On Relations between Informations, Entropies and Bayesian Decisions

Igor Vajda, Jana Zvárová

Abstract: The paper studies ϕ-entropies $H_\phi(X)$ of random messages X defined as maximal ϕ-informations $I_\phi(X;Y)$ about X in observations Y. We show that nonconcave entropies can be obtained in this manner. A class of power entropies $H_\alpha(X)$ parametrized by $\alpha \in \mathbb{R}$ is given which are concave or convex functions of distributions p_X for $\alpha \geq 0$ or $\alpha \leq 0$ respectively. The paper studies also the accuracy of estimation of the errors $e(X)$ of Bayesian decisions about the values of X by means of two special ϕ-entropies: the Shannon entropy $H_1(X)$ and the so-called quadratic entropy $H_2(X)$.

MSC 2000: 94A17, 62C10
Key words: ϕ-divergences, ϕ-informations, ϕ-entropies, Shannon entropy, quadratic entropy, Bayes error

1 Introduction and basic concepts

The models of data and observations in the digital world are usually discrete. Therefore we are interested in random observations with true distributions $p = (p(i) : i \in \mathcal{I})$ and hypothetical distributions $q = (q(i) : i \in \mathcal{I})$ where $\mathcal{I}$ is finite. We drop the observation indices $i \in \mathcal{I}$ with $p(i) + q(i) = 0$, i. e. we suppose $p(i) + q(i) > 0$ for all $i \in \mathcal{I}$.

The divergence of distributions p, q is often expressed by the ϕ-divergence for ϕ from the class $\mathbf{\Phi}$ of real valued functions defined and convex on the interval $(0, \infty)$ which are strictly convex at $t = 1$ with $\phi(1) = 0$. Following Csiszár [1, 2] or Liese and Vajda [7], the ϕ-divergence of distributions p, q can be defined by formula

$$D_\phi(p\|q) = \sum_{i:q(i)>p(i)} q(i)\,\phi\left(\frac{p(i)}{q(i)}\right) + \sum_{i:q(i)<p(i)} p(i)\,\phi^*\left(\frac{q(i)}{p(i)}\right) \tag{1.1}$$

where $\phi^* \in \mathbf{\Phi}$ is conjugated to ϕ in the sense that for all $t \in (0, \infty)$

$$\phi^*(t) = t\phi\left(\frac{1}{t}\right),$$

and the (possibly infinite) values $\phi(0)$, $\phi^*(0)$ needed in (1.1) are obtained as limits of $\phi(t)$, $\phi^*(t)$ for $t \downarrow 0$. It is clear from (1.1) that $D_{\phi^*}(p\|q) = D_\phi(q\|p)$. As well known, the ϕ-divergences take on values between 0 and $\phi(0) + \phi^*(0)$ where $D_\phi(p, q) = 0$ if and only if $p = q$. For this and further properties see Liese and Vajda [7].

Supported by the MSMT grants MSM 0021620839 and 1M0572.

The most simple functions $\phi \in \Phi$ are $\phi_+(t) = (t-1)^+ = \max(t-1,0)$ and $\phi_-(t) = \phi_+^*(t) = (t-1)^- = -\min(t-1,0)$ leading to the *upper variation*

$$D_{\phi_+}(p\|q) = V_+(p\|q) = \sum_{i:p(i)>q(i)} (p(i) - q(i)) \tag{1.2}$$

and *lower variation*

$$D_{\phi_-}(p\|q) = V_-(p\|q) = \sum_{i:p(i)<q(i)} (q(i) - p(i)), \tag{1.3}$$

and their sum $\phi(t) = \phi_+(t) + \phi_-(t) = |t-1|$ which is self-conjugated in the sense $\phi^*(t) = \phi(t)$ and leads to the *total variation*

$$D_\phi(p\|q) = V(p\|q) = \sum_i |p(i) - q(i)|. \tag{1.4}$$

Well known class of ϕ-divergences parametrized by $\alpha \in \mathbb{R}$ is obtained from the power functions

$$\phi_\alpha(t) = \frac{t^\alpha - \alpha(t-1) - 1}{\alpha(\alpha-1)} \quad \text{for } \alpha(\alpha-1) \neq 0 \tag{1.5}$$

and their limits

$$\phi_1(t) = t\ln t - 1 + 1, \quad \phi_0(t) = -\ln t + t - 1 \tag{1.6}$$

which are conjugated by the rule $\phi_\alpha^*(t) = \phi_{1-\alpha}(t)$. In what follows we use the simplified notation $D_\alpha(p\|q) = D_{\phi_\alpha}(p\|q)$. The best known statistical divergence is

$$D_2(p\|q) = \sum_i \frac{p(i)^2}{q(i)} - 1 = \sum_i \frac{(p(i) - q(i))^2}{q(i)}. \tag{1.7}$$

Here and in the sequel the summands with $q(i) = 0$ in the denominator are assumed to be infinite. This divergence will be called *Pearson divergence*. We shall see that it is in some sense information-theoretically neutral and therefore it may be used as a basis of the standardized terminology where

$$D_4(p\|q) = \sum_i \frac{p(i)^4}{q(i)^3} - 1 \tag{1.8}$$

is a *double-Pearson divergence*, the classical information-theoretic divergence (Kullback divergence)

$$D_1(p\|q) = \sum_i p(i) \ln \frac{p(i)}{q(i)} \tag{1.9}$$

is a *half-Pearson divergence* and the squared Hellinger distance

$$D_{1/2}(p\|q) = 4\sum_i \left(\sqrt{p(i)} - \sqrt{q(i)}\right)^2 \tag{1.10}$$

is a *quarter-Pearson divergence.*

In his classical papers on information theory, Shannon introduced probability distributions $p_{X;Y}(i)$, $i \in \mathcal{I} = \mathcal{X} \times \mathcal{Y}$ as models for the situations where an $\mathcal{Y}$-valued observation Y informs about on $\mathcal{X}$-valued message X. As a measure of information he proposed a nonnegative quantity $I(X;Y)$ which is nothing but the half-Pearson divergence $D_1(p_{X,Y}\|p_X p_Y)$ between the joint distribution $p_{X,Y}$ of X, Y and the product $p_X p_Y$ of the marginal distributions

$$p_X(x) = \sum_y p_{X,Y}(x,y), \quad p_Y(y) = \sum_x p(x,y)$$

of X and Y on $\mathcal{X}$ and $\mathcal{Y}$. In other words, the *Shannon information* is

$$I_1(X;Y) = D_1(p_{X,Y}\|p_X p_Y) = \sum_{x,y} p(x,y) \ln \frac{p(x,y)}{p(x)\,p(y)} \tag{1.11}$$

(cf. (1.9)). Here and in the sequel we use the conventions

$$p(x,y) = p_{X,Y}(x,y), \quad p(x) = p_X(x) \quad \text{and} \quad p(y) = p_Y(y) \tag{1.12}$$

which are common in the literature on information theory. We see from (1.11) that the Shannon information is a nonnegative measure of association of the random variables X, Y which is equal zero if and only if X, Y are independent.

A similar measure of association was proposed much earlier by Pearson [9], namely the *mean square contingency*

$$I_2(X;Y) = D_2\left(p_{X,Y}\|p_X p_Y\right) = \sum_{x,y} \frac{(p(x,y) - p(x)\,p(y))^2}{p(x)\,p(y)} \tag{1.13}$$

(cf. (1.7) and (1.12)), used later as a basis in various criteria of statistical association (Cramér, Tschuprow and others). Höffding [5] proposed postulates for measures of association of random pairs (X,Y) based on the measure

$$V(X;Y) = V\left(p_{X,Y}\|p_X p_Y\right) = \sum_{x,y} |p(x,y) - p(x)\,p(y)| \tag{1.14}$$

(cf. (1.4) and (1.12)) called *Höffding coefficient* in Zvárová [13].

Motivated by these proposals and also by earlier papers of A. Rényi, Csiszár [3] and Zvárova [13] introduced the general *ϕ-information*

$$I_\phi(X;Y) = D_\phi\left(p_{X,Y}\|p_X p_Y\right) = \sum_{x,y} p(x)\,p(y)\,\phi\left(\frac{p(x,y)}{p(x)\,p(y)}\right) \tag{1.15}$$

(cf. (1.12)) where we used a simplified form of (1.1) since $p(x)\,p(y) > 0$ follows for all $x \in \mathcal{X}$, $y \in \mathcal{Y}$ from the assumption $p(x)\,p(y) + p(x,y) > 0$. In other words, ϕ-information is the ϕ-divergence of true distribution $p_{X,Y}$ and the hypothetic distribution $p_X p_Y$ which is true only if X and Y are independent.

As observed already by Shannon, the information $I_1(X;Y)$ is maximal if $Y = X$, i. e. if the observed variable is the message X itself. The amount of information $I_1(X;X)$ in the message X is the *Shannon entropy*

$$H_1(X) = -\sum_x p(x) \ln p(x) \quad (\text{in } \textit{nats}). \tag{1.16}$$

This is one element from the family of α-*entropies* defined for arbitrary distributions $p = p_X$ and all $\alpha \in \mathbb{R}$ by the formula

$$H_\alpha(p) \equiv H_\alpha(X) = \sum_x p(x)\, \psi_\alpha(p(x)) \tag{1.17}$$

where

$$\psi_\alpha(\pi) = \frac{1 - \pi^{\alpha-1}}{\alpha - 1}, \quad \pi \in (0,1], \quad \psi_\alpha(0) = \lim_{\pi \downarrow 0} \psi_\alpha(\pi) \tag{1.18}$$

with the limit

$$\psi_1(\pi) = -\ln \pi, \quad \pi \in (0,1], \quad \psi_1(0) = \infty \tag{1.19}$$

are decreasing measures of information in an event of probability $\pi \in [0,1]$ and $0\psi_\alpha(0) = 0$. Therefore the entropies $H_\alpha(X)$ are expected amounts of informations in the individual events $X = x$. If we normalize in (1.18) by $(\alpha - 1)\ln 2$ then the limit entropy $H_1(X)$ for $\alpha \to 1$ will differ from (1.16) by $\log_2$ at the place of $\ln = \log_e$, i. e. the information $H_1(X)$ will be in *bits* instead of *nats*.

In a slightly differently normalized form, the α-entropies with parameters $\alpha > 0$ were introduced by Havrda and Charvát [4]. As we shall see, interesting examples in addition to the Shannon entropy are

$$H_2(X) = 1 - \sum_x p(x)^2 \tag{1.20}$$

called *quadratic entropy* by Vajda [10] and

$$H_0(X) = (\#x : p(x) > 0) - 1 \tag{1.21}$$

indicating the number of possible messages which differ from the delivered one, which may be called *Hartley entropy* (cf. Morales et al [8]).

The α-entropies (1.17) belong to the class of ψ-entropies

$$H_\psi(p) \equiv H_\psi(X) = \sum_x p(x)\, \psi(p(x)) \tag{1.22}$$

for decreasing continuous functions $\psi(\pi)$ of variable $\pi \in (0,1]$ with $\psi(1) = 0$. These general entropies were studied in Vajda [11] who proved that the ψ-entropies satisfy standard desirable properties of information measures if $\pi\psi(\pi)$ ic concave on $(0,1)$. The α-entropies (1.17) are concave in this sense only for $\alpha \geq 0$.

In Section 2 we show that the non-concave α-entropies can be desirable information measures too. Section 3 studies relations between the entropies $H_1(X)$, $H_2(X)$ and the errors $e(X)$ of Bayes decisions about X.

2 Informations and entropies

Let us start this section with the formula

$$\sup_Y I_\phi(X;Y) = I_\phi(X;X) = H_{\tilde{\phi}}(X) \quad \text{(cf. (1.22))} \tag{2.1}$$

for $\tilde{\phi}(t) = \phi^*(t) + \phi(0)\,(1-t)$ proved by Zvárová [13]. This formula says that the ϕ-information obtained by observing a message X distributed by $p(x) = p_X(x)$ is given by the $\tilde{\phi}$-entropy

$$H_{\tilde{\phi}}(p) \equiv H_{\tilde{\phi}}(X) = \sum_x p(x)\,\phi^*(p(x)) + \phi(0)\,H_2(X) \quad \text{(cf. (1.22) and (1.20))}. \tag{2.2}$$

This formula and also the next assertion emphasize the prominent role of the quadratic entropy.

Assertion 2.1. The ϕ-informations $I_\phi(X;Y)$ corresponding to the simple functions $\phi = \phi_+$ and $\phi = \phi_-$ from $\mathbf{\Phi}$ achieve maxima given by the quadratic entropy, i. e.

$$H_{\tilde{\phi}_+}(X) = H_{\tilde{\phi}_-}(X) = H_2(X) \quad \text{(cf. (1.20))}. \tag{2.3}$$

Therefore the entropy $H_{\tilde{\phi}}(X)$ which maximizes the Höffding measure of information (1.14) is $2H_2(X)$.

Proof. For $\phi_+(t)$ we get $\phi_+(0) = 0$ and $\phi_+^*(t) = \phi_-(t) = 1-t$ for all $t \in [0,1]$. Therefore $\tilde{\phi}_+(t) = \psi_2(t)$ from (1.18) so that $H_{\tilde{\phi}_+}(X) = H_2(X)$. For ϕ_- we get $\phi_-(0) = 1$ and $\phi_-^*(t) = \phi_+(t) = 0$ for all $t \in [0,1]$ so that again $\tilde{\phi_-}(t) = \psi_2(t)$ and the rest is as above. The last statement follows from the fact that if $\phi(t) = |t-1|$ then $\tilde{\phi}(t) = \tilde{\phi}_+(t) + \tilde{\phi}_-(t)$. □

In the next assertion we are interested in the entropies

$$\tilde{H}_\alpha(p) \equiv \tilde{H}_\alpha(X) = H_{\tilde{\phi}_\alpha}(X) \tag{2.4}$$

given by (2.2) when $\phi = \phi_\alpha$ which maximize the general *α-informations*

$$I_\alpha(X;Y) = D_\alpha\left(p_{X,Y} \| p_X p_Y\right), \quad \alpha \in \mathbb{R} \tag{2.5}$$

(cf. (1.15) and (1.5)). The trivial case when the Hartley entropy $H_0(X)$ is zero is excluded.

Assertion 2.2. The entropies $\tilde{H}_\alpha(X)$ are infinite for $\alpha \le 0$ and finite, given by

$$\tilde{H}_\alpha(X) = \frac{1}{\alpha} H_{2-\alpha}(X) \tag{2.6}$$

for $\alpha > 0$. This confirms the well known fact that the maximal Shannon information $\tilde{H}_1(X)$ is the Shannon entropy $H_1(X)$. However, this implies also that the maximal Pearson information $\tilde{H}_2(X)$ is half of the Hartley entropy $H_0(X)$ given in (1.21), the maximal quarter-Pearson information (Hellinger information) $\tilde{H}_{1/2}(X)$ is the entropy

$$H_{3/2}(X) = 4 \sum_x p(x) \left(1 - \sqrt{p(x)}\right) \tag{2.7}$$

and the maximal double-Pearson information $\tilde{H}_4(X)$ is the non-concave entropy

$$\frac{1}{2} H_{-2}(X) = \frac{1}{12} \left(\sum_x \frac{1}{p(x)^2} - 1 \right). \tag{2.8}$$

Proof. As mentioned above, the assumption $p(x)\, p(y) - p(x,y) > 0$ implies $p(x) > 0$ for all $x \in \mathcal{X}$. Therefore the sum in (2.2) is finite and $H_2(X)$ is by assumptions positive. Therefore $H_{\tilde{\phi}}(X) = \infty$ if and only if

$$\phi(0) = \phi_\alpha(0) = \infty.$$

From (1.5), (1.6) we see that this takes place for $\alpha \le 0$ while $\phi_\alpha(0) = 1/\alpha$ for $\alpha > 0$. Now assuming $\alpha > 0$ and substituting $\phi(0) = 1/\alpha$ and

$$\phi^*(t) = \phi_{1-\alpha}(t) = \frac{1 - t^{1-\alpha}}{\alpha(1-\alpha)} - \frac{1}{\alpha}(1-t) \quad \text{for } \alpha \ne 1$$

and

$$\phi^*(t) = \phi_0(t) = -\ln t + t - 1 \quad \text{for } \alpha = 1$$

in (2.2) we find the desired form (2.6) for $\tilde{H}_\alpha(X)$. The concrete expressions $\tilde{H}_2(X) = H_0(X)/2$ as well as the expressions (2.7) and (2.8) follow from (2.6) and from the definition of α-entropies in (1.17), (1.18). □

Let us consider for $p = (\pi,\, 1-\pi)$ the nonconcave entropy

$$\tilde{H}_{-2}(p) = h(\pi) = \frac{1}{12} \left(\frac{1}{\pi^2} + \frac{1}{(1-\pi)^2} - 1 \right) \tag{2.9}$$

given by (2.8). Since $\varphi(\pi) = 1/\pi^2$ is convex on $(0,1)$, we get

$$h(\pi) > \frac{1}{12} \left(\frac{2}{(1/2)^2} - 1 \right) = h(1/2) \tag{2.10}$$

for $\pi \neq 1/2$. This as well as the discontinuity $h(\pi) \to \infty$ for $\pi \to 0$ is contrary to what we observe in the case of concave entropies like $H_2(p) = 1 - \pi^2 - (1-\pi)^2$. But nevertheless the information measure $h(\pi)$ of (2.9) is justified. By (2.1) and Assertion 1.2, $h(\pi)$ is the double-Pearson information $I_4(X;X)$, i. e. it is the double-Pearson divergence of the 2×2 contingency tables for $p_{X,X}$ and $p_X\, p_X$ given below.

π	0
0	$1-\pi$

π^2	$\pi(1-\pi)$
$\pi(1-\pi)$	$(1-\pi)^2$

We see that the absolute deviations $|(p(x,y)/p(x)\,p(y)) - 1|$ for $\pi \neq 1/2$ and $\pi = 1/2$ are

$$\frac{1-\pi}{\pi}, \quad 1, \quad 1, \quad \frac{\pi}{1-\pi} \qquad \text{or} \qquad 1,\ 1,\ 1,\ 1$$

respectively, so that the sum of positive powers of the left-hand deviations may be arbitrarily larger than the similar sum on the right-hand side. This helps to understand that if the information in X is measured on the double-Pearson scale by $h(\pi)$ then $h(\pi)$ with π close to zero may be considerably larger than $h(1/2)$. Explicitly one can calculate the double-Pearson divergence of the contingency tables for $\pi = 1/2$ which is smaller than that for $\pi = 1/4$ while the half-Pearson divergence for $\pi = 1/2$ is larger than that for $\pi = 1/4$ (and the ordinary Pearson divergence is in both cases the same). Therefore $h(\pi)$ of (2.9) satisfies (2.10) while the Shannon information

$$h(\pi) = -\pi \ln \pi - (1-\pi)\ln(1-\pi) \tag{2.11}$$

is for $\pi = 1/2$ larger than that for $\pi = 1/4$ and the Pearson information $h(\pi) = 1/2$ is constant for all $\pi \in (0,1)$.

Thus we can summarize that the form of the entropy measuring the information in a message X from a given source $(\mathcal{X},\, p(x))$ depends on the ϕ-divergence used to quantify the information in Y about X. Some ϕ-divergences legitimize in this manner nonconcave entropies.

3 Entropies and Bayesian decisions

In statistical decisions it is desirable to estimate the minimal achievable decision error (the *Bayes error*)

$$e(X) = 1 - \max_x p(x) \quad \text{for} \quad p(x) = p_X(x) \tag{3.1}$$

by an analytically more tractable function. Natural candidates for this role are the measures of information $H_\psi(X)$, in particular $H_1(X)$ and $H_2(X)$ as the most prominent of them. R. M. Fano was the first who found for $e = e(X)$ achievable upper bound in terms of the Shannon entropy $H_1 = H_1(X)$ of (1.16) and Kovalevskij [6] was probably the first who found the corresponding lower bound. These bounds are

$$h(k(1-e)) + k(1-e)\ln k \le H_1 \le h(e) + e\ln(n-1) \tag{3.2}$$

for $h(\pi)$ given by (2.11) and n denoting the number of messages in $\mathcal{X}$. The upper bound holds in the whole range $0 \le e \le (n-1)/n$. The lower bound holds in the subrange

$$\frac{k-1}{k} \le e \le \frac{k}{k+1}, \quad k = 1, \ldots, n-1. \tag{3.3}$$

Assertion 3.1. Achievable bounds for quadratic entropy $H_2 = H_2(X)$ of (1.20) are

$$k(1-e)\,(1+e-(1-e)\,k) \le H_2 \le e\left(2 - \frac{n\,e}{n-1}\right) \tag{3.4}$$

where the lower bound holds for e satisfying (3.3).

Proof. Vajda and Vašek [12] proved that if $H(p)$ is a concave function of probability distributions $p = (p_1, \ldots, p_n)$ then among all p with $e = 1 - \max p_i \in ((k-1)/k,\ k/(k+1)]$ the function $H(p)$ is maximized at $p^+ = (1-e,\ e/(n-1), \ldots, e/(n-1))$ and minimized at $p^- = (1-e, \ldots, 1-e,\ 1-k(1-e), 0, \ldots, 0)$. It is easy to see that $H_2(p^+)$ and $H_2(p^-)$ are the bounds given in (3.4). □

The conjecture of Vajda [10] was that the quadratic entropy provides tighter bounds for the Bayes error than the Shannon entropy. This conjecture was so far neither rejected nor confirmed. It can be rigorously studied using the differences $e_\alpha^{\max}(H) - e_\alpha^{\min}(H)$ between maximal and minimal Bayes errors under the α-entropy $H_\alpha = H$ in the domain $0 \le H \le H_\alpha^{\max}$ and the *average inaccuracies*

$$A_\alpha = \frac{1}{H_\alpha^{\max}} \int_0^{H_\alpha^{\max}} \left[e_\alpha^{\max}(H) - e_\alpha^{\min}(H)\right] \mathrm{d}H \tag{3.5}$$

of the best possible estimates of Bayes errors $e(X)$ on the basis of entropies $H_\alpha(X)$. Applying the formula (3.5) to the Shannon entropy H_1 with $H_1^{\max} = \ln n$ and to the quadratic entropy H_2 with $H_2^{\max} = (n-1)/n$ we get the following result.

Assertion 3.2. For every $n \ge 2$ it holds

$$A_1 = \frac{n-1}{2n} - \frac{1}{2\ln n} \sum_{k=1}^{n-1} \frac{\ln(k+1)}{k(k+1)} \tag{3.6}$$

and

$$A_2 = \frac{n}{3(n-1)} \sum_{k=1}^{n-1} \frac{1}{k(k+1)^2} - \frac{1}{3n}\,. \tag{3.7}$$

Proof. This result follows by a standard integration using the formula

$$A_\alpha = \frac{1}{H_\alpha^{\max}} \int_0^{(n-1)/n} \left[H_\alpha^+(e) - H_\alpha^-(e)\right] \mathrm{d}e \tag{3.8}$$

where H_α^+ and H_α^- are the α-entropy extremes under the Bayes errors $0 \le e \le (n-1)/n$ given for $\alpha = 1$ in (3.2), (3.3) and for $\alpha = 2$ in (3.3), (3.4). □

Next follows a table of values obtained from (3.6) and (3.7).

n	2	3	4	5	6	7	8	9	10	11	12
A_1	0	.092	.0142	.175	.198	.215	.229	.240	.249	.257	.264
A_2	0	.042	.062	.074	.081	.087	.091	.094	.096	.098	.100

From this table we see that the values A_1 are more than 100% above the values of A_2. This rigorously confirms the mentioned conjecture of Vajda [10].

References

[1] I. Csiszár. Informationstheoretische Ungleichung und ihre Anwendung auf den Beweis der Ergodizität von Markoffschen Ketten. *Publ. Math. Inst. Hungarian Acad. Sci.*, Ser. A, 8:85–108, 1963.

[2] I. Csiszár. Information-type measures of difference of probability distributions and indirect observations. *Studia Sci. Math. Hungar*, 2:299–318, 1967.

[3] I. Csiszár. A class of measures of informativity of observation channels. *Periodica Math. Hungar.*, 2:191–213, 1972.

[4] J. Havrda and F. Charvát. Concept of structural a-entropy. *Kybernetika*, 3:30–35, 1967.

[5] W. Höffding. Stochastische Abhängigkeit und Funktionaler Zusammenhang. *Skand. Aktuar. Tidskr.*, 25:200-207, 1942.

[6] V. A. Kovalevskij. The problem of character recognition from the point of view of mathematical statistics. *Character Readers and Pattern Recognition*, Spartan Books, New York, 3–30, 1967.

[7] F. Liese and I. Vajda. *Convex Statistical Distances.* Teubner, Leipzig 1987.

[8] D. Morales, L. Pardo and I. Vajda. Uncertainty of discrete stochastic systems. *IEEE Trans. Systems, Man and Cybernetics*, Part A, 26:681–697, 1996.

[9] K. Pearson. On the theory of contingency and its relation to association and normal correlation. *Drapers Company Research Memoirs*, Biometric Ser. 1, London 1904.

[10] I. Vajda. Bounds on the minimal error probability and checking a finite or countable number of hypotheses. *Information Transmission Problems*, 4:9–17, 1968.

[11] I. Vajda. *Theory of Statistical Inference and Information.* Kluwer, Boston 1989.

[12] I. Vajda and K. Vašek. Majorization concave entropies and comparison of experiments. *Probl. of Control and Inform. Theory*, 14:105–115, 1985.

[13] J. Zvárová. On measures of statistical dependence. Čas. pěst. matemat., 99:15–29, 1974.

Igor Vajda: Institute of Information Theory and Automation , Pod vodárenskou věží 4, Praha, 182 08, Czech Republic, *vajda@utia.cas.cz*

Jana Zvárová: Euro MISE Centre of Charles University and Academy of Sciences of the Czech Republic, Faculty of Mathematics and Physics, Pod vodárenskou věží 2, Prague, 182 07, Czech Republic, *zvarova@cs.cas.cz*