7.13)\qquad I(No(\mu_0, \sigma_0^2), No(\mu, \sigma^2)) = \ln \frac{\sigma_0}{\sigma},$$

$$(7.14)\qquad I(U(\theta_0, \theta_1), No(\mu, \sigma^2)) = \frac{1}{2} \ln \frac{(\theta_1 - \theta_0)^2}{2\pi \sigma^2},$$

$$(7.15)\qquad \bar{I}(No(\mu_0, \sigma_0^2), No(\mu, \sigma^2)) = \frac{1}{2}\left(\frac{\mu - \mu_0}{\sigma_0}\right)^2 + \frac{1}{2}\left[\left(\frac{\sigma}{\sigma_0}\right)^2 - 1\right] + \ln \frac{\sigma_0}{\sigma},$$

$$(7.16)\qquad I\,(U(\theta_0, \theta_1), No(\mu, \sigma^2 \mid (\theta_0, \theta_1))) = \bar{I}(U(\theta_0, \theta_1), No(\mu, \sigma^2 \mid (\theta_0, \theta_1))) =$$

$$= \frac{1}{2} \ln \frac{1}{2\pi e\sigma^2} + \frac{1}{2\left[\Phi\left(\frac{\theta_1 - \mu}{\sigma}\right) - \Phi\left(\frac{\theta_0 - \mu}{\sigma}\right)\right]} \cdot$$

$$\cdot \left[\frac{\theta_1 - \mu}{\sigma^2} e^{-(\theta_1-\mu)^2/2\sigma^2} - \frac{\theta_0 - \mu}{\sigma^2} e^{-(\theta_0-\mu)^2/2\sigma^2}\right] + \ln \frac{\theta_1 - \theta_0}{\Phi\left(\frac{\theta_1 - \mu}{\sigma}\right) - \Phi\left(\frac{\theta_0 - \mu}{\sigma}\right)},$$

where $No(\mu, \sigma^2 \mid (\theta_0, \theta_1))$ denotes the conditional normal probability under the condition $\theta \in (\theta_0, \theta_1)$. If

(7.17) $$\theta_1 > \mu + 3\sigma \quad \text{and} \quad \theta_0 < \mu - 3\sigma$$

then

(7.18) $$I\ (U(\theta_0, \theta_1), No(\mu, \sigma^2 \mid (\theta_0, \theta_1))) = \tilde{I}(U(\theta_0, \theta_1), No(\mu, \sigma^2 \mid (\theta_0, \theta_1))) \doteq$$

$$\doteq \frac{1}{2} \ln \frac{(\theta_1 - \theta_0)^2}{2\pi e\sigma^2}.$$

Proof. (7.13) follows from (3.13) and (7.11), (7.14) from (4.13) and (7.11), (7.15) from (3.12) and (7.12), (7.16) from (4.14), and (7.11) an (7.18) under the condition (7.17) holds according to Theorem 4.2.

Relations of the information $I(P_0, P \mid \mu)$, $\tilde{I}(P_0, P)$ to the Fisher information $I(\theta, \sigma)$ in a statistical measurement model are now summarized in the following

Theorem 7.1. In a strongly regular measurement model we have for large n

(7.19) $$I(P_0, \Pi_{T_n(x),S_n(x)}) \doteq \frac{1}{2} \ln \frac{n\, I(T_n(x), S_n(x))}{2\pi e} - \int_{E_1} p_0 \ln p_0 \, d\theta ,$$

$$\tilde{I}(P_0, \Pi_{T_n(x),S_n(x)}) \doteq \frac{1}{2} \ln \frac{n\, I(T_n(x), S_n(x))}{2\pi e} - \int_{E_1} No(T_n(x), 1/n\, I(T_n(x), S_n(x))) \,.$$

(7.20) $$. \ln p_0 \, d\theta ,$$

where $I(\theta, \sigma)$ is given by (7.2). Further,

(7.21) $$I(No(\mu_0, \sigma^2), \Pi_{T_n(x),S_n(x)}) \doteq \tfrac{1}{2} \ln \left[\sigma_0^2 n\, I(T_n(x), S_n(x))\right],$$

(7.22)

$$\tilde{I}(No(\mu_0, \sigma_0^2), \Pi_{T_n(x),S_n(x)}) \doteq \frac{1}{2}\left[\left(\frac{T_n(x) - \mu_0}{\sigma_0}\right)^2 + \frac{1}{\sigma_0^2 n\, I(T_n(x), S_n(x))} - 1\right] +$$

$$+ \ln [\sigma_0 I(T_n(x), S(x))] \doteq \frac{1}{2}\left[\left(\frac{T_n(x) - \mu_0}{\sigma_0}\right)^2 - 1\right] + I(No(\mu_0, \sigma_0^2), \Pi_{T_n(x), S_n(x)})$$

and, if $P_0 = U(\theta_0, \theta_1)$, then

$$(7.23) \qquad I(U(\theta_0, \theta_1), \Pi_{T_n(x),S_n(x)}) \doteq \frac{1}{2} \ln \frac{(\theta_1 - \theta_0)^2 \, n \, I(T_n(x), S_n(x))}{2\pi \, e} \doteq$$

$$\doteq I(U(\theta_0, \theta_1), \Pi_{T_n(x),S_n(x)}(\cdot|(\theta_0, \theta_1))) \doteq \bar{I}(U(\theta_0, \theta_1), \Pi_{T_n(x),S_n(x)}(\cdot \mid (\theta_0, \theta_1))) .$$

Proof. See (7.7), (7.9), (7.10) and Lemma 7.1 and 7.2.

This theorem implies that, when a strongly regular model with large n is considered, then the information (3.12) or (3.13) is a monotone function of the Fisher information of the model and they both are mutually close.

Let us now consider the example studied in the end of Sec. 1, where the statistical measurement model is given, for a sample size n, by the triple

$$\left(U(0, 1), \left\{ q(x \mid \theta) = \left\langle \begin{matrix} \chi_{\{0,1\}}(x)\theta^x(1 - \theta)^{1-x} & \text{for} \quad \theta \in (0, 1) \\ \text{arbitrary} & \text{for} \quad \theta \notin (0, 1) \end{matrix} \right. \right\}, \; T_n(x) = \frac{\Sigma x + 1}{n + 2}\right).$$

It is easy to verify that this model is strongly regular. As we said in Sec. 1, the Fisher information $I(\theta)$ is equal $1/[\theta(1 - \theta)]$ and P defined on by (i) or (ii) in Sec. 5 for a sample size n is given by

$$(7.24) \qquad P = P_x = Beta\,(\Sigma x + 1, (n - \Sigma x + 1)) \quad \text{or} \quad P = \Pi_{T_n(x)} =$$

$$= No\left(\frac{\Sigma x + 1}{u + 2}, \frac{(\Sigma x + 1)(u - \Sigma x + 1)}{u(u + 2)^2}\right)$$

respectively. It is easy to verify that the parameter θ or parametric function $1/I(\theta)$ in $\Pi_\theta = No(\theta, 1/n\, I(\theta))$ possesses minimum variance estimators $\hat{\theta}_n(x) = \Sigma x/n$ or $\hat{I}_n(x) = \Sigma x(n - \Sigma x)/n(n - 1)$, respectively. Thus, if the actual prior distribution governing the realizations of θ is not uniform but e.g. *Beta* (a, b), where nothing is known about $a > 1$, $b > 1$, then there is a ground on which we can prefer the $(\hat{\theta}_n, \hat{I}_n)$-generated posterior probability

$$(7.25) \qquad P = \Pi_{\hat{\theta}_n(x), \hat{I}_n(x)} \doteq No\left(\frac{\Sigma x}{n}, \frac{\Sigma nx(n - \Sigma x)}{n^2(n - 1)}\right)$$

to the $\Pi_{T_n(x)}$ defined in (7.22).

Compute now the information resulting from the posterior probabilities generated by T_n itself and by the pair $\hat{\theta}_n, \hat{I}_n$. Using (7.21) we obtain the amount of information in a measurement $x = (x_1, \ldots, xn) \in E_n$ in the following form

$$(7.26)\qquad I(U(0,1), \Pi_{T_n(x)}) \doteq \frac{1}{2}\ln\frac{n(n+2)^2}{2\pi\, e(\Sigma x+1)(n-\Sigma x+1)}$$

which is also equal the corresponding information $\bar{I}$ or the "truncated" ifmormation I. Analogically we get

$$(7.27)\qquad I(U(0,1), \Pi_{\theta_n(x),t_n(x)}) \doteq \frac{1}{2}\ln\frac{n^2(n-1)}{2\pi\, e\Sigma x(n-\Sigma x)}\,.$$

Thus in this example both definitions (3.12), (3.13) lead to the same expression for measurement information and the two expressions (7.26), (7.17) obtained on the account of different interpretations of the prior probability $P_0 = U(0, 1)$ asymptotically yield the same quantity

$$I(x_1, \ldots, x_n) = \frac{1}{2}\ln\frac{n}{2\pi\, e\Sigma x/n(1-\Sigma x/n)} \geqq \frac{1}{2}\ln\frac{2n}{\pi\, e}\,.$$

It is interesting to compare this result with a Lemma proved in [9], according to which $H(B_i(\theta, n)) = \frac{1}{2}\,.\ln(2\pi\, e\theta(1-\theta)) + o(1)$ when $n \to \infty$. By this Lemma, with probability one,

$$(7.28)\qquad \lim_{n\to\infty}\frac{I(\xi_1,\ldots,\xi_n) + H(\xi_1+\ldots+\xi_n)}{\ln n} = 1 \quad \text{for every} \quad \theta \in (0, 1)\,,$$

where $\xi_1 + \ldots + \xi_n$ is a sufficient statistic for $\theta \in (0, 1)$ and $H(\xi_1 + \ldots + \xi_n)$ denotes the Shannon entropy of this statistic. Equivalent formulation is that the information $I(x_1, \ldots, x_n)$ in a large number n independent measurements $x_1, \ldots, x_n$ of a binomial parameter $\theta \in (0, 1)$ is equal $\ln n - H(Bi(\Sigma x/n, n))$. This result somewhat resembles the fact valid for Shannon information, namely, that the Shannon information is always a difference between a prior and a posterior entropy.

8. A NORMAL MEASUREMENT MODEL

Let us now consider statistical measurement model

$$(8.1)\qquad (No(\mu_0, \sigma_0^2), \{No(\theta, \sigma^2) : \theta \in E_1, \sigma > 0\}, T_n)\,,$$

where T_n is a Bayes estimator of θ with respect to the prior probability $No(\mu_0, \sigma_0^2)$ and with respect to the quadratic loss function. It is well known that this model with a nuisance parameter σ is strongly regular and that its Fisher informationis given as

$$(8.2)\qquad I(\theta, \sigma) = I(\sigma) = \frac{1}{\sigma^2}\,.$$

It is easy to verify that the posterior probability $P_{x,\sigma}$ for $x = (x_1, \ldots, x_n) \in E_n$ is given by

(8.3) $$P_{x,\sigma} = No\left(\frac{\mu_0 + (\sigma_0/\sigma)^2\,\Sigma x}{1 + (\sigma_0/\sigma)^2\,n}, \frac{\sigma_0^2}{1 + (\sigma_0/\sigma)^2\,n}\right)$$

so that

(8.4) $$T_n(x) = \frac{\mu_0 + (\sigma_0/\sigma)^2\,\Sigma x}{1 + (\sigma_0/\sigma)^2\,n} \quad \text{for every} \quad x \in E_n\,.$$

The unknown σ^2 we can reasonably estimate by the unbiased estimator

(8.5) $$S_n^2(x) = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \Sigma x/n)^2\,.$$

Note that in this model we need not to apply the approximation (7.4) since we know the exact distribution of $T_n(\xi_1, \ldots, \xi_n)$, namely,

(8.6) $$\Pi_{\theta,\sigma} = No\left(\frac{\mu_0 + (\sigma_0/\sigma)^2\,n\theta}{1 + (\sigma_0/\sigma)^2\,n}, \frac{\sigma^2}{n}\right).$$

Now we shall apply (7.12) and (7.15) to calculate the information in a sample $x \in E_n$ given by formulas (3.12), (3.13) for both methods (i), (ii) in Sec. 5 for specification the posterior probability.

Theorem 8.1. If the posterior probability $P = P_{x,\sigma}$ is given by (8.3), then for all $n = 1, 2, \ldots$ the corresponding information $I(x, \sigma)$, $\tilde{I}(x, \sigma)$ given by (3.13), (3.12) is of the following form

(8.7) $$I(x, \sigma) = I(No(\mu_0, \sigma_0^2), P_{x,\sigma}) = \frac{1}{2}\ln\left[1 + \left(\frac{\sigma_0}{\sigma}\right)^2 n\right],$$

(8.8)
$$\tilde{I}(x, \sigma) = \frac{1}{2}\left[\left(\frac{\Sigma x - \mu_0 n}{1 + (\sigma_0/\sigma)^2\,n}\right)^2 \frac{(\sigma_0/\sigma)^4}{\sigma_0^2} - \frac{(\sigma_0/\sigma)^2\,n}{1 + (\sigma_0/\sigma)^2\,n} + \ln\left(1 + \left(\frac{\sigma_0}{\sigma}\right)^2 n\right)\right].$$

For (T_n, S_n)-generated posterior probability $P = \Pi_{T_n(x),S_n(x)}$ the information $I(x)$, $\tilde{I}(x)$ defined by (3.13), (3.12) is for all $n = 1, 2, \ldots$ given by

(8.9) $$I(x) = \frac{1}{2}\ln\frac{n\sigma_0^2}{Sn(x)^2} \quad (\text{see } (8.5))\,,$$

(8.10) $$\tilde{I}(x) = \frac{1}{2}\left[\left(\frac{T_n(x) - \mu_0}{1 + (\sigma_0/S_n(x))^2\,n}\right)^2 \frac{n^2(\sigma_0/S_n(x))^4}{\sigma_0^2} + \frac{S_n(x)^2}{n\sigma_0^2} - 1 + \ln\frac{n\sigma_0^2}{S_n(x)^2}\right]$$

(see (8.4), (8.5)).

Now we shall discuss the results of this theorem. First, for large n, (8.8) and (8.10) become less complex, namely

$$(8.11)\qquad \tilde{I}(x, \sigma) \doteq \frac{1}{2}\left[\left(\frac{\Sigma x/n - \mu_0}{\sigma_0}\right)^2 - 1 + \ln\left(1 + \left(\frac{\sigma_0}{\sigma}\right)^2 n\right)\right],$$

$$(8.12)\qquad \tilde{I}(x) \doteq \frac{1}{2}\left[(\Sigma x/n - \mu_0)^2 - 1 + \ln\frac{n\sigma_0}{S_n(x)^2}\right].$$

This can be understood on the basis that, for large n, $T_n(x)$ in (8.4) becomes approximately equal $\Sigma x/n$ (a relative weight of the prior evidence becomes negligible). If σ is bound to a subset $\Xi \subset E_1$ the maximum of which is much less than σ_0 (if the prior knowledge is weak compared to the posterior) then, for small n, (8.11) and (8.12) can be used while, for large n,

$$(8.13)\quad \tilde{I}(x, \sigma) \doteq \frac{1}{2}\left[\left(\frac{\Sigma x/n - \mu_0}{\sigma_0}\right)^2 - \frac{(\sigma_0/\sigma)^2\, n}{1 + (\sigma_0/\sigma)^2\, n} + \ln\left(1 + \left(\frac{\sigma_0}{\sigma}\right)^2 n\right)\right],$$

$$(8.14)\qquad \tilde{I}(x) \doteq \frac{1}{2}\left[\left(\frac{\Sigma x/n - \mu_0}{\sigma_0}\right)^2 + \frac{S_n(x)^2}{n\sigma_0^2} - 1 + \ln\frac{n\sigma_0^2}{S_n(x)^2}\right].$$

$\tilde{I}(x)$ can be calculated by (8.14) for all $n = 1, 2, \ldots$ and σ_0, σ provided the (T_n', S_n)-generated posterior probability is used with $T_n'(x) = \Sigma x/n$. This is realistic for example when $P_0 = No(\mu_0, \sigma_0^2)$ is not a "true" probability distribution of the parameter θ.

(Received May 12, 1979.)

REFERENCES

[1] K. Eckschlager, V. Štěpánek: Information Theory as Applied to Chemical Analysis. J. Willey, New York 1979.

[2] K. Eckschlager, I. Vajda: Amount of Information of Repeated Higher Precision Analyses. Coll. Czechoslov. Chem. Commun. *39* (1974), 3076.

[3] C. E. Shannon: A mathematical theory of communication. The Bell Syst. Techn. J. *27*, (1948). 379, 623.

[4] S. Kullback: Information Theory and Statistics. J. Willey, New York 1959.

[5] S. Zacks: Theory of Statistical Inference. J. Willey, New York 1972.

[6] I. A. Ibragimov, R. Z. Chasminskij: Asymptotic behaviour of generalized Bayes estimators (in Russian). DAN SSSR *194* (1970), 2, 257–260.

[7] I. A. Ibragimov, R. Z. Chasminskij: On limit behaviour of Bayes maximum likelihood estimators (in Russian). DAN SSSR *198* (1971), 3, 520–523.

[8] A. Perez: L'expérience et l'information puisée dans elle à l'aide des lois limites de la théorie des probabilités. Trans. 2nd Prague Conf. on Inform. Theory, ... NČSAV, Praha 1960.

[9] I. Vajda: Universal versus specific coding. Trans. 8th Prague Conf. on Inform. Theory, ... Academia, Praha 1978.

[10] I. Vajda: χ^αdivergence and generalized Fisher's information. Trans. 6th Prague Conf. on Inform. Theory, ... Academia, Praha 1975.

[11] A. Perez: Notions généralisées d'incertitude, d'entropie et d'information du point de vue de la théorie de martingales. Trans. 1st Prague Conf. on Inform. Theory, ... NČSAV, Praha 1957.

[12] M. M. Bongard: On the concept of "useful information" (in Russian). Problemy Kibernetiki *9* (1963).

[13] D. F. Kerridge: Inaccuracy and inference. J. Roy. Statist. Soc., Ser. *B 23* (1961), 184—194.

[14] C. R. Rao: Linear Statistical Inference and its Applications. J. Willey, New York 1965.

[15] J. Anděl: Základy matematické statistiky. SNTL, Praha 1978.

[16] D. R. Cox, D. V. Hinkley: Theoretical Statistics. Chapman & Hall, London 1974.

Ing. Igor Vajda, CSc., Ústav teorie informace a automatizace ČSAV (Institute of Information theory and Automation — Czechoslovak Academy of Sciences), Pod vodárenskou věží 4, 182 08 Praha 8. Czechoslovakia.

Ing. Karel Eckschlager, DrSc., Ústav anorganické chemie ČSAV (Institute of Anorganic Chemistry — Czechoslovak Academy of Sciences), 250 68 Řež u Prahy. Czechoslovakia.