KYBERNETIKA — VOLUME *48* (2012), NUMBER 4, PAGES 714–749

GENERALIZED INFORMATION CRITERIA FOR BAYES DECISIONS

Domingo Morales and Igor Vajda

This paper deals with Bayesian models given by statistical experiments and standard loss functions. Bayes probability of error and Bayes risk are estimated by means of classical and generalized information criteria applicable to the experiment. The accuracy of the estimation is studied. Among the information criteria studied in the paper is the class of posterior power entropies which include the Shannon entropy as special case for the power $\alpha = 1$. It is shown that the most accurate estimate is in this class achieved by the quadratic posterior entropy of the power $\alpha = 2$. The paper introduces and studies also a new class of alternative power entropies which in general estimate the Bayes errors and risk more tightly than the classical power entropies. Concrete examples, tables and figures illustrate the obtained results.

Keywords: Shannon entropy, alternative Shannon entropy, power entropies, alternative power entropies, Bayes error, Bayes risk, sub-Bayes risk

Classification: 62C10, 62B10

1. INTRODUCTION

In Morales, Pardo and Vajda [13], we systematically studied the subclass of concave and, more generally, Schur concave functions $H(\pi)$ of probability distributions π on finite or countable state spaces Θ which can serve as general measures of *uncertainty of states* of stochastic systems (Θ, π). The corresponding functions $H(\pi_x)$ and their mean values

$$H(\mathcal{E}) = \int_{\mathcal{X}} H(\pi_x)\,\mathrm{d}P(x) \tag{1}$$

characterized *conditional and average conditional uncertainties of states* in the *stochastic observation experiment*

$$\mathcal{E} = \langle(\Theta, \pi),\ \{\pi_x : x \in \mathcal{X}\},\ (\mathcal{X}, \mathcal{S}, P)\rangle \tag{2}$$

where $(\mathcal{X}, \mathcal{S}, P)$ is marginal probability space of an observation X and π_x are conditional probability distributions on Θ corresponding to the values $X = x \in \mathcal{X}$. Particular attention was paid to the separable measures of uncertainty of the form

$$H(\pi) = \sum_{\theta\in\Theta} \phi(\pi(\theta)) \tag{3}$$

for concave functions $\phi(t)$, $0 \le t \le 1$ and the related conditional and average conditional uncertainties $H_\phi(\pi_x)$ and $H_\phi(\mathcal{E})$.

The special separable measure of uncertainty

$$H_1(\pi) = \sum_{\theta\in\Theta} \phi_1(\pi(\theta)) = -\sum_{\theta\in\Theta} \pi(\theta)\ln\pi(\theta) \tag{4}$$

obtained for $\phi_1(t) = -t\log t$ is the classical Shannon entropy used by Shannon to characterize the amount of information in the state from the stochastic state space (Θ, π). The observation X in the experiment $\mathcal{E}$ reduces this information into the residual amount given by the conditional entropy

$$H_1(\mathcal{E}) = \int_{\mathcal{X}} H_1(\pi_x)\,\mathrm{d}P(x) = -\sum_{\theta\in\Theta}\int_{\mathcal{X}} \pi_x(\theta)\ln\pi_x(\theta)\,\mathrm{d}P(x).$$

The difference

$$I_1(\mathcal{E}) = H_1(\pi) - H_1(\mathcal{E})$$

is the information in observation X about the state in the experiment $\mathcal{E}$.

In the light of these basic concepts of the Shannon information theory, we interpret the above considered functions $H(\pi)$ as *generalized informations* in stochastic states, $H(\mathcal{E})$ as *residual informations* in these states and the differences

$$I(\mathcal{E}) = H(\pi) - H(\mathcal{E}) \tag{5}$$

as *informations in the observations* about the states.

By applying the concave function $\phi_2(t) = t(1-t)$ in (3) one obtains an alternative

$$H_2(\pi) = \sum_{\theta\in\Theta} \phi_2(\pi(\theta)) = 1 - \sum_{\theta\in\Theta} \pi^2(\theta) \tag{6}$$

to the Shannon entropy (4) called *quadratic entropy* by Vajda [18]. It is a quadratic measure of information obtained by identifying the state from the source (Θ, π). Relations (1) and (5) define the corresponding quadratic residual information

$$H_2(\mathcal{E}) = \int_{\mathcal{X}} H_2(\pi_x)\,\mathrm{d}P(x) = 1 - \sum_{\theta\in\Theta}\int_{\mathcal{X}} \pi_x^2(\theta)\mathrm{d}P(x)$$

in this state left by the observation from experiment $\mathcal{E}$ and the quadratic information $I_2(\mathcal{E}) = H_2(\pi) - H_2(\mathcal{E})$ contained in the observation about this state. In fact, Cover and Hart [4] and Vajda [18] used independently and in different sense the quadratic measures of informativity $H_2(\pi)$ and $H_2(\mathcal{E})$ as parameters of quality of decisions concerning the states $\theta \in \Theta$ based on observations X from statistical experiments $\mathcal{E}$.

The second of the mentioned papers was interested in the probabilities of error $P_e(\mathcal{E})$ of Bayes decisions $\delta_B : \mathcal{X} \longmapsto \Theta$ in the experiments $\mathcal{E}$ and established the quadratic information bounds

$$\frac{H_2(\mathcal{E})}{1+\sqrt{1-nH_2(\mathcal{E})/(n-1)}} \le P_e(\mathcal{E}) \le H_2(\mathcal{E}) \tag{7}$$

which can be rewritten as

$$P_e(\mathcal{E}) \le H_2(\mathcal{E}) \le 1 - (1 - P_e(\mathcal{E}))^2 - \frac{P_e(\mathcal{E})^2}{n-1}. \tag{8}$$

The paper rigorously proved these bounds including their attainability (tightness) for $n = 2$ and $n = \infty$ (when the term $nH_2(\mathcal{E})/(n-1)$ in (7) is replaced by the limit $H_2(\mathcal{E})$ achieved for $n \to \infty$) and the proof was extended to $3 \le n < \infty$ by Salikhov [15]. These bounds allow to use the computationally simpler residual information $H_2(\mathcal{E})$ rather than $P_e(\mathcal{E})$ for characterization of the quality of achievable decisions, e. g. when investigating the pattern recognition based of various selected features or the learning of machines trained by various teachers or various sizes of empirical data.

The quadratic entropy (6) requires only $n = |\Theta|$ operations of multiplication and summation. It is thus computationally simpler than the Shannon entropy (4) or any of the entropies of Rényi [14]

$$H^{(\alpha)}(\pi) = \frac{1}{\alpha - 1} \ln \sum_{\theta \in \Theta} \pi^\alpha(\theta), \quad \alpha > 0,\ \alpha \ne 1 \tag{9}$$

containing the Shannon entropy as the special limit case $H^{(1)}(\pi) = \lim_{\alpha \to 1} H^{(\alpha)}(\pi) = H_1(\pi)$. Rényi introduced the entropies (9) axiomatically by extending the additivity rule in the axioms suggested earlier by Faddeev [7] to characterize the Shannon's $H_1(\pi)$. However, he emphasized also the alternative to the axiomatic introduction of entropies and measures of information, called *pragmatic approach* by him. It proposes as measures of information arbitrary functionals of stochastic decision models which characterize the optimality of decisions achieved in these models. In this sense for example, Kovalevsky [12] pragmatically supported the Shannon residual information $H_1(\mathcal{E})$ by establishing attainable bounds for $P_e(\mathcal{E})$ in terms of $H_1(\mathcal{E})$ similar to (7), (8) which will be given below and which in some sense inspired the research leading to (7), (8). The Kovalevsky bounds were reinvented and applied in different areas of machine learning or more general information processing by a number of authors, e. g. Tebbe and Dwyer [16] or Feder and Merhav [8]. The quadratic entropy bounds (7), (8) are not only computationally simpler than the Kovalevsky bounds, but also tighter as it is proved in the present paper.

By appropriately modifying the extended additivity rule of Rényi [14], Havrda and Charvát [10] axiomatically introduced the one-one modification

$$H_\alpha(\pi) = \frac{1}{\alpha - 1} \left(1 - \sum_{\theta \in \Theta} \pi^\alpha(\theta) \right), \quad \alpha > 0, \alpha \ne 1 \tag{10}$$

of the Rényi entropies with the limit $H_1(\pi) = \lim_{\alpha \to 1} H_\alpha(\pi)$. Vajda [19] proposed the generalized measures of information $H_\alpha(\pi)$ and $H_\alpha(\mathcal{E})$ obtained by employing the general power informations $H_\alpha(\pi)$, $\alpha > 0$ in (1) as feature extraction criteria for the systems of automatic pattern recognition. He also formulated the problem of investigating for which powers $\alpha > 0$ one can find the bounds of the type (7) – (8) and for which of them will be the most tight. These criteria were cited later by many authors, e. g. Kanal [11], Devijver and Kittler [5] or Devroye et al. [6], and the bounds of the type (7), (8) were later found, tightened or applied by Toussaint [17], Ben Bassat [1], Ben Bassat and Raviv [2] and Harremoës and Topsøe [9].

Vajda and Vašek [20] found a method for obtaining attainable bounds of the type (7), (8) for arbitrary Schur concave entropies (1) and applied them to derive in a simple and rigorous manner the attainable upper and lower bounds of the prior Bayes probability of error $P_e(\pi)$ for given prior power information $H_\alpha(\pi), \alpha > 0$ and similar bounds for the average posterior Bayes probability error $P_e(\mathcal{E})$ for given residual power information $H_\alpha(\mathcal{E}), \alpha > 0$. Their results were applied later in Morales, Pardo and Vajda [13] and Vajda and Zvárová [21]. It is to be noted that the error probability $P_e(\mathcal{E})$ is related to the residual information $H_\alpha(\mathcal{E})$ rather than to the information $H_\alpha(\pi) - H_\alpha(\mathcal{E})$ gained by observing the experiment outcome because it is isotone with (i. e. proportionally related to) the former and antitone with the latter.

In this paper we introduce a new class of *adjoint power informations* $\tilde{H}_\alpha(\pi)$ and $\tilde{H}_\alpha(\mathcal{E})$ for the powers $\alpha > 0$ where the adjoint Shannon informations $\tilde{H}_1(\pi)$ and $\tilde{H}_1(\mathcal{E})$ differ from the classical Shannon informations $H_1(\pi)$ and $H_1(\mathcal{E})$. We use the results of Vajda and Vašek [20] to obtain for these measures of information the attainable bounds of the Bayes error $P_e(\mathcal{E})$ and also the bounds of the prior Bayes loss L_B and Bayes risk $R_B(\mathcal{E})$ for a class of common loss functions. The main issue addressed in the paper is the accuracy of specification of the Bayes error $P_e(\mathcal{E})$ by the residual informations $H_\alpha(\mathcal{E})$ and $\tilde{H}_\alpha(\mathcal{E})$ of all powers $\alpha > 0$. Perhaps the most interesting of the obtained results is the fact that the quadratic residual information $H_2(\mathcal{E})$ specifies the Bayes error most accurately in the class of all power residual informations $H_\alpha(\mathcal{E})$, $\alpha > 0$. However, we also show that the accuracies achieved by the alternative power residual informations $\tilde{H}_\alpha(\mathcal{E})$, $\alpha > 0$ uniformly dominate those achieved by $H_\alpha(\mathcal{E})$, $\alpha > 0$.

Basic concepts and results are in Sections 2 – 4. The main results are in Section 5 and 6.

2. GENERAL LOSS MODEL

Consider the classical model of Bayesian decision theory (cf. e. g. Berger [3]) with state of nature θ from a finite set Θ, prior probability distributions of states $\pi = (\pi(\theta) > 0 : \theta \in \Theta)$ and observations (random samples) X conditionally distributed by probability measures P_θ on a measurable observation space $(\mathcal{X}, \mathcal{S})$ depending on the states $\theta \in \Theta$. We restrict ourselves to the important situation where the purpose of decision is identification of the unknown state θ. Thus our decisions (actions in the sense of Berger) are states θ from the action space Θ, and the loss functions are of the form

$$L : \Theta \times \Theta \mapsto [0, \infty) \quad \text{where} \quad \max_{\theta \in \Theta} L(\theta, \theta) = 0, \quad \min_{\hat{\theta} \in \Theta} \max_{\theta \in \Theta} L(\theta, \hat{\theta}) > 0. \tag{11}$$

Thus we deal with the Bayesian model given by a statistical experiment

$$\mathcal{E} = \langle \pi, \mathcal{P} = \{P_\theta : \theta \in \Theta\} \rangle \tag{12}$$

and a nontrivial loss function (11).

This is the standard decision-theoretic model of many real situations, in particular of the

(1) *pattern recognition* where the states of nature θ represent various possible patterns (images) and $L(\theta,\hat{\theta}) > 0$ is the loss incurred by the wrong identifications $\hat{\theta}$ of these patterns,

(2) *classification* where the states θ represent various classes of objects and $L(\theta,\hat{\theta}) > 0$ is the loss of misclassification, and

(3) *information transmission* where the states θ represent various possible messages transmitted via communication channel $(\Theta, \{P_\theta : \theta \in \Theta\}, \mathcal{X})$ with input alphabet Θ, output alphabet $\mathcal{X}$ and transition probability distributions P_θ describing distortion of messages by the channel noise.

These concrete interpretations and their various combinations appear also in the *detection theory* and *stochastic control theory.*

Let us briefly review basic concepts of Bayesian decision theory applicable in the present model. The *expected loss* of an individual identification action $\hat{\theta} \in \Theta$ is

$$\mathcal{L}(\pi,\hat{\theta}) = \sum_{\theta\in\Theta} L(\theta,\hat{\theta})\,\pi(\theta). \tag{13}$$

Each individual action $\theta_\pi \in \Theta$ with the property

$$\theta_\pi = \operatorname{argmin}_{\hat{\theta}} \mathcal{L}(\pi,\hat{\theta}) \tag{14}$$

is said to be *Bayes action* (Bayes decision without data) and the minimal a priori expected loss

$$L_B(\pi) = \mathcal{L}(\pi,\theta_\pi) \tag{15}$$

is a *prior Bayes loss.* Observation data $x \in \mathcal{X}$ are assumed to be used for identification by means of *identification rules*

$$\delta = \mathcal{X} \mapsto \Theta. \tag{16}$$

Technically, they are assumed to be $\mathcal{S}$-measurable and P_θ-integrable for all $\theta \in \Theta$. The *risk function* of the identification rule (16) is

$$R(\theta,\delta) = \int_{\mathcal{X}} L(\theta,\delta(x))\,\mathrm{d}P_\theta(x), \quad \theta\in\Theta$$

and its expected value

$$\mathcal{R}(\pi,\delta) = \sum_{\theta\in\Theta} R(\theta,\delta)\pi(\theta) = \sum_{\theta\in\Theta}\int_{\mathcal{X}} L(\theta,\delta(x))\pi(\theta)\,\mathrm{d}P_\theta(x) \tag{17}$$

is simply denoted as *risk.* The minimizer

$$\delta_B = \operatorname{argmin}_\delta \mathcal{R}(\pi,\delta) \tag{18}$$

is the *Bayes identification rule* and

$$R_B = R_B(\mathcal{E},L) = \mathcal{R}(\pi,\delta_B) \tag{19}$$

the *Bayes risk* of identification in the model under consideration specified by the experiment $\mathcal{E}$ and loss function L.

Under the above specified assumptions, the Bayes identification rule exists and is given by a relatively simple explicit formula. To demonstrate this and to find the Bayes identification rule formula, take first into account the marginal probability distribution

$$P = \sum_{\theta \in \Theta} \pi(\theta) P_\theta \tag{20}$$

on the observation space $(\mathcal{X}, \mathcal{S})$ which dominates each conditional distribution P_θ in the sense $P(S) = 0$ implies $P_\theta(S) = 0$ for $S \in \mathcal{S}$. Hence there exists the Radon–Nikodym density

$$p_\theta(x) = \frac{\mathrm{d}P_\theta(x)}{\mathrm{d}P(x)}$$

defined for all data $x \in \mathcal{X}$, with values uniquely given except possibly for a set $S_\theta \in \mathcal{S}$ with $P(S_\theta) = 0$ (i. e. for P-almost all in symbols P-a.e. on $\mathcal{X}$). Then

$$\pi_x = (\pi_x(\theta) \stackrel{\triangle}{=} \pi(\theta) p_\theta(x) : \theta \in \Theta) \tag{21}$$

is the conditional (posterior) probability distribution on Θ given data x. Indeed, by the definition of Radon–Nikodym densities, $p_\theta(x)$

$$\min_\theta \pi_x(\theta) \geq 0 \quad \text{and} \quad \sum_\theta \pi_x(\theta) = \frac{\mathrm{d}P(x)}{\mathrm{d}P(x)} = 1 \qquad P\text{-a.e. on } \mathcal{X}.$$

Obviously, the statistical experiment (12) is equivalently described by the conditional distributions (21) for $x \in \mathcal{X}$ and the marginal distribution (20),

$$\mathcal{E} = \langle \pi, \mathcal{P} = \{P_\theta : \theta \in \Theta\} \rangle \equiv \langle P, \Pi = \{\pi_x : x \in \mathcal{X}\} \rangle. \tag{22}$$

Using the posterior distribution (21) and the concept of expected loss (13), we can rewrite the risk formula (17) into the simple form

$$\mathcal{R}(\pi_x, \delta) = \int_{\mathcal{X}} \mathcal{L}(\pi_x, \delta(x))\, \mathrm{d}P(x). \tag{23}$$

From here and from (18) we see that an identification rule δ is Bayes (in symbols $\delta = \delta_B$) if and only if for P-almost all data $x \in \mathcal{X}$ the data based action $\delta_B(x)$ is Bayes for the posterior distribution, π_x, i. e. coincides with some θ_{π_x} defined in accordance with (14). Thus the Bayes identification rule can equivalently be defined P-a.e. on $\mathcal{X}$ by the formula

$$\delta_B(x) = \theta_{\pi_x} \equiv \mathrm{argmin}_{\hat{\theta}} \sum_{\theta \in \Theta} L(\theta, \hat{\theta}) \pi_x(\theta). \tag{24}$$

From here we deduce also that the Bayes risk R_B is the expected *posterior Bayes loss* given data x, denoted $L_B(\pi_x)$ and defined by (15) with the prior distribution π replaced by the posterior distribution π_x. In other words, we deduce that

$$\begin{aligned} R_B = \mathcal{R}(\pi, \delta_B) \quad &= \int_{\mathcal{X}} \mathcal{L}(\pi_x, \theta_{\pi_x})\, \mathrm{d}P(x) \qquad \text{(cf. (23), (24))} \\ &= \int_{\mathcal{X}} L_B(\pi_x)\, \mathrm{d}P(x). \end{aligned} \tag{25}$$

3. RELATIONS TO ZERO-ONE LOSS MODEL

A prominent role in the applications of the model of previous section plays the error loss function

$$L_e : \Theta \times \Theta \mapsto \{0,1\}, \qquad L_e(\theta,\hat\theta) = \begin{cases} 1 & \text{if } \hat\theta \neq \theta, \\ 0 & \text{if } \hat\theta = \theta. \end{cases} \tag{26}$$

Here the general expected loss $\mathcal{L}(\pi,\hat\theta)$ reduces to the *prior probability of error* of the identification action $\hat\theta \in \Theta$,

$$\mathcal{L}_e(\pi,\hat\theta) = \sum_{\theta\in\Theta} L_e(\theta,\hat\theta)\pi(\theta) = 1 - \pi(\hat\theta) \tag{27}$$

The Bayes identification action θ_π thus minimizes this probability of error over $\hat\theta \in \Theta$. This means that the prior Bayes expected loss $L_B(\pi)$ given by (15) is the minimal prior probability of error given by the formula

$$e_B(\pi) = 1 - \pi(\theta_\pi), \tag{28}$$

and called simply *prior Bayes error*. Similarly the posterior Bayes expected loss $L_B(\pi_x)$ for data $x \in \mathcal{X}$ is in this case the minimal posterior probability of error

$$e_B(\pi_x) = 1 - \pi_x(\theta_{\pi_x}) \tag{29}$$

called simply *posterior Bayes error*, as the Bayes identification action $\theta_{\pi_x} \in \Theta$ minimizes over $\hat\theta \in \Theta$ the posterior error probability $1-\pi(\hat\theta)$. Finally by (25) and the equality $L_B(\pi_x) = e_B(\pi_x)$, the Bayes risk $R_B = R_B(\mathcal{E}, L)$ of (19) achieved under the special loss function $L = L_e$ coincides with the *Bayes error* (average minimal posterior probability of error) depending only on the experiment $\mathcal{E}$ and given by the formula

$$e_B = e_B(\mathcal{E}) = \int_{\mathcal{X}} e_B(\pi_x)\,\mathrm{d}P(x). \tag{30}$$

As mentioned in the introduction, our intention is to evaluate or estimate performances of Bayes identification rules in the general loss function models by means of known performances of such rules in the simpler error loss function models. The rest of this section is devoted to the research of this eventuality. The achieved results serve in the next section to establish new bounds for the Bayes risk R_B based partly on the bounds for the Bayes error probability e_B established in previous literature and partly on new such bounds established in the next section.

In the general loss model (11) the proper losses are positive between

$$L^- = \min\{L(\theta,\hat\theta) : \theta,\hat\theta \in \Theta,\, L(\theta,\hat\theta) > 0\},$$

and

$$L^+ = \max\{L(\theta,\hat\theta) : \theta,\hat\theta \in \Theta\} \geq L^-$$

We characterize them by two parameters called *median loss* and *loss dispersion*

$$\Lambda = \frac{L^+ + L^-}{2} \quad \text{and} \quad \Delta = (L^+ - \Lambda). \tag{31}$$

Example 3.1. Let the state space $\Theta = \{1, \ldots, n\}$ represents classification of satellite ship images and let the loss function (11) be given as the matrix

$$\big(L(\theta, \hat{\theta})\big)_{\theta,\hat{\theta}=1}^{n} = \begin{pmatrix} 0 & 4/5 & 4/5 & \ldots & 4/5 & 1 \\ 4/5 & 0 & 4/5 & \ldots & 4/5 & 1 \\ 4/5 & 4/5 & 0 & \ldots & 4/5 & 1 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 4/5 & 4/5 & 4/5 & \ldots & 0 & 1 \\ 6/5 & 6/5 & 6/5 & \ldots & 6/5 & 0 \end{pmatrix}$$

where states $1 \le \theta \le n-1$ represent merchant ships of $n-1$ different nations and the state $\theta = n$ represents a pirate ship. Here

$$L^- = 4/5, \quad L^+ = 6/5 \quad \text{and} \quad (\Delta, \Lambda) = (1/5, 1)\,.$$

The following theorem helps to extend the information bounds obtained for the Bayes probability of errors to the Bayes losses and risks.

Theorem 3.1. If the general loss model of Section 2 has median loss Λ and the loss dispersion $\Delta \ge 0$, then

(i) the prior Bayes loss L_B and the prior Bayes error e_B satisfy the relation

$$|L_B(\pi) - e_B(\pi)\,\Lambda| \le e_B(\pi)\,\Delta/2,$$

(ii) for P-almost all $x \in \mathcal{X}$, the posterior Bayes loss $L_B(\pi_x)$ and the posterior Bayes error $e_B(\pi_x)$ satisfy the relation

$$|L_B(\pi_x) - e_B(\pi_x)\,\Lambda| \le e_B(\pi_x)\,\Delta/2, \tag{32}$$

(iii) the Bayes risk R_B and the Bayes error satisfy the relation

$$|R_B - e_B\,\Lambda| \le e_B\,\Delta/2.$$

Proof. (I) It follows from the minimax assumption in (11) that $e_B(\pi) = 0$ if and only if $L_B(\pi) = 0$. Thus for $e_B(\pi) = 0$ (i) holds and we can restrict ourselves to π with $e_B(\pi) > 0$. By (31), $L(\theta, \hat{\theta}) > 0$ implies $L(\theta, \hat{\theta}) \in [L^-, L^+]$ where either $L(\theta, \hat{\theta}) \in [\Lambda, L^+]$ in which case

$$L(\theta, \hat{\theta}) - \Lambda \le L^+ - \Lambda = \Delta/2$$

or $L(\theta, \hat{\theta}) \in [L^-, \Lambda]$ in which case

$$\Lambda - L(\theta, \hat{\theta}) \le \Lambda - L^- = \Delta/2.$$

Hence

$$|L(\theta, \hat{\theta}) - \Lambda| \le \Delta/2 \quad \text{for all} \quad \theta, \hat{\theta} \in \Theta \;\; \text{with} \;\; L(\theta, \hat{\theta}) > 0. \tag{33}$$

Further, by (13) and (15),

$$\mathcal{L}(\pi, \hat{\theta}) = \sum_{\theta \neq \hat{\theta}} L(\theta, \hat{\theta})\pi(\theta) \quad \text{and} \quad L_B(\pi) = \sum_{\theta \neq \theta_\pi} L(\theta, \theta_\pi)\pi(\theta). \tag{34}$$

Therefore multiplying the left side of (33) by $\pi(\theta)/e_B(\pi)$, summing over all $\theta \neq \theta_\pi$ and using the Jensen inequality, we get

$$\left| \frac{1}{e_B(\pi)} \sum_{\theta \neq \theta_\pi} L(\theta, \hat{\theta})\pi(\theta) - \Lambda \right| \leq \frac{\Delta}{2}.$$

It remains to apply (34) to complete the proof of (i).

(II) Since π_x given in Section 2 are probability distributions on Θ for P-almost all $x \in \mathcal{X}$, (ii) follows from (i).

(III) Integrating both sides of (32) over $\mathcal{X}$ with respect to the measure P and using once more the Jensen inequality, we get

$$\left| \int_{\mathcal{X}} L_B(\pi_x)\,\mathrm{d}P(x) - \Lambda \int_{\mathcal{X}} e_B(\pi_x)\,\mathrm{d}P(x) \right| \leq \frac{\Delta}{2} \int_{\mathcal{X}} e_B(\pi_x)\,\mathrm{d}P(x).$$

The desired result of (iii) follows from here and from the formulas (25) and (30). □

4. GENERALIZED INFORMATION CRITERIA

In this section and in the rest of the paper, $n = |\Theta|$ denotes the number of states in Θ. We study estimates of Bayes errors $e_B(\pi)$, $e_B(\pi_x)$ and $e_B = e_B(\mathcal{E})$ (or more generally, the Bayes losses $L_B(\pi)$, $L_B(\pi_x)$ and Bayes risks $R_B = R_B(\mathcal{E})$) by means of measures of information $H(\pi)$ contained in states of nature θ from stochastic sources (Θ, π), and by residual informations (average conditional informations)

$$H(\mathcal{E}) = \int_{\mathcal{X}} H(\pi_x)\,\mathrm{d}P(x)$$

in the states from the systems of conditional stochastic sources $\mathcal{E} = \{(\Theta, \pi_x) : x \in \mathcal{X}\}$ where x are observed data (realizations of random observation X) from the observation probability space $(\mathcal{X}, \mathcal{S}, P)$. For details about these concepts and notations see sections 2 and 3.

One of the most widely used information measures is the *Shannon entropy* (here measured in *nats* instead of *bits*)

$$H_1(\pi) = \sum_{\theta \in \Theta} \phi_1(\pi(\theta)), \quad \phi_1(t) = -t \ln t.$$

In Section 1 we mentioned their generalizations called *power informations*

$$H_\alpha(\pi) = \sum_{\theta \in \Theta} \phi_\alpha(\pi(\theta)), \qquad \alpha > 0 \tag{35}$$

where

$$\phi_\alpha(t) = \begin{cases} \frac{1}{\alpha-1}\left[t(1-t^{\alpha-1})\right] & \text{if} \quad \alpha \neq 1 \\ \lim_{\alpha\to 1}\phi_\alpha(t) = -t\ln t & \text{if} \quad \alpha = 1. \end{cases} \tag{36}$$

Hence

$$H_\alpha(\pi) = \begin{cases} \frac{1}{\alpha-1}\left[1-\sum_{\theta\in\Theta}\pi(\theta)^\alpha\right] & \text{if} \quad \alpha \neq 1 \\ \lim_{\alpha\to 1} H_\alpha(\pi) = -\sum_{\theta\in\Theta}\pi(\theta)\ln\pi(\theta) & \text{if} \quad \alpha = 1. \end{cases} \tag{37}$$

As argued in Morales, Pardo and Vajda [13], the desired information-theoretic properties of the power informations follow from the strict concavity of functions $\phi_\alpha(t)$ on $[0,1]$ and from their extremal values $\phi_\alpha(0) = \phi_\alpha(1) = 0$. A characteristic property is the *information preservation law*:

$$0 = H_\alpha(\pi_D) \leq H_\alpha(\pi T^{-1}) \leq H_\alpha(\pi) \leq H_\alpha(\pi_U) = (1-n^{1-\alpha})/(\alpha-1)$$

where $T : \Theta \mapsto \mathcal{T}$ is a mapping of states $\theta \in \Theta$ into strings of events $\tau = T(\theta) \in \mathcal{T}$ representing a processing of information in the states θ and leading to the new distribution

$$\pi T^{-1}(\tau) = \sum_{\theta : T(\theta)=\tau} \pi(\theta)$$

on the space of information processing outcomes $\mathcal{T}$. The remaining symbols π_D, π_U stand for the Dirac and uniform probability distributions on Θ. This law thus says that the information processing cannot increase the information which is maximal when the states are equiprobable and minimal equal zero when the state is a fixed constant, and that the information is preserved if and only if the results of processing of different states are different, i. e. the information processing preserves the identifiability of states.

The strict concavity and zero extremal values are preserved by passing to the *adjoint power functions* $\tilde{\phi}_\alpha(t) = \phi_\alpha(1-t)$ so that the same information-theoretic properties are shared by the corresponding *adjoint power informations*

$$\tilde{H}_\alpha(\pi) = \sum_{\theta\in\Theta} \tilde{\phi}_\alpha(\pi(\theta)), \quad \alpha > 0 \tag{38}$$

given explicitly by the formulas

$$\tilde{H}_\alpha(\pi) = \begin{cases} \frac{1}{\alpha-1}\left[n-1-\sum_{\theta\in\Theta}(1-\pi(\theta))^\alpha\right] & \text{if} \quad \alpha \neq 1 \\ \lim_{\alpha\to 1}\tilde{H}_\alpha(\pi) = -\sum_{\theta\in\Theta}(1-\pi(\theta))\ln(1-\pi(\theta)) & \text{if} \quad \alpha = 1 \end{cases} \tag{39}$$

where n denotes the number of states in Θ.

Since $H_\alpha(\pi)$ and $\tilde{H}_\alpha(\pi)$ measures the information in the state from Θ distributed by π, they must be closely related tot the minimal error probability $e_B(\pi)$ of identification of the state on the basis of distribution π. Further, the Bayes error $e_B = e_B(\mathcal{E})$ in the general experiment $\mathcal{E}$ (cf. (22)) is the average minimal error probability

$$e_B(\mathcal{E}) = \int_{\mathcal{X}} e_B(\pi_x)\,\mathrm{d}P(x) \quad \text{(cf. (30))}, \tag{40}$$

so that it must be similarly related to the residual informations $H_\alpha(\mathcal{E})$ and $\tilde{H}_\alpha(\mathcal{E})$ defined as the stochastic mixtures

$$H_\alpha(\mathcal{E}) = \int_{\mathcal{X}} H_\alpha(\pi_x)\,\mathrm{d}P(x) \quad \text{and} \quad \tilde{H}_\alpha(\mathcal{E}) = \int_{\mathcal{X}} \tilde{H}_\alpha(\pi_x)\,\mathrm{d}P(x). \tag{41}$$

In what follows we investigate this relation.

In the next theorem we evaluate for all $\alpha > 0$ and $n = |\Theta|$ the upper and lower power information bounds

$$\mathcal{H}_\alpha^+(e_B) = \max_{e_B(\mathcal{E})=e_B} H_\alpha(\mathcal{E}) \quad \text{and} \quad \mathcal{H}_\alpha^-(e_B) = \min_{e_B(\mathcal{E})=e_B} H_\alpha(\mathcal{E}) \tag{42}$$

by means of the auxiliary function

$$h(t) = -t\ln t - (1-t)\ln(1-t), \quad 0 \le t \le 1 \ \text{ where } \ 0\ln 0 = 0 \tag{43}$$

and the auxiliary constants

$$a_{\alpha,k} = \begin{cases} \frac{1-k^{1-\alpha}}{\alpha-1} & \text{if } \alpha \neq 1 \\ \lim_{\alpha\to 1} a_{\alpha,k} = \ln k & \text{if } \alpha = 1 \end{cases} \quad \text{and} \quad c_k = \frac{k-1}{k}, \quad 1 \le k \le n \tag{44}$$

as well as

$$b_{\alpha,k} = \frac{a_{\alpha,k+1} - a_{\alpha,k}}{c_{k+1} - c_k}, \quad 1 \le k \le n-1. \tag{45}$$

In (42) and in the rest of the paper we use the fact that the range of the Bayesian errors $e(\pi)$ and e_B is the interval

$$0 \le e(\pi),\ e_B \le c_n. \tag{46}$$

In the proof of the next theorem are used the formulas

$$H_\alpha^+(e) = \frac{1-(n-1)^{1-\alpha}e^\alpha - (1-e)^\alpha}{\alpha-1}, \quad 0 \le e \le c_n \tag{47}$$

$$H_\alpha^-(e) = \frac{1-[1-k(1-e)]^\alpha - k(1-e)^\alpha}{\alpha-1}, \quad c_k \le e \le c_{k+1}, \quad 1 \le k \le n-1 \tag{48}$$

and their limits

$$H_1^+(e) = h(e) + e\ln(n-1), \quad 0 \le e \le c_n \tag{49}$$

$$H_1^-(e) = h(k(1-e)) + k(1-e)\ln k, \quad c_k \le e \le c_{k+1}, \quad 1 \le k \le n-1 \tag{50}$$

for the attainable upper and lower power information bounds

$$H_\alpha^+(e) = \max_{e(\pi)=e} H_\alpha(\pi) \quad \text{and} \quad H_\alpha^-(e) = \min_{e(\pi)=e} H_\alpha(\pi) \tag{51}$$

(for details about these bounds see Theorem 2 in Morales et al. [13]).

Theorem 4.1. The power information bounds (42) are for every $0 \le e_B \le c_n$ explicitly given by

$$\mathcal{H}_\alpha^+(e_B) = \begin{cases} H_\alpha^+(e_B) = \frac{1}{\alpha-1}\left[1-(n-1)^{1-\alpha}e_B^\alpha - (1-e_B)^\alpha\right] & \text{if } \alpha \neq 1 \\ H_1^+(e_B) = h(e_B) + e_B \ln(n-1) & \text{if } \alpha = 1 \end{cases} \tag{52}$$

(cf. (47), (50)) and

$$\mathcal{H}_\alpha^-(e_B) = \begin{cases} a_{\alpha,k} + b_{\alpha,k}(e_B - c_k) & \text{if } c_k \le e_B \le c_{k+1},\ 1 \le k \le n-1,\ \alpha \neq 1 \\ a_{\alpha,n} e_B / c_n, & \text{if } \alpha = 1. \end{cases} \tag{53}$$

The bounds $\mathcal{H}_\alpha^+(e_B)$ and $\mathcal{H}_\alpha^-(e_B)$ coincide only at the endpoints $c_1 = 0$ and c_n of the domain of e_B where

$$\mathcal{H}_\alpha^+(0) = \mathcal{H}_\alpha^-(0) = 0 \quad \text{and } \mathcal{H}_\alpha^+(c_n) = \mathcal{H}_\alpha^-(c_n) = a_{\alpha,n} > 0. \tag{54}$$

P r o o f. Consider an arbitrary $\alpha > 0$, arbitrary constants $0 \le \tilde{c} < c \le c_n$ and arbitrary distributions $\pi, \tilde{\pi}$ such that $e(\pi) = c$ and $\tilde{e}(\tilde{\pi}) = \tilde{c}$. Then the linear function

$$tH_\alpha(\pi) + (1-t)H_\alpha(\tilde{\pi}) \text{ of variable } 0 \le t \le 1$$

must be bounded from above by the function $\mathcal{H}_\alpha^+(tc + (1-t)\tilde{c})$ and bounded from below by the function $\mathcal{H}_\alpha^-(tc + (1-t)\tilde{c})$. This implies that $\mathcal{H}_\alpha^+$ must be concave and $\mathcal{H}_\alpha^-$ convex on the interval $[\tilde{c}, c] \subseteq [0,1]$. At the same time it follows from (41), (42) and (51) that $\mathcal{H}_\alpha^+$ must be minimal but above H_α^+ and $\mathcal{H}_\alpha^-$ must be maximal but below H_α^-. Since H_α^+ is concave itself, this implies $\mathcal{H}_\alpha^+ = H_\alpha^+$ so that (52) follow from (47) and (49). On the other hand, H_α^- given by (48) and (50) is piecewise concave in the intervals between the cutpoints c_k, $1 \le k \le n-1$. The piecewise linear function $\Phi_\alpha(t)$ of variable $t \in [0, c_n]$ connecting the points $[c_k, H_\alpha^-(c_k)] \equiv [c_k, a_{\alpha,k}]$ for $1 \le k \le n$ is

$$\Phi_\alpha(t) = a_{\alpha,k} + b_{\alpha,k}(t - c_k) \quad \text{for} \quad c_k \le t \le c_{k+1}, \quad 1 \le k \le n-1. \tag{55}$$

This function is convex (concave) if the sequence

$$\frac{\Phi_\alpha(c_k)}{c_k} = \frac{a_{\alpha,k}}{c_k} = \begin{cases} \frac{k\left(1-k^{1-\alpha}\right)}{(\alpha-1)(k-1)} & \text{if } \alpha \neq 1 \\ \lim_{\alpha\to 1} a_{\alpha,k} = \frac{k}{k-1}\ln k & \text{if } \alpha = 1 \end{cases}$$

is increasing (decreasing) for $k = 2, 3, \ldots$ Obviously, it is constant equal 1 if $\alpha = 2$, increasing if $0 < \alpha < 2$ and decreasing if $\alpha > 2$. Therefore $\mathcal{H}_\alpha^-(e_B) = \Phi_\alpha(e_B)$ if $0 < \alpha \le 2$ and $\mathcal{H}_\alpha^-(e_B)$ is linear in the variable e_B, equal $[\Phi_\alpha(c_n) - \Phi_\alpha(0)]\, e_B/c_n \equiv a_{\alpha,n} e_B / c_n$, if $\alpha > 2$. This proves (53). The last assertion including relations (54) is clear from what has already been proved. □

Figures C.1 and C.2 of Appendix C present the curves $\mathcal{H}_\alpha^\pm(e_B)$ as functions of variable e_B for $\alpha = 1/2, 3/4, 1$ and $\alpha = 2, 3, 4$. We see from these figures that the lower bounds $\mathcal{H}_\alpha^-(e_B)$ for $\alpha \ge 2$ are linear in the variable e_B.

Remark 4.1. Relation (49) is the well known Fano bound of information theory and (47) is its extension obtained previously in Vajda [18] for $\alpha = 2$ and in Morales et al. [13] and other references mentioned there for remaining $\alpha > 0$.

Remark 4.2. It is easy to verify that all power information bounds (47) – (53) are continuous functions strictly increasing (on their definition domain $0 \le e,\ e_B \le c_n$) from the minimum 0 to the maximum $a_{\alpha,n}$. Therefore the inverse functions

$$e_\alpha^{\mp}(H) = \max_{H_\alpha^{\pm}(e) \le H} e \quad \text{and} \quad e_{B,\alpha}^{\mp}(H) = \max_{H_\alpha^{\pm}(e_B) \le H} e_B \tag{56}$$

(notice the reversed order of $\pm$ and $\mp$ here!) are for all $\alpha > 0$ continuously increasing (on their definition domain $0 \le H \le a_{\alpha,n}$) from the common minimum 0 to the common maximum c_n at the endpoints of the domain, and with different values

$$e_\alpha^-(H) < e_\alpha^+(H) \quad \text{and} \quad e_{B,\alpha}^-(H) < e_{B,\alpha}^+(H) \tag{57}$$

between the endpoints. The values $e_\alpha^{\pm}(H_\alpha(\pi))$, $e_\alpha^{\pm}(H_\alpha(\pi_x))$ and $e_{B,\alpha}^{\pm}(H_\alpha(\mathcal{E}))$ are attainable upper and lower estimates of the prior, posterior and average Bayes errors $e(\pi)$, $e(\pi_x)$ and $e_B = e_B(\mathcal{E})$ based on the prior, posterior and overall power information measures $H_\alpha(\pi), H_\alpha(\pi_x)$ and $H_\alpha(\mathcal{E})$.

The next theorem evaluates the upper and lower adjoint power information bounds

$$\tilde{\mathcal{H}}_\alpha^+(e_B) = \max_{e_B(\mathcal{E}) = e_B} \tilde{H}_\alpha(\mathcal{E}) \quad \text{and} \quad \tilde{\mathcal{H}}_\alpha^-(e_B) = \min_{e_B(\mathcal{E}) = e_B} \tilde{H}_\alpha(\mathcal{E}). \tag{58}$$

It uses the same c_k as Theorem 4.1 and for every $\alpha > 0$ also the constants

$$\tilde{a}_{\alpha,k} = \begin{cases} \frac{k-1}{\alpha-1}\left[1 - \left(\frac{k-1}{k}\right)^{\alpha-1}\right] & \text{if } \alpha \ne 1 \\ \lim_{\alpha\to 1} \tilde{a}_{\alpha,k} = (1-k)\ln\frac{k-1}{k} & \text{if } \alpha = 1 \end{cases} \quad \text{for } 1 \le k \le n \tag{59}$$

and

$$\tilde{b}_{\alpha,k} = \frac{\tilde{a}_{\alpha,k+1} - \tilde{a}_{\alpha,k}}{c_{k+1} - c_k}, \quad 1 \le k \le n-1 \tag{60}$$

where $0 \ln 0 = 0$ in (59).

Theorem 4.2. Let $\alpha > 0$ be arbitrary fixed. The adjoint power information bounds (58) are for every $0 \le e_B \le c_n$ explicitly given by

$$\tilde{\mathcal{H}}_\alpha^+(e_B) = \begin{cases} \frac{1}{\alpha-1}\left[n - 1 - e_B^\alpha - (n-1)\left(1 - \frac{e_B}{n-1}\right)^\alpha\right] & \text{if } \alpha \ne 1 \\ \lim_{\alpha\to 1} \tilde{\mathcal{H}}_\alpha^+(e_B) = -e_B \ln e_B - (n-1-e_B)\ln\left(\frac{n-1-e_B}{n-1}\right) & \text{if } \alpha = 1 \end{cases} \tag{61}$$

and

$$\tilde{\mathcal{H}}_\alpha^-(e_B) = \begin{cases} \tilde{a}_{\alpha,k} + \tilde{b}_{\alpha,k}(e_B - c_k) & \text{if } c_k < e_B < c_{k+1}, 1 \le k \le n-1, \alpha > 2 \\ \tilde{a}_{\alpha,n} e_B / c_n & \text{if } 0 < \alpha \le 2. \end{cases} \tag{62}$$

The bounds $\tilde{\mathcal{H}}_\alpha^+(e_B)$ and $\tilde{\mathcal{H}}_\alpha^-(e_B)$ coincide only at the endpoints $c_1 = 0$ and c_n of the domain of e_B where

$$\tilde{\mathcal{H}}_\alpha^+(0) = \tilde{\mathcal{H}}_\alpha^-(0) = 0 \quad \text{and } \tilde{\mathcal{H}}_\alpha^+(c_n) = \tilde{\mathcal{H}}_\alpha^-(c_n) = \tilde{a}_{\alpha,n} > 0. \tag{63}$$

P r o o f.
(I) By Theorem 1 in Vajda and Vašek [20], for every $0 \le e \le c_n$

$$e(\pi) = e \quad \text{implies} \quad \tilde{H}_\alpha^-(e) \le \tilde{H}_\alpha(\pi) \le \tilde{H}_\alpha^+(e) \tag{64}$$

where the lower and upper bounds $H_\alpha^\pm(e)$ are attained by the informations $H_\alpha(\pi^\pm)$ for the special distributions

$$\pi^+ = \left(1-e, \frac{e}{n-1}, \frac{e}{n-1}, \dots, \frac{e}{n-1}\right)$$

and

$$\pi^- = (1-e, 1-e, \dots, 1-e, 1-k(1-e), 0, 0, \dots, 0)$$

provided $c_k \le e \le c_{k+1}$ for $1 \le k \le n-1$. Hence for $\alpha \ne 1$

$$\tilde{H}_\alpha^+(e) = \tilde{H}_\alpha(\pi^+) = \frac{1}{\alpha-1}\left[n-1-e^\alpha-(n-1)\left(1-\frac{e}{n-1}\right)^\alpha\right] \tag{65}$$

and

$$\tilde{H}_\alpha^-(e) = \tilde{H}_\alpha(\pi^-) = \frac{k-ke^\alpha-k^\alpha(1-e)^\alpha}{\alpha-1} \tag{66}$$

when $c_k \le e \le c_{k+1}$ and $1 \le k \le n-1$. For $\alpha = 1$ we get

$$\tilde{H}_1^+(e) = \tilde{H}_1(\pi^+) = \lim_{\alpha\to1} \tilde{H}_\alpha^+(e) = -e\ln e - (n-1-e)\ln\left(\frac{n-1-e}{n-1}\right) \tag{67}$$

and

$$\tilde{H}_1^-(e) = \tilde{H}_1^-(\pi^-) = \lim_{\alpha\to1} \tilde{H}_\alpha^-(e) = -ke - k(1-e)\ln[k(1-e)] \tag{68}$$

on the intervals $c_k \le e \le c_{k+1}$ for $1 \le k \le n-1$.

(II) Consider now arbitrary parameter $\alpha > 0$, arbitrary constants $0 \le \tilde{c} < c \le c_n$ and arbitrary distributions $\pi, \tilde{\pi}$ such that $e(\pi) = c$ and $\tilde{e}(\tilde{\pi}) = \tilde{c}$. Then the linear function

$$t\tilde{H}_\alpha(\pi) + (1-t)\tilde{H}_\alpha(\tilde{\pi}) \text{ of variable } 0 \le t \le 1$$

must be bounded above by the function $\tilde{\mathcal{H}}_\alpha^+(tc + (1-t)\tilde{c})$ and bounded below by the function $\tilde{\mathcal{H}}_\alpha^-((tc + (1-t)\tilde{c})$. Similarly as in the previous proof, this implies that $\tilde{\mathcal{H}}_\alpha^+$ must be concave and $\tilde{\mathcal{H}}_\alpha^-$ convex on the interval $[\tilde{c}, c] \subseteq [0,1]$. At the same time $\tilde{\mathcal{H}}_\alpha^+$ must be minimal but above $\tilde{H}_\alpha^+$ and $\tilde{\mathcal{H}}_\alpha^-$ must be maximal but below $\tilde{H}_\alpha^-$. Since $\tilde{H}_\alpha^+$ is concave itself, this implies $\tilde{\mathcal{H}}_\alpha^+ = \tilde{H}_\alpha^+$ so that (61) follows from (65) and (67). On the other hand, $\tilde{H}_\alpha^-$ given by (66) and (68) is piecewise concave in the intervals between the cutpoints c_k, $1 \le k \le n-1$. The piecewise linear function $\tilde{\Phi}_\alpha(t)$ of variable $t \in [0, c_n]$ connecting the points $[c_k, \tilde{H}_\alpha^-(c_k)] \equiv [c_k, \tilde{a}_k]$ for $1 \le k \le n$ is

$$\tilde{\Phi}_\alpha(t) = \tilde{a}_{\alpha,k} + \tilde{b}_{\alpha,k}(t - c_k) \quad \text{for} \quad c_k \le t \le c_{k+1}, \quad 1 \le k \le n-1.$$

This function is convex (concave) if the sequence

$$\frac{\tilde{\Phi}_\alpha(c_k)}{c_k} = \frac{\tilde{a}_{\alpha,k}}{c_k} = \begin{cases} \frac{k}{\alpha-1}\left[1 - \left(\frac{k-1}{k}\right)^{\alpha-1}\right] & \text{if } \alpha \ne 1 \\ \lim_{\alpha\to 1} \tilde{a}_{\alpha,k} = -k \ln \frac{k-1}{k} & \text{if } \alpha = 1 \end{cases}$$

is increasing (decreasing) for $k = 2, 3, \ldots$ Obviously, it is constant equal 1 if $\alpha = 2$, decreasing if $0 < \alpha < 2$ and increasing if $\alpha > 2$. Therefore $\mathcal{H}_\alpha^-(e_B) = \Phi_\alpha(e_B)$ if $\alpha > 2$ and $\mathcal{H}_\alpha^-(e_B)$ is linear in e_B equal $[\Phi_\alpha(c_n) - \Phi_\alpha(0)]\, e_B/c_n \equiv a_n e_B/c_n$ if $0 < \alpha \le 2$. This proves (62). The last assertion including the equations (63) follow from what was already proved above. □

Remark 4.3. The information bounds of Theorem 4.2 seem to be a new result.

In Figures C.3 and C.4 of Appendix C we see the curves $\tilde{\mathcal{H}}_\alpha^\pm(e_B)$ as functions of variable e_B for $\alpha = 1/2, 1, 2$ and $\alpha = 3, 5, 8$.

Remark 4.4. It is deductible from Figures C.3, C.4, and easily verified also formally, that all adjoint power information bounds (61) – (68) are for all $\alpha > 0$ continuous functions strictly increasing on their definition domain $0 \le e$, $e_B \le c_n$ from the minimum 0 to the maximum $\tilde{a}_{\alpha,n}$. Therefore the inverse functions

$$\tilde{e}_\alpha^\mp(\tilde{H}) = \max_{\tilde{H}_\alpha^\pm(e) \le \tilde{H}} e \quad \text{and} \quad \tilde{e}_{B,\alpha}^\mp(\tilde{H}) = \max_{\tilde{H}_\alpha^\pm(e_B) \le \tilde{H}} e_B \tag{69}$$

(notice the reversed order of $\pm$ and $\mp$!) are continuously increasing on their definition domain $0 \le \tilde{H} \le \tilde{a}_{\alpha,n}$ from 0 to c_n at the endpoints but achieving different values

$$\tilde{e}_\alpha^-(\tilde{H}) < \tilde{e}_\alpha^+(\tilde{H}) \quad \text{and} \quad \tilde{e}_{B,\alpha}^-(\tilde{H}) < \tilde{e}_{B,\alpha}^+(\tilde{H}) \tag{70}$$

between the endpoints. Similarly as in Remark 4.2, by plugging the prior, posterior and overall alternative power information measures $\tilde{H}_\alpha(\pi)$, $\tilde{H}_\alpha(\pi_x)$ and $\tilde{H}_\alpha(\mathcal{E})$ in (70) we obtain the attainable upper and lower estimates $\tilde{e}_\alpha^\pm(\tilde{H}_\alpha(\pi))$, $\tilde{e}_\alpha^\pm(\tilde{H}_\alpha(\pi_x))$ and $\tilde{e}_{B,\alpha}^\pm(\tilde{H}_\alpha(\mathcal{E}))$ of the prior, posterior and average Bayes errors $e(\pi)$, $e(\pi_x)$ and $e_B = e_B(\mathcal{E})$. These estimates are compared with those of Remark 4.2 in the next section.

5. INACCURACIES OF INFORMATION CRITERIA

Section 4 shows that the Bayes decision errors

$$e \in \{e(\pi), e(\pi_x), e_B(\mathcal{E})\} \tag{71}$$

depend on the levels achieved by the respective information measures

$$H_\alpha \in \{H_\alpha(\pi), H_\alpha(\pi_x), H_\alpha(\mathcal{E})\} \text{ and } \tilde{H}_\alpha \in \{\tilde{H}_\alpha(\pi), \tilde{H}_\alpha(\pi_x), \tilde{H}_\alpha(\mathcal{E})\} \tag{72}$$

and vice versa. We remind that the range of the errors e is the interval $[0, c_n]$ and the range of the power information H_α or the adjoint power information $\tilde{H}_\alpha$ is the interval $[0, a_{\alpha,n}]$ or $[0, \tilde{a}_{\alpha,n}]$ respectively, where

$$c_n = \frac{n-1}{n}, \quad a_{\alpha,n} = \begin{cases} (n^{1-\alpha} - 1)/(1-\alpha) & \text{if } \alpha \neq 1 \\ \ln n & \text{if } \alpha = 1 \end{cases}$$

and

$$\tilde{a}_{\alpha,n} = \begin{cases} \frac{n-1}{1-\alpha}\left[\left(\frac{n}{n-1}\right)^{1-\alpha} - 1\right] & \text{if } \alpha \neq 1 \\ (n-1)\ln\frac{n}{n-1} & \text{if } \alpha = 1. \end{cases}$$

This section studies the inaccuracies of estimation of the information measures (72) by means of the errors (71) and vice versa. For simplicity, we restrict ourselves to the Bayes errors and residual informations

$$e_B = e_B(\mathcal{E}) \text{ and } H_\alpha = H_\alpha(\mathcal{E}), \quad \tilde{H}_\alpha = \tilde{H}_\alpha(\mathcal{E})$$

and the corresponding bounds

$$\mathcal{H}_\alpha^-(e_B) \leq \mathcal{H}_\alpha^+(e_B), \quad \tilde{\mathcal{H}}_\alpha^-(e_B) \leq \tilde{\mathcal{H}}_\alpha^+(e_B) \tag{73}$$

and

$$e_{B,\alpha}^-(H_\alpha) < e_{B,\alpha}^+(H_\alpha), \quad \tilde{e}_{B,\alpha}^-(\tilde{H}_\alpha) < \tilde{e}_{B,\alpha}^+(\tilde{H}_\alpha) \tag{74}$$

established by Theorems 4.1, 4.2 and their remarks. Alternative results for the prior Bayes errors and power informations

$$e = e(\pi) \text{ and } H_\alpha = H_\alpha(\pi), \quad \tilde{H}_\alpha = \tilde{H}_\alpha(\pi),$$

and for the corresponding bounds $H_\alpha^\pm(e), \tilde{H}_\alpha^\pm(e)$ and $e_\alpha^\pm(H_\alpha), \tilde{e}_\alpha^\pm(\tilde{H}_\alpha)$ mentioned or established in previous section, are obtained in asimilar manner.

By (73), under a given Bayes decision error e_B the corresponding residual informations H_α and $\tilde{H}_\alpha$ are restricted to the intervals $[\mathcal{H}_\alpha^-(e_B), \mathcal{H}_\alpha^+(e_B)]$ and $[\tilde{\mathcal{H}}_\alpha^-(e_B), \tilde{\mathcal{H}}_\alpha^+(e_B)]$ which are their tight estimates in the sense that all values of the intervals are achievable by the informations. Therefore the interval lengths $\mathcal{H}_\alpha^+(e_B) - \mathcal{H}_\alpha^-(e_B)$ and $\tilde{\mathcal{H}}_\alpha^+(e_B) - \tilde{\mathcal{H}}_\alpha^-(e_B)$ are realistic local measures of inaccuracy of these estimates and the *average inaccuracies*

$$AI_n(H_\alpha|e_B) = \frac{1}{c_n}\int_0^{c_n} [\mathcal{H}_\alpha^+(e) - \mathcal{H}_\alpha^-(e))]\,\mathrm{d}e \tag{75}$$

and

$$AI_n(\tilde{H}_\alpha|e_B) = \frac{1}{c_n}\int_0^{c_n} [\tilde{\mathcal{H}}_\alpha^+(e) - \tilde{\mathcal{H}}_\alpha^-(e)]\,\mathrm{d}e \tag{76}$$

are natural and realistic global measures of accuracy of these estimates. They can be used to select the versions of the conditional entropies H_α and $\tilde{H}_\alpha$ most accurately determined by the Bayes decision error e_B.

Similarly, under given conditional entropies H_α and $\tilde{H}_\alpha$ the Bayes decision error e_B is restricted to the intervals $[e^-_{B,\alpha}(H_\alpha), e^+_{B,\alpha}(H_\alpha)]$ and $[\tilde{e}^-_{B,\alpha}(\tilde{H}_\alpha), \tilde{e}^+_{B,\alpha}(\tilde{H}_\alpha)]$ where all values are achievable. Hence these intervals represent the most tight estimates of these entropies by means of the error e_B. The interval lengths $e^+_{B,\alpha}(H_\alpha) - e^-_{B,\alpha}(H_\alpha)$ and $\tilde{e}^+_{B,\alpha}(\tilde{H}_\alpha) - \tilde{e}^-_{B,\alpha}(\tilde{H}_\alpha)$ are suitable local measures of inaccuracy of these estimates and the *average inaccuracies*

$$AI_{n,\alpha}(e_B|H_\alpha) = \frac{1}{a_{\alpha,n}} \int_0^{a_{\alpha,n}} [e^+_{B,\alpha}(H) - e^-_{B,\alpha}(H)] \,\mathrm{d}H \tag{77}$$

and

$$AI_{n,\alpha}(e_B|\tilde{H}_\alpha) = \frac{1}{\tilde{a}_{\alpha,n}} \int_0^{\tilde{a}_{\alpha,n}} [\tilde{e}^+_{B,\alpha}(\tilde{H}) - \tilde{e}^-_{B,\alpha}(\tilde{H})] \,\mathrm{d}\tilde{H} \tag{78}$$

are natural global measures of accuracy of these estimation procedures. They can be used to select the versions of the conditional entropies H_α and $\tilde{H}_\alpha$ most suitable for estimation the Bayes decision error e_B.

Lemma 5.1. The power information bounds $\mathcal{H}^\pm_\alpha(e_B)$ satisfy the integral formulas

$$\int_0^{c_n} \mathcal{H}^+_\alpha(e)\,\mathrm{d}e = \begin{cases} \frac{1}{\alpha-1}\left[\frac{n-1}{n} - \frac{n^\alpha+n-2}{(\alpha+1)n^\alpha}\right] & \text{if } \alpha \neq 1 \\ \frac{1}{2n}\left[n-1+(n-2)\ln n\right] & \text{if } \alpha = 1 \end{cases} \tag{79}$$

and

$$\int_0^{c_n} \mathcal{H}^-_\alpha(e)\,\mathrm{d}e = \begin{cases} \frac{1}{2(\alpha-1)} \sum_{k=1}^{n-1} \frac{2-k^{1-\alpha}-(k+1)^{1-\alpha}}{k(k+1)} & \text{if } 0<\alpha<2, \alpha\neq 1 \\ \frac{1}{2}\sum_{k=1}^{n-1} \frac{\ln[k(k+1)]}{k(k+1)} & \text{if } \alpha = 1 \\ \frac{(n-1)(1-n^{1-\alpha})}{2(\alpha-1)n} & \text{if } \alpha \geq 2. \end{cases} \tag{80}$$

Proof. For $\alpha \neq 1$ the result of (79) follows by a routine integration of the power functions of $e = e_B$ appearing in the formula (52) for the upper bound $\mathcal{H}^+_\alpha(e) = \mathcal{H}^+_\alpha(e_B)$. For $\alpha = 1$ this result can be obtained by taking the limit for $\alpha \to 1$ in the already proved version of the formula (79) for $\alpha \neq 1$ since the integrand is bounded and continuous in the parameter α from the neighborhood of $\alpha = 1$. An alternative possibility is to integrate the function $\mathcal{H}^+_1(e)$ with the use of the formula

$$\int x \ln x \,\mathrm{d}x = \frac{x^2}{2}\left(\ln x - \frac{1}{2}\right) \tag{81}$$

obtained by differentiating the function $x^2 \ln x$. The upper and lower formulas of (80) follow by a routine integration of the linear or piecewise linear functions of $e = e_B$

appearing in the formulas (53) for the lower bound $\mathcal{H}_\alpha^-(e) = \mathcal{H}_\alpha^-(e_B)$. The middle formula of (80) can be obtained similarly as above, by taking the limit for $\alpha \to 1$ in the already proved upper formula of (80). Alternatively, we can integrate the piecewise linear function $\mathcal{H}_1^-(e) = \mathcal{H}_1^-(e_B)$ of (50). Details can be found in Appendix A. □

Formula (79) was obtained previously by Vajda and Zvárová [21]. Formula (80) is new as well as both formulas of the next lemma.

Lemma 5.2. The alternative power entropy bounds $\tilde{\mathcal{H}}_\alpha^\pm(e_B)$ satisfy the integral formulas

$$\int_0^{c_n} \tilde{\mathcal{H}}_\alpha^+(e_B)\,\mathrm{d}e_B = \begin{cases} \frac{1}{\alpha-1}\left[\frac{(n-1)^2}{n} - \frac{(n-1)^2}{\alpha+1} + \frac{n(n-2)}{\alpha+1}\left(\frac{n-1}{n}\right)^{\alpha+1}\right] & \text{if } \alpha \neq 1 \\ \frac{(n-1)^2}{2n}\left[1 + (n-2)\ln\frac{n-1}{n}\right] & \text{if } \alpha = 1 \end{cases} \tag{82}$$

and

$$\int_0^{c_n} \tilde{\mathcal{H}}_\alpha^-(e_B)\,\mathrm{d}e_B = \begin{cases} \frac{(n-1)^2}{2n(\alpha-1)}\left[1 - \left(\frac{n-1}{n}\right)^{\alpha-1}\right] & \text{if } 0 < \alpha \leq 2, \alpha \neq 1 \\ \frac{(n-1)^2}{2n}\ln\frac{n}{n-1} & \text{if } \alpha = 1 \\ \frac{1}{2(\alpha-1)}\sum_{k=1}^{n-1}\frac{2k-1-(k-1)\left(\frac{k-1}{k}\right)^{\alpha-1} - k\left(\frac{k}{k+1}\right)^{\alpha-1}}{k(k+1)} & \text{if } \alpha > 2. \end{cases} \tag{83}$$

P r o o f. Similarly as in the previous proof, for $\alpha \neq 1$ the result of (82) follows by a routine integration of the power functions of $e = e_B$ appearing in the formula (61) for the upper bound $\tilde{\mathcal{H}}_\alpha^+(e_B)$. For $\alpha = 1$ this result can be obtained by taking the limit for $\alpha \to 1$ in the already proved version of the formula (82) for $\alpha \neq 1$ since the integrand $\tilde{\mathcal{H}}_\alpha^+(e) = \tilde{\mathcal{H}}_\alpha^+(e_B)$ is bounded and continuous in the parameter α from the neighborhood of $\alpha = 1$. Again, an alternative is to integrate $\tilde{\mathcal{H}}_1^+(e)$ using (81). The upper and lower formulas of (83) follow by a routine integration of the linear or piecewise linear functions of $e = e_B$ appearing in the formulas (62) for the lower bound $\tilde{\mathcal{H}}_\alpha^-(e) = \tilde{\mathcal{H}}_\alpha^-(e_B)$. The middle formula of (83) can be obtained by taking the limit for $\alpha \to 1$ in the already proved upper formula of (83) since the integrand $\tilde{\mathcal{H}}_\alpha^-(e)$ is bounded and continuous in the parameter α from the neighborhood of $\alpha = 1$. Details can be found in Appendix A. □

Theorem 5.1. The average inaccuracies $AI_n(H_\alpha|e_B)$ and $AI_n(\tilde{H}_\alpha|e_B)$ of the estimation of the power informations $H_\alpha = H_\alpha(\mathcal{E})$ and $\tilde{H}_\alpha = \tilde{H}_\alpha(\mathcal{E})$ by means of the Bayes error $e_B = e_B(\mathcal{E})$ are given by the formulas

$$AI_n(H_\alpha|e_B) = \frac{1}{c_n}\left(\int_0^{c_n}\mathcal{H}_\alpha^+(e)\,\mathrm{d}e - \int_0^{c_n}\mathcal{H}_\alpha^-(e)\,\mathrm{d}e\right) \tag{84}$$

and

$$AI_n(\tilde{H}_\alpha|e_B) = \frac{1}{c_n}\left(\int_0^{c_n}\tilde{\mathcal{H}}_\alpha^+(e)\,\mathrm{d}e - \int_0^{c_n}\tilde{\mathcal{H}}_\alpha^-(e)\,\mathrm{d}e\right) \tag{85}$$

where the integrals are given by lemmas 5.1 and 5.2.

P r o o f. Clear from (75), (76) and lemmas 5.1 and 5.2. □

Theorem 5.2. The average inaccuracies $AI_n(e_B|H_\alpha)$ and $AI_n(e_B|\tilde{H}_\alpha)$ of estimation of the Bayes error $e_B = e_B(\mathcal{E})$ by means of the power informations $H_\alpha = H_\alpha(\mathcal{E})$ and $\tilde{H}_\alpha = \tilde{H}_\alpha(\mathcal{E})$ are given by the formulas

$$AI_{n,\alpha}(e_B|H_\alpha) = \frac{1}{a_{\alpha,n}}\left(\int_0^{c_n} \mathcal{H}_\alpha^+(e)\,\mathrm{d}e - \int_0^{c_n} \mathcal{H}_\alpha^-(e)\,\mathrm{d}e\right) \tag{86}$$

and

$$AI_{n,\alpha}(e_B|\tilde{H}_\alpha) = \frac{1}{\tilde{a}_{\alpha,n}}\left(\int_0^{c_n} \tilde{\mathcal{H}}_\alpha^+(e)\,\mathrm{d}e - \int_0^{c_n} \tilde{\mathcal{H}}_\alpha^-(e)\,\mathrm{d}e\right) \tag{87}$$

where the integrals are given by lemmas 5.1 and 5.2.

P r o o f. By the definitions of functions (77) and (78), the area $c_n.a_{\alpha,n}$ of the rectangle $(0, c_n)\otimes(0, a_{\alpha,n})$ representing the domain of e_B (range of $e_{B,\alpha}^-(H_\alpha)$) and range of $\mathcal{H}_\alpha^+(e_B)$ (domain of H_α) must be the sum of integrals

$$\int_0^{c_n} \mathcal{H}_\alpha^+(e)\,\mathrm{d}e + \int_0^{a_{\alpha,n}} e_{B,\alpha}^-(H)\,\mathrm{d}H.$$

Similarly we get

$$c_n.a_{\alpha,n} = \int_0^{c_n} \mathcal{H}_\alpha^-(e)\,\mathrm{d}e + \int_0^{a_{\alpha,n}} e_{B,\alpha}^+(H)\,\mathrm{d}H,$$

$$c_n.\tilde{a}_{\alpha,n} = \int_0^{c_n} \tilde{\mathcal{H}}_\alpha^+(e)\,\mathrm{d}e + \int_0^{\tilde{a}_{\alpha,n}} \tilde{e}_{B,\alpha}^-(\tilde{H})\,\mathrm{d}\tilde{H}$$

and

$$c_n.\tilde{a}_{\alpha,n} = \int_0^{c_n} \tilde{\mathcal{H}}_\alpha^-(e)\,\mathrm{d}e + \int_0^{\tilde{a}_{\alpha,n}} \tilde{\mathcal{E}}_\alpha^+(\tilde{H})\,\mathrm{d}\tilde{H}.$$

The desired relations are clear from here and from definitions (77), (78). □

Functions $AI_n(H_\alpha|e_B)$, $AI_n(\tilde{H}_\alpha|e_B)$, $AI_{n,\alpha}(e_B|H_\alpha)$ and $AI_{n,\alpha}(e_B|\tilde{H}_\alpha)$ of variable $0 < \alpha < 8$ for the selected values of $n = 2, 4, 8$ and 20 are shown in Figures D.1 – D.4 of Appendix D and the numerical values for $2 \le n \le 1000$ are in Tables B.1 – B.4 of Appendix B. We see from these results that $AI_{n,\alpha}(e_B|H_\alpha)$ and $AI_{n,\alpha}(e_B|\tilde{H}_\alpha)$ are minimized at $\alpha = 2$ for all $n \ge 2$. The minima of $AI_n(\tilde{H}_\alpha|e_B)$ are achieved at $\alpha = 2$ only for $n > 4$. The remaining minima of $AI_n(\tilde{H}_\alpha|e_B)$, as well as all minima of $AI_n(H_\alpha|e_B)$, are achieved at infinite α.

Conclusion 5.1 The fact that the average inaccuracy $AI_{n,\alpha}(e_B|H_\alpha)$ is minimized at $\alpha = 2$ indicates that among the various information criteria H_α including the Shannon's H_1 used in the literature to estimate the Bayes error e_B, the most accurate is the quadratic entropy H_2 suggested in Vajda [18]. This result answers the problem posed in Vajda [19].

Conclusion 5.2 By comparing Figures D.3 and D.4 or Tables B.3 and B.4 one can see that the inaccuracies $AI_{n,\alpha}(e_B|\tilde{H}_\alpha)$ are slightly lower than $AI_{n,\alpha}(e_B|H_\alpha)$ for almost all powers α and state space sizes n. The only exception is the optimal power $\alpha = 2$ where $AI_{n,\alpha}(e_B|\tilde{H}_\alpha)$ coincides with $AI_{n,\alpha}(e_B|H_\alpha)$. Therefore the alternative power entropies $\tilde{H}_\alpha$ are in general slightly better than the classical power entropies H_α for estimating the Bayes decision errors and the Bayes risks, but the optimal versions for $\alpha = 2$ are equivalent to H_2.

6. INFORMATION CRITERIA IN GENERAL MODEL

In this section are proposed new estimates of Bayes risk obtained by plugging into the estimates of Section 5 the bounds obtained in Section 3. These estimates together with the results on optimality of the information criteria appearing in these estimates obtained in Section 4 represent the main results of this paper.

Throughout this section we consider the general decision situation of Section 2 with losses (11) on a space Θ of size

$$n = |\Theta|$$

and with an experiment $\mathcal{E}$ (cf. (11)). The losses are characterized by the median loss and the loss range

$$\Lambda > 0, \quad \Delta \geq 0 \quad \text{cf. (31)} \tag{88}$$

and the whole decision situation is characterized by the prior Bayes loss, posterior Bayes loss and Bayes risk

$$L_B(\pi),\;\; L_B(\pi_x) \;\text{ and }\; R_B = R_B(\mathcal{E}) \quad \text{(cf. (15) and (19))} \tag{89}$$

respectively.

In the next theorem the knowledge about experiment $\mathcal{E}$ is represented by the power information measures

$$H_\alpha(\pi),\;\; H_\alpha(\pi_x) \;\text{ and }\; H_\alpha(\mathcal{E}) \;\text{ for some } \alpha > 0 \text{ (cf. (37) and (39))} \tag{90}$$

respectively. We study the tight upper bounds

$$H_\alpha^+(L_B|\Lambda,\Delta) = \max_{L_B(\pi)=L_B} H_\alpha(\pi) \equiv \max_{L_B(\pi_x)=L_B} H_\alpha(\pi_x) \tag{91}$$

$$\mathcal{H}_\alpha^+(R_B|\Lambda,\Delta) = \max_{R_B(\mathcal{E})=R_B} H_\alpha(\mathcal{E}) \tag{92}$$

and the tight lower bounds

$$H_\alpha^-(L_B|\Lambda,\Delta) = \min_{L_B(\pi)=L_B} H_\alpha(\pi) \equiv \min_{L_B(\pi_x)=L_B} H_\alpha(\pi_x), \tag{93}$$

$$\mathcal{H}_\alpha^-(R_B|\Lambda,\Delta) = \min_{R_B(\mathcal{E})=R_B} H_\alpha(\mathcal{E}) \tag{94}$$

for these entropies at given values of the prior and posterior Bayes losses and the Bayes risk appearing in (89), respectively.

Theorem 6.1. The bounds (91) – (94) are given in the whole definition domains

$$0 \leq L_B \leq c_n(\Lambda + \Delta/2) \quad \text{and} \quad 0 \leq R_B \leq c_n(\Lambda + \Delta/2) \tag{95}$$

by the formulas

$$H_\alpha^\pm(L_B|\Lambda,\Delta) = H_\alpha^\pm\left(\frac{L_B}{1 \mp \Delta/2}\right) \quad \text{and} \quad \mathcal{H}_\alpha^\pm(R_B|\Lambda,\Delta) = \mathcal{H}_\alpha^\pm\left(\frac{R_B}{1 \mp \Delta/2}\right) \tag{96}$$

for $H_\alpha^\pm(\cdot)$, $\mathcal{H}_\alpha^\pm(\cdot)$ defined in the domain $[0, c_n]$ by (47) – (50) and extended to $t > c_n$ by

$$H_\alpha^+(t) = H_\alpha^+(c_n) \equiv a_{\alpha,n}, \quad \mathcal{H}_\alpha^+(t) = \mathcal{H}_\alpha^+(c_n) \equiv a_{\alpha,n} \tag{97}$$

where

$$c_n = \frac{n}{n-1} \quad \text{and} \quad a_{\alpha,n} = \begin{cases} \frac{n^{1-\alpha}-1}{1-\alpha} & \text{if } \alpha \neq 1 \\ \ln n & \text{if } \alpha = 1 \end{cases} \quad \text{(cf. (44))}. \tag{98}$$

P r o o f . By Theorem 3.1, The Bayes errors $e_B(\pi)$, $e_B(\pi_x)$ and $e_B = e_B(\mathcal{E})$ are restricted by the bounds

$$\frac{L_B(\pi)}{\Lambda + \Delta/2} \leq e_B(\pi) \leq \max\left\{\frac{L_B(\pi)}{1 - \Delta/2}, c_n\right\} \tag{99}$$

$$\frac{L_B(\pi_x)}{\Lambda + \Delta/2} \leq e_B(\pi_x) \leq \max\left\{\frac{L_B(\pi_x)}{1 - \Delta/2}, c_n\right\} \tag{100}$$

$$\frac{R_B(\mathcal{E})}{\Lambda + \Delta/2} \leq e_B(\mathcal{E}) \leq \max\left\{\frac{R_B(\mathcal{E})}{1 + \Delta/2}, c_n\right\} \tag{101}$$

for c_n given by (98) and these bounds are tight. Applying these bounds in the definitions (91) – (94) of $H_\alpha^\pm(L_B|\Lambda,\Delta)$ and $\mathcal{H}_\alpha^\pm(R_B|\Lambda,\Delta)$ and using the definitions (47) – (50) of $H_\alpha^\pm(e)$ and $\mathcal{H}_\alpha^\pm(e_B)$ we get the desired formulas (96). The new bounds (91) – (94) are attained because the initial bounds (47) – (50) were proved to be attained. □

From the bounds of Theorem 6.1 we obtain the tight upper bounds

$$L_{B,\alpha}^+(H|\Lambda,\Delta) = \max_{H_\alpha(\pi)=H} L_B(\pi) = \max_{H_\alpha(\pi_x)=H} L_B(\pi_x) \tag{102}$$

$$R_{B,\alpha}^+(H|\Lambda,\Delta) = \max_{H_\alpha(\mathcal{E})=H} R_B(\mathcal{E}) \tag{103}$$

and the tight lower bounds

$$L_{B,\alpha}^-(H|\Lambda,\Delta) = \min_{H_\alpha(\pi)=H} L_B(\pi) = \min_{H_\alpha(\pi_x)=H} L_B(\pi_x) \tag{104}$$

$$R_{B,\alpha}^-(H|\Lambda,\Delta) = \min_{H_\alpha(\mathcal{E})=H} R_B(\mathcal{E}) \tag{105}$$

of the Bayes losses and risks (89) in the models with loss parameters Λ, Δ and given values of the power informations (90).

Corollary 6.1. The tight upper and lower bounds (102) – (105) are given in the corresponding definition domains

$$0 \le H \le a_{\alpha,n}, \quad \alpha > 0 \tag{106}$$

of the power informations (90) by the formulas

$$L^{\pm}_{B,\alpha}(H|\Lambda,\Delta) = e^{\pm}_{\alpha}(H)(\Lambda \pm \Delta/2), \quad R^{\pm}_{B,\alpha}(H|\Lambda,\Delta) = e^{\pm}_{B,\alpha}(H)(\Lambda \pm \Delta/2) \tag{107}$$

for $e^{\pm}_{\alpha}(H)$, $e^{\pm}_{B,\alpha}(H)$ defined by (56).

Now we deal with the situation where the knowledge about experiment $\mathcal{E}$ is represented by adjoint power information measures

$$\tilde{H}_\alpha(\pi), \;\; \tilde{H}_\alpha(\pi_x) \;\text{ and }\; \tilde{H}_\alpha(\mathcal{E}) \;\text{ for some } \alpha > 0 \text{ (cf. (39))} \tag{108}$$

respectively. We study the tight upper bounds

$$\tilde{H}^{+}_{\alpha}(L_B|\Lambda,\Delta) = \max_{L_B(\pi)=L_B} \tilde{H}_\alpha(\pi) \equiv \max_{L_B(\pi_x)=L_B} \tilde{H}_\alpha(\pi_x) \tag{109}$$

$$\tilde{\mathcal{H}}^{+}_{\alpha}(R_B|\Lambda,\Delta) = \max_{R_B(\mathcal{E})=R_B} \tilde{H}_\alpha(\mathcal{E}) \tag{110}$$

and the tight lower bounds

$$\tilde{H}^{-}_{\alpha}(L_B|\Lambda,\Delta) = \min_{L_B(\pi)=L_B} \tilde{H}_\alpha(\pi) \equiv \min_{L_B(\pi_x)=L_B} \tilde{H}_\alpha(\pi_x), \tag{111}$$

$$\tilde{\mathcal{H}}^{-}_{\alpha}(R_B|\Lambda,\Delta) = \min_{R_B(\mathcal{E})=R_B} \tilde{H}_\alpha(\mathcal{E}) \tag{112}$$

of these information measures for given values of the prior and posterior Bayes losses and the Bayes risk appearing in (89).

Theorem 6.2. The bounds (91) – (94) are given in the whole definition domains

$$0 \le L_B \le c_n(\Lambda + \Delta/2) \quad \text{and} \quad 0 \le R_B \le c_n(\Lambda + \Delta/2) \tag{113}$$

by the formulas

$$\tilde{H}^{\pm}_{\alpha}(L_B|\Lambda,\Delta) = \tilde{H}^{\pm}_{\alpha}\left(\frac{L_B}{1 \mp \Delta/2}\right) \quad \text{and} \quad \tilde{\mathcal{H}}^{\pm}_{\alpha}(R_B|\Lambda,\Delta) = \tilde{\mathcal{H}}^{\pm}_{\alpha}\left(\frac{R_B}{1 \mp \Delta/2}\right) \tag{114}$$

for $\tilde{H}^{\pm}_{\alpha}(\cdot)$, $\mathcal{H}^{\pm}_{\alpha}(\cdot)$ defined in the domain $[0, c_n]$ by (61), (62) and for $\tilde{H}^{+}_{\alpha}(\cdot)$, $\mathcal{H}^{+}_{\alpha}(\cdot)$ extended to $t > c_n$ by

$$\tilde{H}^{+}_{\alpha}(t) = \tilde{H}^{+}_{\alpha}(c_n) \equiv \tilde{a}_{\alpha,n}, \quad \tilde{\mathcal{H}}^{+}_{\alpha}(t) = \tilde{\mathcal{H}}^{+}_{\alpha}(c_n) \equiv \tilde{a}_{\alpha,n} \tag{115}$$

where

$$c_n = \frac{n}{n-1} \quad \text{and} \quad \tilde{a}_{\alpha,n} = \begin{cases} \frac{n-1}{1-\alpha}\left[\left(\frac{n}{n-1}\right)^{1-\alpha} - 1\right] & \text{if } \alpha \ne 1 \\ (n-1)\ln\frac{n}{n-1} & \text{if } \alpha = 1 \end{cases} \quad \text{(cf. (44)).} \tag{116}$$

P r o o f. By Theorem 3.1, the Bayes errors $e_B(\pi)$, $e_B(\pi_x)$ and $e_B = e_B(\mathcal{E})$ are restricted by the bounds (99) – (101) for c_n given by (98) and these bounds are tight. Applying these bounds in the definitions (109) – (112) of $\tilde{H}^{\pm}_{\alpha}(L_B|\Lambda,\Delta)$ and $\tilde{\mathcal{H}}^{\pm}_{\alpha}(R_B|\Lambda,\Delta)$ and using the definitions (61), (62) of $\tilde{H}^{\pm}_{\alpha}(e)$ and $\tilde{\mathcal{H}}^{\pm}_{\alpha}(e_B)$ we get the desired formulas (114). The new bounds (114) are attained because the initial bounds (61), (62) were proved to be attained. □

From the bounds of Theorem 6.2 we obtain the tight upper bounds

$$L^{+}_{B,\alpha}(\tilde{H}|\Lambda,\Delta) = \max_{\tilde{H}_\alpha(\pi)=\tilde{H}} L_B(\pi) \equiv \max_{\tilde{H}_\alpha(\pi_x)=\tilde{H}} L_B(\pi_x) \tag{117}$$

$$R^{+}_{B,\alpha}(\tilde{H}|\Lambda,\Delta) = \max_{\tilde{H}_\alpha(\mathcal{E})=\tilde{H}} R_B(\mathcal{E}) \tag{118}$$

and the tight lower bounds

$$L^{-}_{B,\alpha}(\tilde{H}|\Lambda,\Delta) = \min_{\tilde{H}_\alpha(\pi)=\tilde{H}} L_B(\pi) \equiv \min_{\tilde{H}_\alpha(\pi_x)=\tilde{H}} L_B(\pi_x) \tag{119}$$

$$R^{-}_{B,\alpha}(\tilde{H}|\Lambda,\Delta) = \min_{\tilde{H}_\alpha(\mathcal{E})=\tilde{H}} R_B(\mathcal{E}) \tag{120}$$

of the Bayes losses and risks (89) in models with parameters Λ, Δ and given values of the power entropies (108).

Corollary 6.2. The attainable upper bounds (117) – (120) are given in the definitions domains

$$0 \le \tilde{H} \le \tilde{a}_{\alpha,n}, \quad \alpha > 0 \tag{121}$$

of the adjoint power informations (108) by the formulas

$$L^{\pm}_{B,\alpha}(\tilde{H}|\Lambda,\Delta) = e^{\pm}_{\alpha}(\tilde{H})(\Lambda \pm \Delta/2), \quad R^{\pm}_{B,\alpha}(\tilde{H}|\Lambda,\Delta) = e^{\pm}_{B,\alpha}(\tilde{H})(\Lambda \pm \Delta/2) \tag{122}$$

for $e^{\pm}_{\alpha}(\tilde{H})$, $e^{\pm}_{B,\alpha}(\tilde{H})$ defined by (69).

Conclusion 6.1 Conclusion 5.1 indicates that the average inaccuracy of the interval estimates $\left[R^{-}_{B,\alpha}(H_\alpha|\Lambda,\Delta), R^{+}_{B,\alpha}(H_\alpha|\Lambda,\Delta)\right]$ of the Bayes risk $R_B = R_B(\mathcal{E})$ by means of the power informations $H_\alpha = H_\alpha(\mathcal{E})$ is minimized at the power $\alpha = 2$.

Conclusion 6.2 Conclusion 5.2 indicates that the average inaccuracy of the interval estimates $\left[R^{-}_{B,\alpha}(\tilde{H}_\alpha|\Lambda,\Delta), R^{+}_{B,\alpha}(\tilde{H}_\alpha|\Lambda,\Delta)\right]$ of the Bayes risk $R_B = R_B(\mathcal{E})$ by means of the adjoint power informations $\tilde{H}_\alpha = \tilde{H}_\alpha(\mathcal{E})$ is minimized at the power $\alpha = 2$. Moreover, the alternative power entropies $\tilde{H}_\alpha$ give in general better estimates than the classical power entropies H_α, except the optimal power $\alpha = 2$ where both estimates coincide.

Figures E.1 and E.2 in Appendix E illustrate the power information bounds $H^{\pm}_{\alpha}(L_B|\Lambda,\Delta)$ for the power parameters $\alpha = 1$ and $\alpha = 2$ and the loss function parameters $(\Lambda,\Delta) = (1,0)$ and $(\Lambda,\Delta) = (1,2/5)$ taken from the concrete situation of

Example 3.1. Similar illustrations of the bounds $\mathcal{H}_\alpha^\pm(R_B|\Lambda,\Delta)$ for the same power parameters and loss function parameters are in Figures E.3 and E.4. Inverse functions to the bounds of Figures E.1 – E.4, readable in this figure too, illustrate the corresponding prior and Bayes risk bounds $L_{B,\alpha}^\pm(H|\Lambda,\Delta)$ and $R_{B,\alpha}^\pm(H|\Lambda,\Delta)$.

APPENDIX A.

In this appendix we prove formulas of (79) and of (80) from Lemma 5.1 and formulas of (82) and of (83) from Lemma 5.2.

Lemma 5.1, formulas of (79).

(i) If $\alpha > 0$, $\alpha \neq 1$ then using the upper formula of (52) we obtain

$$\begin{aligned}\int_0^{c_n} \mathcal{H}_\alpha^+(e)\,\mathrm{d}e &= \int_0^{c_n} \frac{1-(n-1)^{1-\alpha}e^\alpha-(1-e)^\alpha}{\alpha-1}\,\mathrm{d}e \\ &= \frac{1}{\alpha-1}\left[e-(n-1)^{1-\alpha}\frac{e^{\alpha+1}}{\alpha+1}+\frac{(1-e)^{\alpha+1}}{\alpha+1}\right]_0^{\frac{n-1}{n}} \\ &= \frac{1}{\alpha-1}\left[\frac{n-1}{n}-\frac{n^\alpha+n-2}{(\alpha+1)n^\alpha}\right].\end{aligned}$$

(ii) If $\alpha = 1$ then we apply in the lower formula of (52) the relations

$$\int x\ln x\,\mathrm{d}x = \frac{x^2}{2}\left[\ln x-\frac{1}{2}\right], \quad \int(1-x)\ln(1-x)\,\mathrm{d}x = -\frac{(1-x)^2}{2}\left[\ln(1-x)-\frac{1}{2}\right]$$

and obtain

$$\begin{aligned}\int_0^{c_n} \mathcal{H}_1^+(e)\,\mathrm{d}e &= \int_0^{c_n} \left[e\ln(n-1)-e\ln e-(1-e)\ln(1-e)\right]\mathrm{d}e \\ &= \left[\frac{e^2}{2}\ln(n-1)-\frac{e^2}{2}\big(\ln e-\frac{1}{2}\big)+\frac{(1-e)^2}{2}\big(\ln(1-e)-\frac{1}{2}\big)\right]_0^{\frac{n}{n-1}} \\ &= \frac{1}{2}\big(\frac{n-1}{n}\big)^2\ln(n-1)-\frac{1}{2}\big(\frac{n-1}{n}\big)^2\big(\ln\frac{n-1}{n}-\frac{1}{2}\big)+\frac{1}{2}\frac{1}{n^2}\big(\ln\frac{1}{n}-\frac{1}{2}\big)+\frac{1}{4} \\ &= \frac{1}{2}\big(\frac{n-1}{n}\big)^2\ln n+\frac{1}{4}\big(\frac{n-1}{n}\big)^2-\frac{1}{2n^2}\ln n-\frac{1}{4n^2}+\frac{1}{4} \\ &= \frac{n-2}{2n}\ln n+\frac{n-2}{4n}+\frac{1}{4}=\frac{1}{2n}\left\{n-1+(n-2)\ln n\right\}.\end{aligned}$$

Lemma 5.1, formulas of (80).

(i) If $0 < \alpha < 2$ then using the upper formula of (53) we obtain

$$\int_0^{c_n} \mathcal{H}_\alpha^-(e)\,\mathrm{d}e = \sum_{k=1}^{n-1} \int_{c_k}^{c_{k+1}} [a_{\alpha,k}+b_{\alpha,k}(e-c_k)]\,\mathrm{d}e = \sum_{k=1}^{n-1} \left[a_{\alpha,k}e+b_{\alpha,k}\frac{e^2}{2}-b_{\alpha,k}c_k e\right]_{c_k}^{c_{k+1}}$$

$$= \sum_{k=1}^{n-1} \left\{a_{\alpha,k}(c_{k+1}-c_k) + b_{\alpha,k}\frac{c_{k+1}^2-c_k^2}{2} - b_{\alpha,k}c_k(c_{k+1}-c_k)\right\}$$

$$= \sum_{k=1}^{n-1} \left\{\frac{1-k^{1-\alpha}}{(\alpha-1)k(k+1)} + \frac{a_{\alpha,k+1}-a_{\alpha,k}}{2}(c_{k+1}-c_k)\right\}$$

$$= \sum_{k=1}^{n-1} \left\{\frac{1-k^{1-\alpha}}{(\alpha-1)k(k+1)} + \frac{k^{1-\alpha}-(k+1)^{1-\alpha}}{(\alpha-1)2k(k+1)}\right\}$$

$$= \frac{1}{2(\alpha-1)} \sum_{k=1}^{n-1} \frac{2-k^{1-\alpha}-(k+1)^{1-\alpha}}{k(k+1)}.$$

(ii) If $\alpha \geq 2$, then using the lower formula of (53) we obtain

$$\int_0^{c_n} \mathcal{H}_\alpha^-(e)\,\mathrm{d}e = \frac{a_{\alpha,n}}{c_n}\int_0^{c_n} e\,\mathrm{d}e = \frac{a_{\alpha,n}}{c_n}\frac{c_n^2}{2} = \frac{a_{\alpha,n}c_n}{2} = \frac{(n-1)(1-n^{1-\alpha})}{2(\alpha-1)n}.$$

Lemma 5.2, formulas (82).

(i) If $\alpha > 0, \alpha \neq 1$, then by the upper formula of (61)

$$\int_0^{c_n} \tilde{\mathcal{H}}_\alpha^+(e)\,\mathrm{d}e = \int_0^{c_n} \frac{(n-1)-e^\alpha-(n-1)^{1-\alpha}(n-1-e)^\alpha}{\alpha-1}\,\mathrm{d}e \quad (x = n-1-e)$$

$$= \frac{1}{\alpha-1}\left\{\frac{(n-1)^2}{n} - \frac{1}{\alpha+1}\left(\frac{n-1}{n}\right)^{\alpha+1} - (n-1)^{1-\alpha}\int_{n-1-c_n}^{n-1} x^\alpha\,\mathrm{d}x\right\}$$

$$= \frac{1}{\alpha-1}\left\{\frac{(n-1)^2}{n} - \frac{1}{\alpha+1}\left(\frac{n-1}{n}\right)^{\alpha+1}\right.$$

$$\left. - \frac{(n-1)^{1-\alpha}}{\alpha+1}\left[(n-1)^{\alpha+1} - \frac{(n-1)^{2(\alpha+1)}}{n^{\alpha+1}}\right]\right\}$$

$$= \frac{1}{\alpha-1}\left[\frac{(n-1)^2}{n} - \frac{(n-1)^2}{\alpha+1} + \frac{n(n-2)}{\alpha+1}\left(\frac{n-1}{n}\right)^{\alpha+1}\right].$$

(ii) If $\alpha = 1$, then by the lower formula of (61)

$$\int_0^{c_n} \tilde{\mathcal{H}}_1^+(e)\,\mathrm{d}e = \int_0^{c_n} (-e\ln e - (n-1-e)\ln(n-1-e) + (n-1-e)\ln(n-1))\,\mathrm{d}e$$

$$= \frac{1}{2}\left[-e^2(\ln e - \frac{1}{2}) + (n-1-e)^2(\ln(n-1-e) - \frac{1}{2}) - (n-1-e)^2\ln(n-1)\right]_0^{\frac{n-1}{n}}$$

$$= -\frac{1}{2}(\frac{n-1}{n})^2\big[\ln\frac{n-1}{n} - \frac{1}{2}\big] + \frac{1}{2}\frac{(n-1)^4}{n^2}\big[\ln\frac{(n-1)^2}{n} - \frac{1}{2}\big] - \frac{1}{2}\frac{(n-1)^4}{n^2}\ln(n-1)$$

$$+0 - \frac{(n-1)^2}{2}\big[\ln(n-1) - \frac{1}{2}\big] + \frac{(n-1)^2}{2}\ln(n-1)$$

$$= -\frac{1}{2}(\frac{n-1}{n})^2\ln(n-1) + \frac{1}{2}(\frac{n-1}{n})^2\ln n + \frac{1}{4}(\frac{n-1}{n})^2 + \frac{(n-1)^4}{n^2}\ln(n-1)$$

$$-\frac{1}{2}\frac{(n-1)^4}{n^2}\ln n - \frac{1}{4}\frac{(n-1)^4}{n^2} - \frac{1}{2}\frac{(n-1)^4}{n^2}\ln(n-1) + \frac{(n-1)^2}{4}$$

$$= \frac{1}{2}(\frac{n-1}{n})^2 n(n-2)\ln(n-1) - \frac{1}{2}(\frac{n-1}{n})^2 n(n-2)\ln n + \frac{1}{4}(\frac{n-1}{n})^2 2n$$

$$= \frac{(n-1)^2}{2n}\left[1 + (n-2)\ln\frac{n-1}{n}\right].$$

Lemma 5.2, formulas (83).

(i) If $0 < \alpha < 2, \alpha \neq 1$ then by the lower formula of (62) and by the definition of $\tilde{a}_{\alpha,n}$ in (59) we have that

$$\int_0^{c_n} \tilde{\mathcal{H}}_\alpha^-(e)\,\mathrm{d}e = \int_0^{c_n} \frac{1}{c_n}\,\tilde{a}_{\alpha,n}e\,\mathrm{d}e = \frac{(n-1)^2}{2n(\alpha-1)}\left[1 - \left(\frac{n-1}{n}\right)^{\alpha-1}\right].$$

(ii) If $\alpha = 1$ then by the lower formula of (62) and by the definition of $\tilde{a}_{\alpha,n}$ in (59) we have that

$$\int_0^{c_n} \tilde{\mathcal{H}}_\alpha^-(e)\,\mathrm{d}e = \int_0^{c_n} \frac{1}{c_n}\,\tilde{a}_{\alpha,n}e\,\mathrm{d}e = \frac{(n-1)^2}{2n}\ln\frac{n}{n-1}.$$

(iii) If $\alpha > 2$ then by the upper formula of (62)

$$\int_0^{c_n} \tilde{\mathcal{H}}_\alpha^-(e)\,\mathrm{d}e = \sum_{k=1}^{n-1}\int_{c_k}^{c_{k-1}} [\tilde{a}_{\alpha,k} + \tilde{b}_{\alpha,k}(e - c_k)]\,\mathrm{d}e = \sum_{k=1}^{n-1}\frac{\tilde{a}_{\alpha,k} + \tilde{a}_{\alpha,k+1}}{2k(k+1)}.$$

By (59)

$$\tilde{a}_{\alpha,k} = \frac{k-1}{\alpha-1}\big[1 - (\frac{k-1}{k})^{\alpha-1}\big]$$

so that

$$\tilde{a}_{\alpha,k} + \tilde{a}_{\alpha,k+1} = \frac{1}{\alpha-1}\left[2k - 1 - \frac{(k-1)^\alpha}{k^{\alpha-1}} - \frac{k^\alpha}{(k+1)^{\alpha-1}}\right]$$

and, consequently,

$$\int_0^{c_n} \tilde{\mathcal{H}}_\alpha^-(e)\,\mathrm{d}e = \frac{1}{2(\alpha-1)}\sum_{k=1}^{n-1}\frac{2k - 1 - (k-1)\big(\frac{k-1}{k}\big)^{\alpha-1} - k\big(\frac{k}{k+1}\big)^{\alpha-1}}{k(k+1)}.$$

APPENDIX B.

In this appendix we present tables with average and alternative average inaccuracies for selected α and n.

n	0.1	0.2	0.5	1	1.5	2	3	5	8
2	0.428	0.370	0.252	0.153	0.107	0.083	0.063	0.049	0.040
3	1.046	0.884	0.557	0.291	0.171	0.111	0.083	0.064	0.048
4	1.687	1.398	0.832	0.394	0.210	0.125	0.094	0.070	0.050
5	2.339	1.908	1.085	0.477	0.238	0.133	0.100	0.073	0.052
6	2.997	2.413	1.320	0.547	0.259	0.139	0.104	0.075	0.052
7	3.657	2.911	1.542	0.608	0.275	0.143	0.107	0.076	0.053
8	4.318	3.403	1.751	0.662	0.289	0.146	0.109	0.077	0.053
9	4.979	3.888	1.951	0.710	0.301	0.148	0.111	0.078	0.054
10	5.639	4.368	2.142	0.754	0.310	0.150	0.113	0.079	0.054
20	12.147	8.907	3.737	1.052	0.366	0.158	0.119	0.081	0.055
30	18.482	13.105	5.002	1.235	0.393	0.161	0.121	0.082	0.055
40	24.673	17.073	6.085	1.368	0.409	0.163	0.122	0.082	0.055
50	30.746	20.871	7.047	1.472	0.420	0.163	0.123	0.082	0.055
100	59.898	38.240	10.865	1.802	0.449	0.165	0.124	0.083	0.055
200	114.685	68.720	16.321	2.139	0.470	0.166	0.124	0.083	0.055
300	166.832	96.250	20.529	2.338	0.479	0.166	0.125	0.083	0.056
400	217.298	122.004	24.083	2.480	0.485	0.166	0.125	0.083	0.056
500	266.539	146.506	27.218	2.590	0.489	0.166	0.125	0.083	0.056
1000	501.137	257.689	39.535	2.934	0.499	0.167	0.125	0.083	0.056

Table B.1. Average inaccuracies $AI_n(H_\alpha|e_B)$ for selected α and n.

n	0.1	0.2	0.5	1	1.5	2	3	5	8
2	0.428	0.370	0.252	0.153	0.107	0.083	0.063	0.049	0.040
3	0.439	0.389	0.285	0.189	0.139	0.111	0.104	0.099	0.088
4	0.444	0.397	0.298	0.205	0.155	0.125	0.126	0.130	0.122
5	0.446	0.402	0.306	0.215	0.164	0.133	0.140	0.150	0.148
6	0.448	0.404	0.311	0.221	0.170	0.139	0.149	0.164	0.167
7	0.449	0.406	0.314	0.225	0.175	0.143	0.156	0.175	0.182
8	0.450	0.408	0.317	0.228	0.178	0.146	0.161	0.184	0.194
9	0.450	0.409	0.319	0.231	0.180	0.148	0.165	0.190	0.204
10	0.451	0.410	0.320	0.233	0.182	0.150	0.168	0.196	0.212
20	0.453	0.413	0.327	0.242	0.191	0.158	0.183	0.222	0.253
30	0.453	0.414	0.329	0.244	0.194	0.161	0.187	0.231	0.268
40	0.454	0.415	0.330	0.246	0.196	0.162	0.190	0.236	0.276
50	0.454	0.415	0.331	0.247	0.196	0.163	0.191	0.238	0.281
100	0.454	0.416	0.332	0.248	0.198	0.165	0.194	0.244	0.291
200	0.454	0.416	0.333	0.249	0.199	0.166	0.196	0.247	0.296
300	0.454	0.416	0.333	0.249	0.199	0.166	0.196	0.248	0.297
400	0.454	0.416	0.333	0.250	0.200	0.166	0.197	0.248	0.298
500	0.454	0.417	0.333	0.250	0.200	0.166	0.197	0.249	0.299
1000	0.455	0.417	0.333	0.250	0.200	0.167	0.197	0.249	0.300

Table B.2. Alternative average inaccuracies $AI_n(\tilde{H}_\alpha|e_B)$ for selected α and n.

n	0.1	0.2	0.5	1	1.5	2	3	5	8
2	0.222	0.200	0.152	0.111	0.091	0.083	0.083	0.106	0.142
3	0.372	0.335	0.254	0.176	0.134	0.111	0.125	0.172	0.222
4	0.459	0.413	0.312	0.213	0.157	0.125	0.150	0.210	0.264
5	0.517	0.465	0.351	0.237	0.172	0.133	0.167	0.234	0.289
6	0.560	0.504	0.380	0.255	0.182	0.139	0.179	0.250	0.306
7	0.592	0.533	0.401	0.268	0.190	0.143	0.188	0.262	0.317
8	0.619	0.557	0.419	0.279	0.196	0.146	0.194	0.271	0.326
9	0.640	0.576	0.433	0.287	0.200	0.148	0.200	0.278	0.333
10	0.658	0.592	0.446	0.295	0.204	0.150	0.205	0.283	0.339
20	0.751	0.678	0.511	0.334	0.224	0.158	0.226	0.308	0.364
30	0.790	0.714	0.540	0.351	0.232	0.161	0.234	0.317	0.372
40	0.812	0.735	0.557	0.362	0.237	0.163	0.238	0.321	0.376
50	0.826	0.748	0.569	0.369	0.240	0.163	0.240	0.323	0.379
100	0.859	0.780	0.598	0.387	0.247	0.165	0.245	0.328	0.384
200	0.880	0.801	0.618	0.402	0.252	0.166	0.248	0.331	0.386
300	0.888	0.809	0.627	0.409	0.254	0.166	0.248	0.332	0.387
400	0.892	0.813	0.632	0.413	0.255	0.166	0.249	0.332	0.388
500	0.895	0.816	0.636	0.416	0.256	0.166	0.249	0.332	0.388
1000	0.901	0.823	0.645	0.424	0.257	0.167	0.250	0.333	0.388

Table B.3. Average inaccuracies $AI_{n,\alpha}(e_B|H_\alpha)$ for selected α and n.

n	0.1	0.2	0.5	1	1.5	2	3	5	8
2	0.222	0.200	0.152	0.111	0.091	0.083	0.083	0.106	0.142
3	0.299	0.271	0.211	0.155	0.127	0.111	0.125	0.165	0.218
4	0.338	0.307	0.241	0.179	0.145	0.125	0.144	0.190	0.247
5	0.361	0.329	0.259	0.193	0.156	0.133	0.155	0.203	0.262
6	0.377	0.343	0.271	0.202	0.163	0.139	0.163	0.212	0.270
7	0.388	0.354	0.280	0.209	0.168	0.143	0.168	0.218	0.276
8	0.396	0.362	0.287	0.214	0.172	0.146	0.171	0.222	0.280
9	0.403	0.368	0.292	0.218	0.175	0.148	0.174	0.225	0.283
10	0.408	0.373	0.296	0.221	0.178	0.150	0.177	0.228	0.285
20	0.431	0.395	0.315	0.235	0.189	0.158	0.187	0.239	0.294
30	0.439	0.402	0.321	0.240	0.193	0.161	0.191	0.243	0.296
40	0.443	0.406	0.324	0.243	0.194	0.162	0.192	0.245	0.297
50	0.445	0.408	0.326	0.244	0.196	0.163	0.193	0.246	0.298
100	0.450	0.412	0.330	0.247	0.198	0.165	0.195	0.248	0.299
200	0.452	0.414	0.331	0.249	0.199	0.166	0.196	0.249	0.300
300	0.453	0.415	0.332	0.249	0.199	0.166	0.197	0.249	0.300
400	0.453	0.416	0.332	0.249	0.199	0.166	0.197	0.249	0.300
500	0.454	0.416	0.333	0.249	0.200	0.166	0.197	0.249	0.301
1000	0.454	0.416	0.333	0.250	0.200	0.167	0.197	0.250	0.301

Table B.4. Alternative average inaccuracies $AI_{n,\alpha}(e_B|\tilde{H}_\alpha)$ for selected α and n.

APPENDIX C.

In this appendix we present figures of power and adjoint power information bounds.

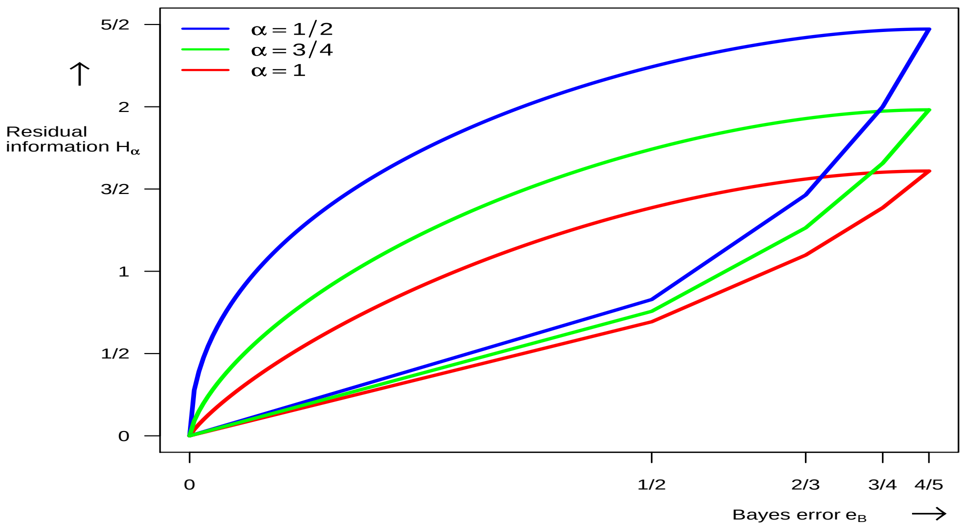

Fig. C.1. Power information bounds $\mathcal{H}_\alpha^{\pm}(e_B)$ as functions of Bayes error $0 \le e_B \le (n-1)/n$ for $n = 5$ and powers $\alpha \le 1$.

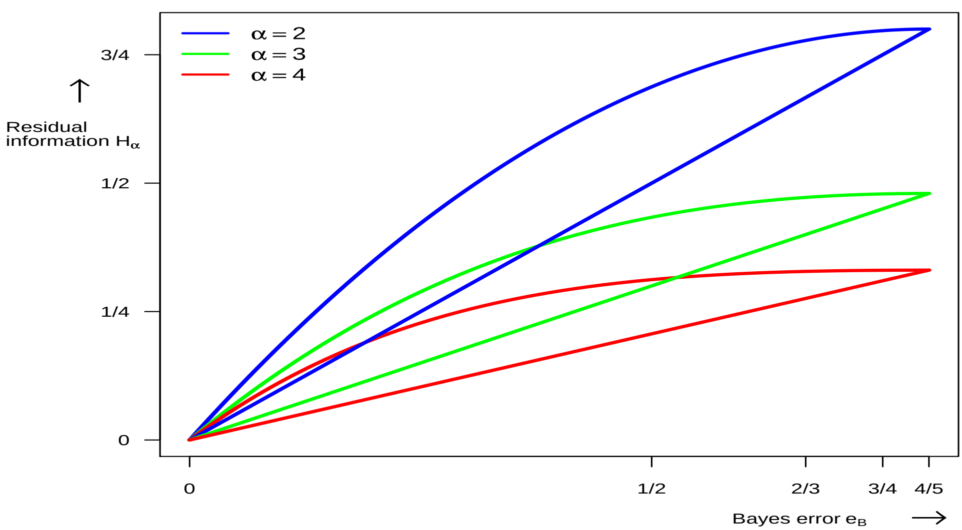

Fig. C.2. The same as in Figure C.1 for powers $\alpha > 1$.

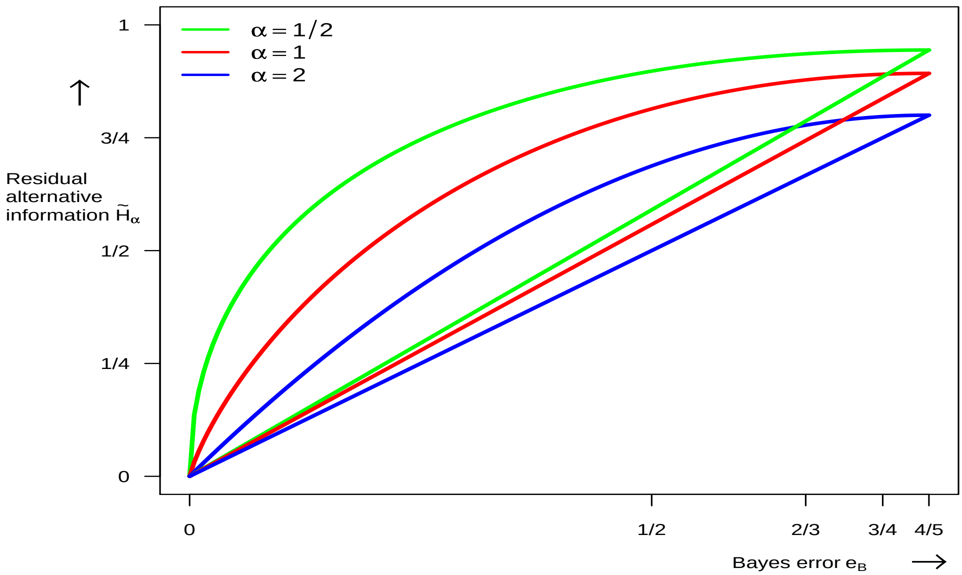

Fig. C.3. Adjoint power information bounds $\tilde{\mathcal{H}}_\alpha^{\pm}(e_B)$ as functions of Bayes error $0 \le e_B \le (n-1)/n$ for $n = 5$ and powers $\alpha \le 2$.

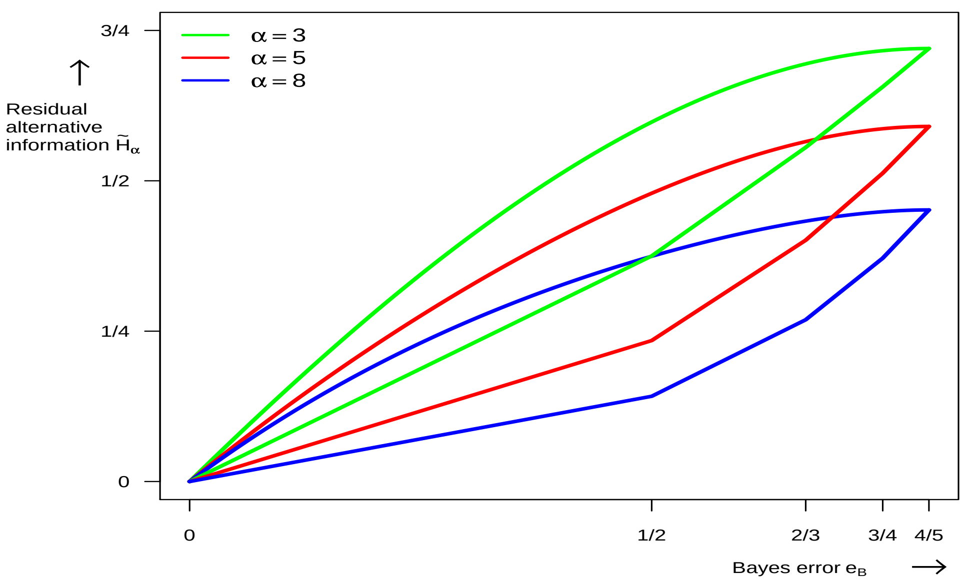

Fig. C.4. The same as in Figure C.3 for powers $\alpha > 2$.

APPENDIX D.

In this appendix we present figures of average and alternative average inaccuracies.

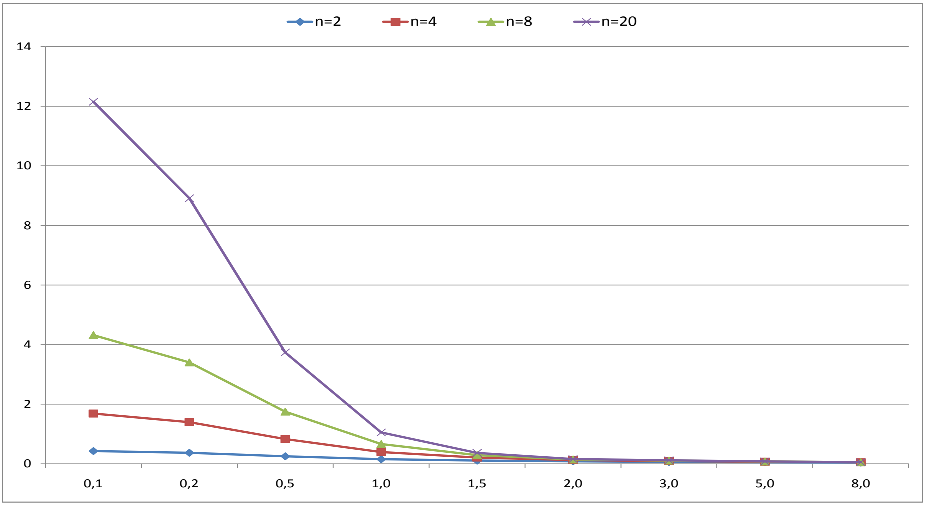

Fig. D.1. Average inaccuracies $AI_n(H_\alpha|e_B)$ for selected n as functions of α.

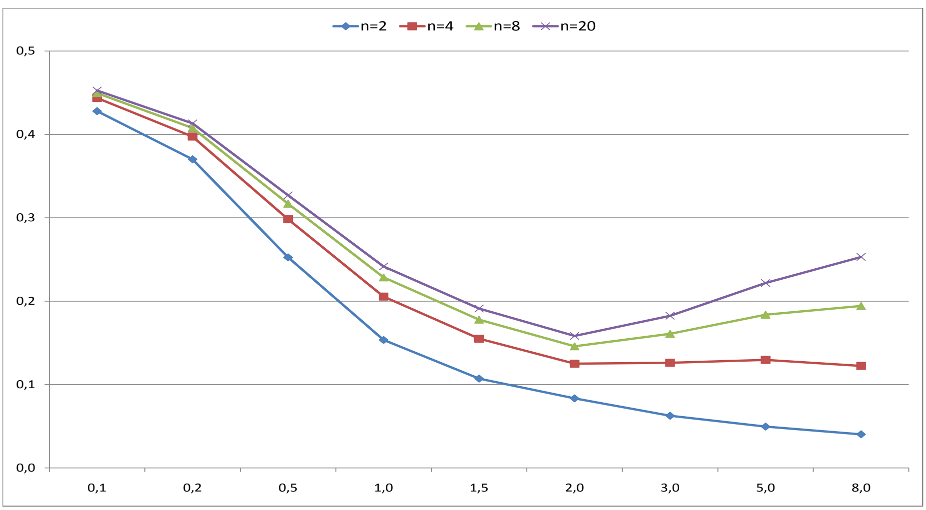

Fig. D.2. Alternative average inaccuracies $AI_n(\tilde{H}_\alpha|e_B)$ for selected n as functions of α.

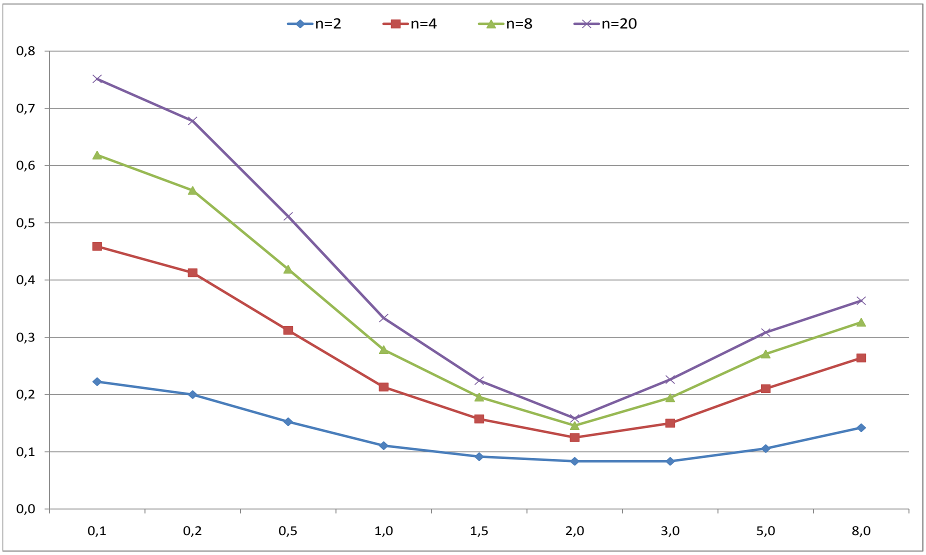

Fig. D.3. Average inaccuracies $AI_{n,\alpha}(e_B|H_\alpha)$ for selected n as functions of α.

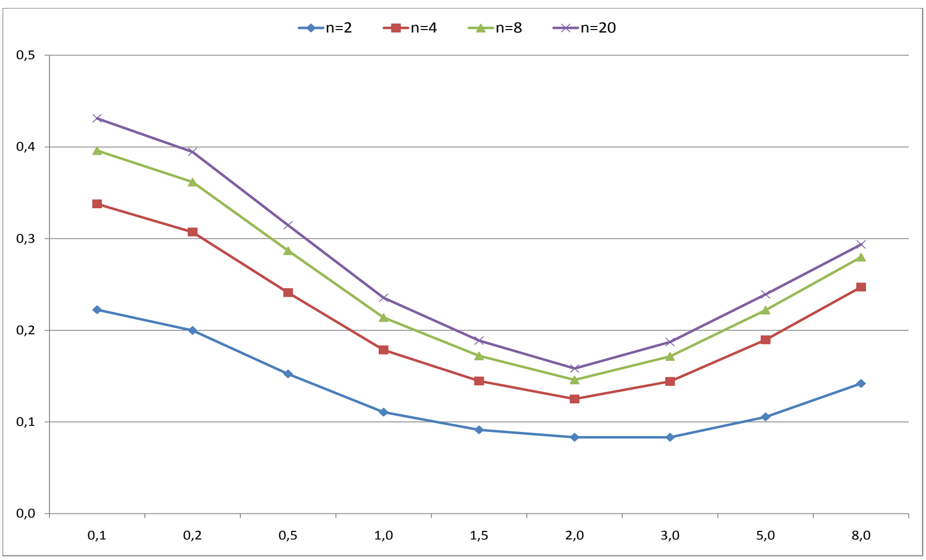

Fig. D.4. Alternative average inaccuracies $AI_{n,\alpha}(e_B|\tilde{H}_\alpha)$ for selected n as functions of α.

APPENDIX E.

In this appendix we present figures of Shannon and quadratic information bounds.

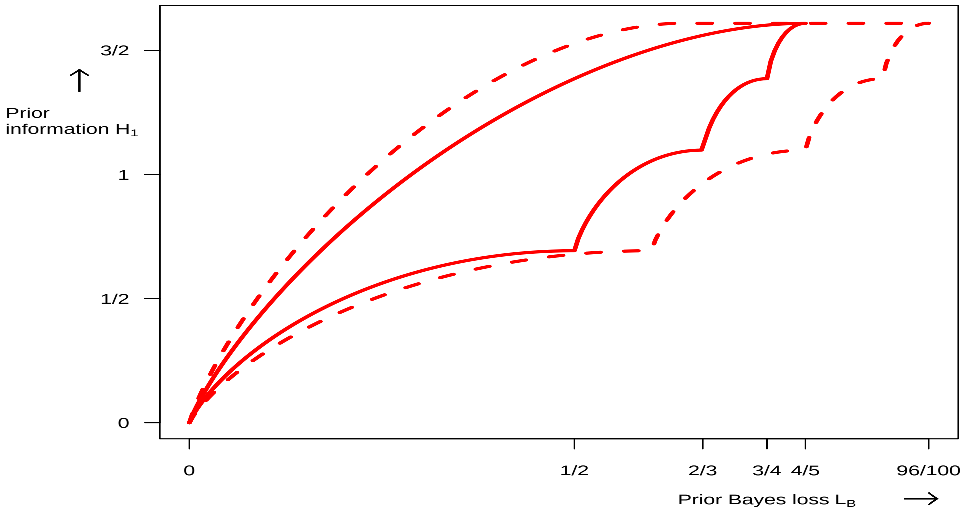

Fig. E.1. Shannon information bounds $H_1^{\pm}(L_B|\Lambda,\Delta)$ for the prior Bayes loss L_B under the loss function parameters $\Lambda = 1$ and $\Delta = 0$ (full line) or $\Delta = 2/5$ (interrupted line).

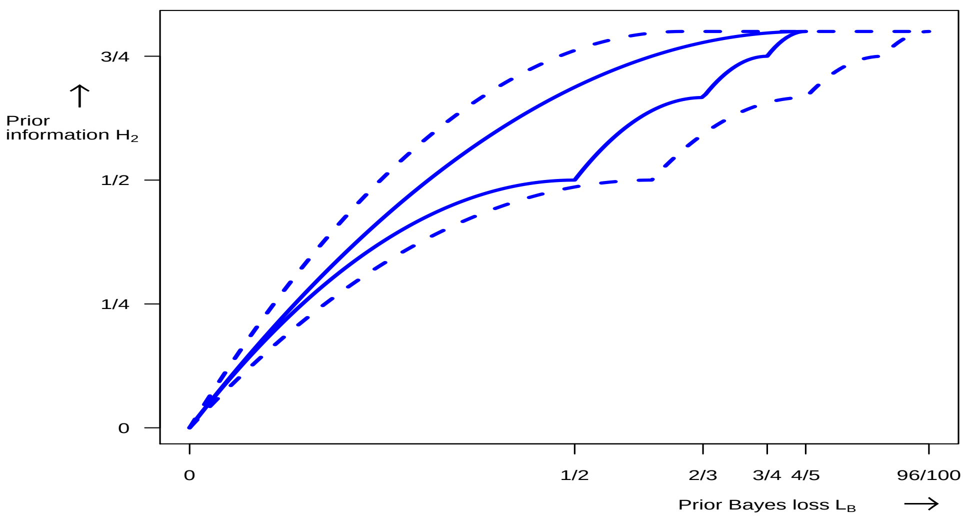

Fig. E.2. Quadratic information bounds $H_2^{\pm}(R_B|\Lambda,\Delta)$ for the prior Bayes loss L_B under the same conditions function as in Figure E.1.

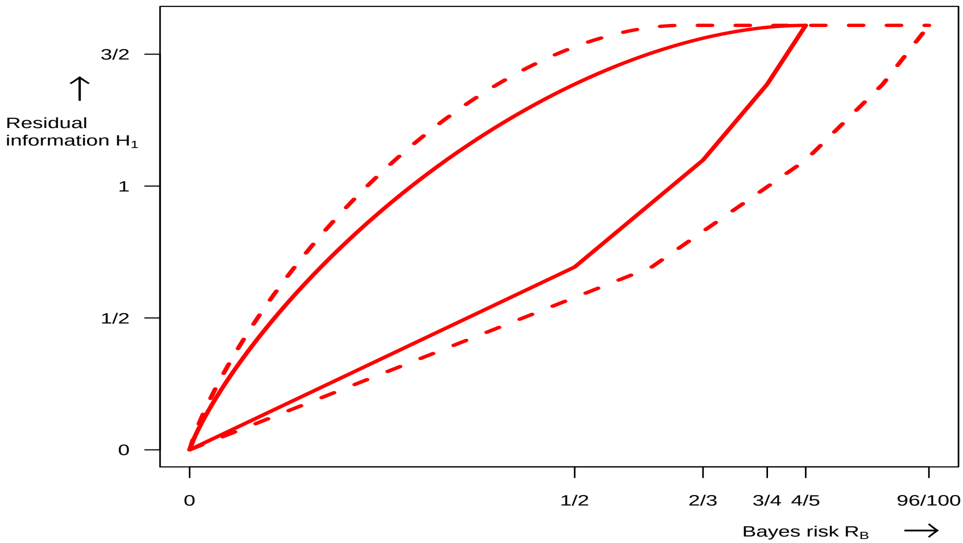

Fig. E.3. Shannon information bounds $\mathcal{H}_1^{\pm}(R_B|\Lambda, \Delta)$ for the Bayes risk R_B under the loss function parameters $\Lambda = 1$ and $\Delta = 0$ (full line) or $\Delta = 2/5$ (interrupted line).

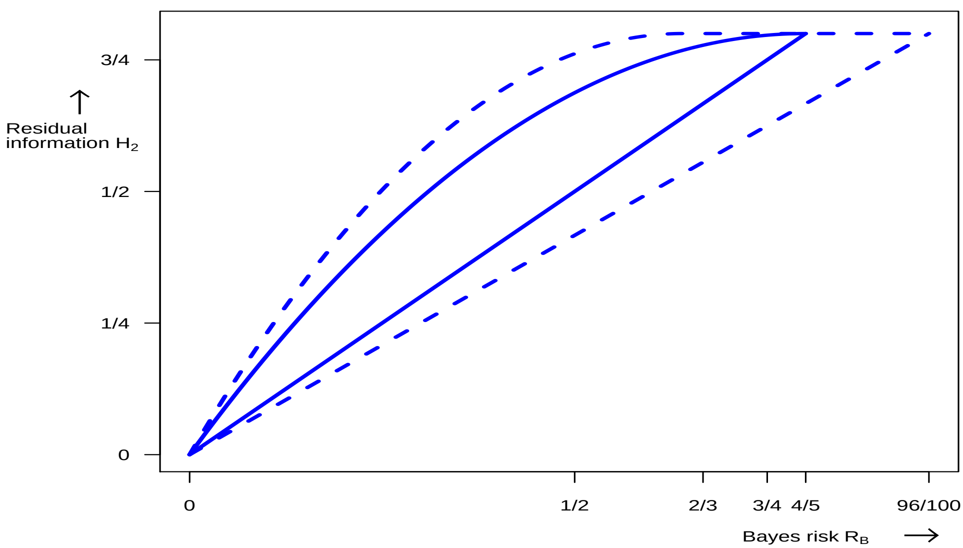

Fig. E.4. Quadratic information bounds $\mathcal{H}_2^{\pm}(R_B|\Lambda, \Delta)$ for the Bayes risk R_B under the same conditions as in Figure E.2.

ACKNOWLEDGEMENT

This work was partially supported by the grants MTM2009-09473 and GA ČR 102/07/1131.

(Received November 24, 2010)

REFERENCES

[1] M. Ben Bassat: f-entropies, probability of error, and feature selection. Inform. Control *39* (1978), 227–242.

[2] M. Ben Bassat and J. Raviv: Rényi's entropy and probability of error. IEEE Trans. Inform. Theory *24* (1978), 324–331.

[3] J. O. Berger: Statistical Decision Theory and Bayesian Analysis. Second edition. Springer, Berlin 1986.

[4] T. M. Cover and P. E. Hart: Nearest neighbor pattern classification. IEEE Trans. Inform. Theory *13* (1967), 21–27.

[5] P. Devijver and J. Kittler: Pattern Recognition. A Statistical Approach. Prentice Hall, Englewood Cliffs, New Jersey 1982.

[6] L. Devroye, L. Györfi, and G. Lugosi: A Probabilistic Theory of Pattern Recognition 1996. Springer, Berlin 1996.

[7] D. K. Faddeev: Zum Begriff der Entropie einer endlichen Wahrscheinlichkeitsschemas. Vol. I. Deutscher Verlag der Wissenschaften, Berlin 1957.

[8] M. Feder and N. Merhav: Relations between entropy and error probability. IEEE Trans. Inform. Theory *40* (1994), 259–266.

[9] P. Harremoës and F. Topsøe: Inequalities between entropy and index of coincidence derived from information diagrams. IEEE Trans. Inform. Theory *47* (2001), 2944–2960.

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

[11] L. Kanal: Patterns in pattern recognittion. IEEE Trans. Inform. Theory *20* (1974), 697–707.

[12] V. A. Kovalevsky: The problem of character recognition from the point of view of mathematical statistics. In: Reading Automata and Pattern Recognition (in Russian) (Naukova Dumka, Kyjev, ed. 1965). English translation in: Character Readers and Pattern Recognition, Spartan Books, New York 1968, pp. 3–30.

[13] D. Morales, L. Pardo, and I. Vajda: Uncertainty of discrete stochastic systems: general theory and statistical inference. IEEE Trans. System, Man and Cybernetics, Part A *26* (1996), 1–17.

[14] A. Rényi: Proceedings of 4th Berkeley Symp. on Probab. Statist. University of California Press, Berkeley, California 1961.

[15] N. P. Salikhov: Confirmation of a hypothesis of I. Vajda (in Russian). Problemy Peredachi Informatsii *10* (1974), 114–115.

[16] D. L. Tebbe and S. J. Dwyer III: Uncertainty and probability of error. IEEE Trans. Inform. Theory *14* (1968), 516–518.

[17] G. T. Toussaint: A generalization of Shannon's equivocation and the Fano bound. IEEE Trans. System, Man and Cybernetics *7* (1977), 300–302.

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

[19] I. Vajda: A contribution to informational analysis of patterns. In: Methodologies of Pattern Recognition (M. S. Watanabe, ed.), Academic Press, New York 1969.

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

[21] I. Vajda and J. Zvárová: On generalized entropies, Bayesian decisions and statistical diversity. Kybernetika *43* (2007), 675–696.

Domingo Morales, Operations Research Center, University Miguel Hernández de Elche, Elche, 03202. Spain.

e-mail: d.morales@umh.es

Igor Vajda, Institute of Information Theory and Automation – Academy of Sciences of the Czech Republic, Pod Vodárenskou věží 4, 182 08 Praha 8. Czech Republic.

e-mail: vajda@utia.cas.cz