Akademie věd České republiky
Ústav teorie informace a automatizace

Academy of Sciences of the Czech Republic
Institute of Information Theory and Automation

RESEARCH REPORT

D. Morales and I. Vajda:

Generalized Information Criteria for Optimal Bayes Decisions

No. 2274 January 2010

ÚTIA AV ČR, P. O. Box 18, 182 08 Prague, Czech Republic
Telex: 122018 atom c, Fax: (+42) (2) 688 4903
E-mail: utia@utia.cas.cz

This report constitutes an unrefereed manuscript which is intended to be submitted for publication. Any opinions and conclusions expressed in this report are those of the author(s) and do not necessarily represent the views of the Institute.

Generalized information criteria for optimal Bayes decisions*

D. Morales[1] and I. Vajda[2]

[1] Operations Research Center, University Miguel Hernández of Elche.

[2]Institute of Information Theory and Automation, Academy of Sciences of the Czech Republic, Prague.

January, 2010

Abstract

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 includes the Shannon entropy as a 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.

Key words: Shannon entropy, Alternative Shannon entropy, Power entropies, Alternative power entropies, Bayes error, Bayes risk, Sub-Bayes risk.

1. INTRODUCTION

In Morales, Pardo and Vajda (1996), we systematically studied *generalized measures of uncertainty* of stochastic systems with finite or countable state spaces Θ and probability distributions π on Θ, and *generalized measures of informativity* of random observations X with sample probability spaces $(\mathcal{X}, \mathcal{S}, P)$ and posterior distributions π_x on Θ when

* Supported by grants MTM2009-09473 and GAČR 102/07/1131

$X = x \in \mathcal{X}$. We investigated the general entropies $H(\pi)$ as appropriate concave or Schur concave functions of stochastic vectors π. As general *characteristics of informativity* of the whole stochastic observation experiment

$$\mathcal{E} = \langle(\Theta, \pi),\ (\mathcal{X}, \mathcal{S}, P)\rangle \tag{1.1}$$

we proposed the corresponding conditional entropies

$$H(\mathcal{E}) = \int_{\mathcal{X}} H(\pi_x) \mathrm{d}P(x) \tag{1.2}$$

closely related to the general information measures

$$I(\mathcal{E}) = H(\pi) - H(\mathcal{E}). \tag{1.3}$$

Particular attention was paid to the entropies of the form

$$H_\phi(\pi) = \sum_{\theta \in \Theta} \phi(\pi(\theta)) \tag{1.4}$$

for concave functions $\phi(t)$, $0 \leq t \leq 1$.

For $\phi(t) = -t \log t$ we obtain from (1.4) the classical Shannon entropy

$$H_1(\pi) = -\sum_{\theta \in \Theta} \pi(\theta) \ln \pi(\theta) \tag{1.5}$$

and from (1.2) and (1.3) the classical Shannon conditional entropy and Shannon information. For $\phi(t) = t(1-t)$ we obtain from (1.4) the alternative to the Shannon entropy

$$H_2(\pi) = 1 - \sum_{\theta \in \Theta} \pi^2(\theta) \tag{1.6}$$

called the *quadratic entropy* by Vajda (1968), and from (1.2) and (1.3) the corresponding quadratic conditional entropy $H_2(\mathcal{E})$ and quadratic information $I_2(\mathcal{E})$. In fact, Cover and Hart (1967) and Vajda (1968) introduced independently $H_2(\mathcal{E})$ as a measure of quality of decisions concerning the states $\theta \in \Theta$ achievable on the basis of observations X in the statistical experiments $\mathcal{E}$.

Vajda (1968) estimated the probability of error $P_e(\mathcal{E})$ of the Bayes decisions $\delta_B : \mathcal{X} \longmapsto \Theta$ by means of the quadratic entropy $H_2(\mathcal{E})$ as follows

$$\frac{H_2(\mathcal{E})}{1 + \sqrt{1 - H_2(\mathcal{E})}} \leq P_e(\mathcal{E}) \leq H_2(\mathcal{E}). \tag{1.7}$$

Obviously, the accuracy of this estimation increases with decreasing level of the entropy $H_2(\mathcal{E})$. This opens the possibility to replace the Bayesian characteristic $P_e(\mathcal{E})$ of decision situations $\mathcal{E}$ by the more smooth and computationally simpler information criterion $H_2(\mathcal{E})$

e.g. in feature selection procedures. The bounds (1.7) can be rewritten to the simpler equivalent form

$$P_e(\mathcal{E}) \leq H_2(\mathcal{E}) \leq P_e(\mathcal{E})\,(2 - P_e(\mathcal{E})) \equiv 1 - (1 - P_e(\mathcal{E}))^2 \tag{1.8}$$

and it was proved in Vajda (1968) that these bounds are attainable in the class of statistical experiments $\mathcal{E}$ with state spaces Θ of arbitrary sizes $|\Theta|$. For fixed sizes $|\Theta| = n$ the lower bound (1.7) was replaced by the more tight attainable bound

$$\frac{H_2(\mathcal{E})}{1 + \sqrt{1 - nH_2(\mathcal{E})/\,(n-1)}} \leq P_e(\mathcal{E}) \tag{1.9}$$

which is equivalent to

$$H_2(\mathcal{E}) \leq 1 - (1 - P_e(\mathcal{E}))^2 - \frac{P_e(\mathcal{E})^2}{n-1} \tag{1.10}$$

but the rigorous proof for $n \geq 3$ was given only later by Salichov (1974). If $n \to \infty$ then these bounds reduce to the previous bounds (1.7), (1.8).

The quadratic entropy (1.6) requires the operation of multiplication and summation, and is thus computationally simpler than the Shannon entropy (1.5) and also than the more general entropies of Rényi (1961)

$$HR_\alpha(\pi) = \frac{1}{\alpha - 1} \ln \sum_{\theta \in \Theta} \pi^\alpha(\theta), \quad \alpha > 0, \alpha \neq 1 \tag{1.11}$$

containing the Shannon entropy as the special limit case $H_1(\pi) = HR_1(\pi) \triangleq \lim_{\alpha \to 1} HR_\alpha(\pi)$. Rényi introduced the entropies axiomatically by extending and parameterizing by α the additivity rule in the axioms used earlier by Faddeev (1957) to characterize the Shannon's $H_1(\pi)$. However, he emphasized also the alternative "pragmatic approach" to motivate $H_1(\pi)$ and its extensions as characteristics of various statistical decision problems. In this sense for example Kovalevsky (1965) used $H_1(\mathcal{E})$ to obtain similar bounds as (1.8), (1.10) to characterize the error probability $P_e(\mathcal{E})$ in pattern recognition problems which inspired among other the work of Vajda (1968). The bounds of Kovalevsky were later reinvented and applied in different areas of statistical decisions and information processing by several authors, e.g. Tebbe and Dwyer (1968) or Feder and Merhav (1994).

By appropriately modifying the extended additivity rule of Rényi (1961), Havrda and Charvát (1967) axiomatically introduced the one-one modification

$$H_\alpha(\pi) = \frac{1}{\alpha - 1}\left(1 - \sum_{\theta \in \Theta} \pi^\alpha(\theta)\right), \quad \alpha > 0, \alpha \neq 1 \tag{1.12}$$

of the Rényi entropies with the limit $H_1(\pi) = \lim_{\alpha \to 1} H_\alpha(\pi)$. Vajda (1969) used the generalized informativity $H_\alpha(\mathcal{E})$ obtained by employing the general power entropy $H_\alpha(\pi)$ in (1.2) to evaluate bounds of the type (1.8), (1.10) and proposed the conditional power entropy $H_\alpha(\mathcal{E})$ as a generalized feature extraction criterion. This criterion was cited later by many authors, e.g. Kanal (1974), Devijver and Kittler (1982) or Devroye et al. (1996), and the bounds of the type (1.8), (1.10) were later completed, modified or tightened by Toussaint (1977), Ben Bassat (1978), Ben Bassat and Raviv (1978) and Harremoes and Topsoe (2001).

Vajda and Vašek (1985) found a method for obtaining attainable bounds of the type (1.8), (1.10) for arbitrary Schur concave entropy (1.2). These were applied later in Morales, Pardo and Vajda (1996) and Vajda and Zvárová (2007). Here we use the results of these papers to obtain some new attainable bounds for the probability of error $P_e(\mathcal{E})$ and to apply them in in estimation of the Bayes risks $R_B(\mathcal{E})$ in given experiments $\mathcal{E}$ with standard loss functions. The attention is focused on the accuracy of approximation of the Bayes probabilities of errors $P_e(\mathcal{E})$ and the related Bayes risks by information criteria of the common power type $H_\alpha(\mathcal{E})$. Perhaps the most interesting of the obtained results is the fact that the quadratic entropy $H_2(\mathcal{E})$ provides the most accurate estimate in the class of all power entropies $H_\alpha(\mathcal{E})$, $\alpha > 0$. Basic concepts and auxiliary 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 (1986)) 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

$$L : \Theta \times \Theta \mapsto [0, \infty) \quad \text{where} \quad \max_{\theta \in \Theta} L(\theta, \theta) = 0, \quad \min_{\hat{\theta} \in \Theta} \max_{\theta \in \Theta} L(\theta, \hat{\theta}) > 0. \tag{2.1}$$

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

$$\mathcal{E} = \langle \pi, \mathcal{P} = \{P_\theta : \theta \in \Theta\} \rangle \tag{2.2}$$

and a nontrivial loss function (2.1).

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
(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. *Expected loss* of an individual identification action $\hat{\theta} \in \Theta$ is

$$\mathcal{L}(\pi, \hat{\theta}) = \sum_{\theta \in \Theta} L(\theta, \hat{\theta})\, \pi(\theta). \tag{2.3}$$

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

$$\theta_\pi = \operatorname{argmin}_{\hat{\theta}} \mathcal{L}(\pi, \hat{\theta}) \tag{2.4}$$

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

$$L_B(\pi) = \mathcal{L}(\pi, \theta_\pi) \tag{2.5}$$

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

$$\delta = \mathcal{X} \mapsto \Theta. \tag{2.6}$$

Technically, they are assumed to be $\mathcal{S}$-measurable and P_θ-integrable for all $\theta \in \Theta$. *Risk function* of the identification rule (2.6) is

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

and its expected value

$$\mathcal{R}(\pi, \delta) = \sum_{\theta \in \Theta} R(\theta, \delta)\pi(\theta) = \sum_{\theta \in \Theta} \int_{\mathcal{X}} L(\theta, \delta(x))\pi(\theta)\, dP_\theta(x) \tag{2.7}$$

is simply a *risk*. The minimizer

$$\delta_B = \operatorname{argmin}_\delta \mathcal{R}(\pi, \delta) \tag{2.8}$$

is the *Bayes identification rule* and

$$R_B = R_B(\mathcal{E}, L) = \mathcal{R}(\pi, \delta_B) \tag{2.9}$$

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

It is known that in this model 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

$$P = \sum_{\theta \in \Theta} \pi(\theta) P_\theta \tag{2.10}$$

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

$$p_\theta(x) = \frac{dP_\theta(x)}{dP(x)}$$

defined for all data $x \in \mathcal{X}$, with values uniquely given except possibly 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

$$\pi_x = (\pi_x(\theta) \stackrel{\triangle}{=} \pi(\theta) p_\theta(x) : \theta \in \Theta) \tag{2.11}$$

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

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

Obviously, the statistical experiment (2.2) is equivalently described by the conditional distributions (2.11) for $x \in \mathcal{X}$ and the marginal distribution (2.10),

$$\mathcal{E} = \langle \pi, \mathcal{P} = \{P_\theta : \theta \in \Theta\} \rangle \equiv \langle P, \Pi = \{\pi_x : x \in \mathcal{X}\} \rangle. \tag{2.12}$$

Using the posterior distribution (2.11) and the concept of expected loss (2.3), we can rewrite the risk formula (2.7) into the simple form

$$\mathcal{R}(\pi_x, \delta) = \int_{\mathcal{X}} \mathcal{L}(\pi_x, \delta(x))\, dP(x). \tag{2.13}$$

From here and from (2.8) 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 (2.4). Thus the Bayes identification rule can equivalently be defined P-a.e. on $\mathcal{X}$ by the formula

$$\delta_B(x) = \theta_{\pi_x} \equiv \text{argmin}_{\hat{\theta}} \sum_{\theta \in \Theta} L(\theta, \hat{\theta}) \pi_x(\theta). \tag{2.14}$$

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 (2.5) with the prior distribution π replaced by the posterior distribution π_x. In other words, we deduce that

$$\begin{aligned} R_B = \mathcal{R}(\pi, \delta_B) &= \int_{\mathcal{X}} \mathcal{L}(\pi_x, \theta_{\pi_x})\, dP(x) \qquad (\text{cf. } (2.13), (2.14)) \\ &= \int_{\mathcal{X}} L_B(\pi_x)\, dP(x). \end{aligned} \tag{2.15}$$

3. RELATIONS TO ZERO-ONE LOSS MODEL

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

$$L_e : \Theta \times \Theta \mapsto \{0,1\}, \qquad L_e(\theta,\hat{\theta}) = \begin{cases} 1 & \text{if } \hat{\theta} \neq \theta, \\ 0 & \text{if } \hat{\theta} = \theta. \end{cases} \tag{3.1}$$

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

$$\mathcal{L}_e(\pi,\hat{\theta}) = \sum_{\theta\in\Theta} L_e(\theta,\hat{\theta})\pi(\theta) = 1 - \pi(\hat{\theta}) \tag{3.2}$$

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 (2.5) is the minimal prior probability of error given by the formula

$$e_B(\pi) = 1 - \pi(\theta_\pi), \tag{3.3}$$

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

$$e_B(\pi_x) = 1 - \pi_x(\theta_{\pi_x}) \tag{3.4}$$

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 (2.15) and the equality $L_B(\pi_x) = e_B(\pi_x)$, the Bayes risk $R_B = R_B(\mathcal{E}, L)$ of (2.9) 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

$$e_B = e_B(\mathcal{E}) = \int_{\mathcal{X}} e_B(\pi_x)\, dP(x). \tag{3.5}$$

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 (2.1) the proper losses are positive between

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

and

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

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

$$\Lambda = \frac{L^+ + L^-}{2} \quad \text{and} \quad \Delta = (L^+ - \Lambda). \tag{3.6}$$

Obviously, $\Delta = 0$ if and only if $L(\theta,\hat\theta) = \Lambda\, L_e(\theta,\hat\theta)$ and the model has zero-one losses if and only if

$$(\Delta, \Lambda) = (0, 1).$$

Example 3.1. Let the state space $\Theta = \{1, \ldots, n\}$ represents classification of satellite ship images and let the loss function (2.1) 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

$$L^- = 4/5, \quad L^+ = 6/5 \quad \text{and} \quad (\Delta, \Lambda) = (2/5, 1)\,.$$

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

$$|L_B(\pi_x) - e_B(\pi_x)\,\Lambda| \le e_B(\pi_x)\,\Delta/2, \tag{3.7}$$

(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 minmax assumption in (2.1) 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 (3.6), $L(\theta, \hat{\theta}) > 0$ implies $L(\theta, \hat{\theta}) \in [L^-, L^+]$ where either $L(\theta, \hat{\theta}) \in [\Lambda, L^+]$ in which case

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

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

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

Hence

$$|L(\theta, \hat{\theta}) - \Lambda| \leq \Delta/2 \quad \text{for all} \quad \theta, \hat{\theta} \in \Theta \ \text{ with } \ L(\theta, \hat{\theta}) > 0. \tag{3.8}$$

Further, by (2.3) and (2.5),

$$\mathcal{L}(\pi, \hat{\theta}) = \sum_{\theta \neq \hat{\theta}} L(\theta, \hat{\theta})\pi(\theta) \ \text{ and } \ L_B(\pi) = \sum_{\theta \neq \theta_\pi} L(\theta, \theta_\pi)\pi(\theta). \tag{3.9}$$

Therefore multiplying the left side of (3.8) 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 (3.9) 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 (3.7) over $\mathcal{X}$ with respect to the measure P and using once more the Jensen inequality, we get

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

The desired result of (iii) follows from here and from the formulas (2.15) and (3.5). ■

Denote for a while by δ_e the Bayes identifier in the simpler error loss model, to distinguish it from the Bayes identifier δ_B in the general loss model of the previous section. By definition, $\delta_e(x)$ maximizes the posterior probability $\pi_x(\theta)$ on Θ under observation $x \in \mathcal{X}$. Therefore $L(\delta_e(x), \hat{\theta})$ is the lowest loss among all losses $L(\theta, \hat{\theta})$ resulting from the decision $\hat{\theta}$. If we replace in the definition of the Bayes identification $\hat{\theta} = \delta_B(x)$ the posteriori expected loss

$$\mathcal{L}(\pi_x, \hat{\theta}) = \sum_{\theta \in \Theta} L(\theta, \hat{\theta})\pi_x(\theta) \quad \text{(c.f. (2.14) and (2.4))}$$

by the a posteriori most probable loss $L(\delta_e(x), \hat{\theta})$ then the corresponding identifier

$$\delta_{SB}(x) = \text{argmin}_{\hat{\theta}} L(\delta_e(x), \hat{\theta}) \tag{3.10}$$

is an interesting alternative to the Bayes identifier $\delta_B(x)$. We call it a *sub-Bayes identifier.* It is simpler than $\delta_B(x)$ since it minimizes one particular loss profile $L(\delta_e(x), \hat{\theta})$ while $\delta_B(x)$ minimizes the mixture

$$\sum_{\theta \in \Theta} L(\theta, \hat{\theta}) \pi_x(\theta) \qquad \text{(cf. (2.14)}$$

of all loss profiles. It seems to be a suitable alternative to the Bayes δ_B when fast on-line decisions $\delta : \mathcal{X} \to \Theta$ are needed in the situations with fixed experiment $\mathcal{E}$ and loss function $L(\theta, \hat{\theta})$ fluctuating out of the control of the decision maker.

The following Theorem 3.2 deals with the *sub-Bayes risk*

$$R_{SB} = \mathcal{R}(\pi, \delta_{SB}). \tag{3.11}$$

Theorem 3.2. Consider the general loss model of Section 3 with median loss $\Lambda > 0$ and loss dispersion $\Delta \geq 0$.

(i) For P-almost all $x \in \mathcal{X}$ the posterior Bayes loss $L_B(\pi_x) = \mathcal{L}(\pi_x, \delta_B(x))$ and the posterior sub-Bayes loss $\mathcal{L}(\pi_x, \delta_{SB}(x))$ satisfy the relation

$$0 \leq \mathcal{L}(\pi_x, \delta_{SB}(x)) - \mathcal{L}(\pi_x, \delta_B(x)) \leq e_B(\pi_x)\, \Delta,$$

where $e_B(\pi_x)$ is the posterior Bayes error (3.4).

(ii) The Bayes risk R_B and the sub-Bayes risk R_{SB} satisfy the relation

$$0 \leq R_{SB} - R_B \leq e_B\, \Delta,$$

where e_B is the Bayes error (3.5).

Proof. (I) Since for P-almost all $x \in \mathcal{X}$

$$\delta_B(x) = \text{argmin}_{\hat{\theta}} \mathcal{L}(\pi_x, \hat{\theta}),$$

the left inequality in (i) is clear. By (2.3)

$$\mathcal{L}(\pi_x, \delta_{SB}(x)) = L(\delta_e(x), \delta_{SB}(x))\, \pi_x(\delta_e(x)) + M(x)$$

for

$$M(x) = \sum_{\theta \neq \delta_e(x)} L(\theta, \delta_{SB}(x)) \leq (\Lambda + \Delta/2)\, [1 - \pi_x(\delta_e(x))].$$

Similarly,

$$\mathcal{L}(\pi_x, \delta_B(x)) = L(\delta_e(x), \delta_B(x))\, \pi_x(\delta_e(x)) + N(x) \leq L(\delta_e(x), \delta_{SB}(x))\, \pi_x(\delta_e(x)) + N(x)$$

for

$$N(x) = \sum_{\theta \neq \delta_e(x)} L(\theta, \delta_B(x)) \geq (\Lambda - \Delta/2)\, [1 - \pi_x(\delta_e(x))].$$

Therefore

$$\mathcal{L}(\pi_x, \delta_{SB}) - \mathcal{L}(\pi_x, \delta_B) \leq [1 - \pi_x(\delta_e(x))]\,\Delta$$

and (i) follows from (3.4) where θ_{π_x} is nothing but the Bayes identifier $\delta_e(x)$.

(II) By (3.3)

$$R_B = \int_{\mathcal{X}} \mathcal{L}(\pi_x, \delta_B(x))\, dP(x)$$

and by (3.11) and (2.7)

$$R_{SB} = \int_{\mathcal{X}} \mathcal{L}(\pi_x, \delta_{SB}(x))\, dP(x).$$

Thus (ii) obviously follows from the already proved inequality in (i) and from the formula (3.5) for the Bayes error e_B. ■

4. GENERALIZED INFORMATION CRITERIA

In this section and in the rest of the paper we denote by $n = |\Theta|$ 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 information criteria $H(\pi)$, $H(\pi_x)$ and

$$H = H(\mathcal{E}) = \int_{\mathcal{X}} H(\pi_x)\, dP(x)$$

measuring the uncertainties (entropies) of realizations of states of nature θ from individual stochastic sources (Θ, π), (Θ, π_x), or from systems of such sources $\mathcal{E} = \{(\Theta, \pi_x) : x \in \mathcal{X}\}$ where x are data (realizations of random observations X with the sample space $(\mathcal{X}, \mathcal{S}, P)$) For details about these concepts and notations see sections 2 and 3.

Classical Shannon information criteria are based on the *Shannon entropy* (here measured in *nats* instead of *bits*)

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

In Section 1 we mentioned their generalizations based on the *power entropies*

$$H_\alpha(\pi) = \sum_{\theta \in \Theta} \phi_\alpha(\pi(\theta)), \qquad \alpha > 0 \tag{4.1}$$

where for $\alpha \neq 1$

$$\phi_\alpha(t) = \begin{cases} \frac{1}{\alpha - 1}\,[t(1 - t^{\alpha - 1})] & \text{if} \quad \alpha \neq 1 \\ \lim_{\alpha \to 1} \phi_\alpha(t) = -t \ln t & \text{if} \quad \alpha = 1. \end{cases} \tag{4.2}$$

Hence

$$H_\alpha(\pi) = \begin{cases} \frac{1}{\alpha - 1}\,[1 - \sum_{\theta \in \Theta} \pi(\theta)^\alpha] & \text{if} \quad \alpha \neq 1 \\ \lim_{\alpha \to 1} H_\alpha(\pi) = -\sum_{\theta \in \Theta} \pi(\theta) \ln \pi(\theta) & \text{if} \quad \alpha = 1. \end{cases} \tag{4.3}$$

As argued in Morales, Pardo and Vajda (1996), the desired information-theoretic properties of the power entropies follow from the concavity of functions $\phi_\alpha(t)$ on $[0,1]$ and from their extremal values $\phi_\alpha(0) = \phi_\alpha(1) = 0$. As an example we can take the *information processing property*

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

where $T : \Theta \mapsto \mathcal{T}$ is a mapping which leads to the new distribution

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

on the new states $\tau \in \mathcal{T}$ and as such represents an information processing on the state space. The remaining symbols π_D, π_U stand for the Dirac and uniform probability distributions on Θ. The concavity argument applies also to the *alternative power functions* $\tilde{\phi}_\alpha(t) = \phi_\alpha(1-t)$ so that the same information-theoretic properties are shared by the corresponding *alternative power entropies*

$$\tilde{H}_\alpha(\pi) = \sum_{\theta \in \Theta} \tilde{\phi}_\alpha(\pi(\theta)), \quad \alpha > 0, \tag{4.4}$$

i.e.

$$\tilde{H}_\alpha(\pi) = \begin{cases} \frac{1}{\alpha - 1}\left[n_\pi - 1 - \sum_{\theta \in \Theta}(1 - \pi(\theta))^\alpha\right] & \text{if} \quad \alpha \neq 1 \\ \\ \lim_{\alpha \to 1} \tilde{H}_\alpha(\pi) = -\sum_{\theta \in \Theta}(1 - \pi(\theta)) \ln(1 - \pi(\theta)) & \text{if} \quad \alpha = 1 \end{cases} \tag{4.5}$$

where n_π denotes the number of states in Θ supporting the prior distribution π,

$$n_\pi = \#\{\theta \in \Theta : \pi(\theta) > 0\}.$$

Similarly as the classical Shannon entropy, the generalized entropies $H_\alpha(\pi)$ and $\tilde{H}_\alpha(\pi)$ are measures of the information obtained by observing the state from Θ a priori distributed by π. One can thus expect that the minimal error probability $e_B(\pi)$ of identification of this state on the basis of π is intimately related to these entropies. Since the Bayes error $e_B = e_B(\mathcal{E})$ in the general experiment $\mathcal{E}$ (c.f. (2.12)) is the average minimal error probability

$$e_B(\mathcal{E}) = \int_{\mathcal{X}} e_B(\pi_x) \mathrm{d}P(x) \quad \text{(c.f. (3.5))}, \tag{4.6}$$

it must be similarly related to the average generalized entropies $H_\alpha(\mathcal{E})$ and $\tilde{H}_\alpha(\mathcal{E})$ defined as analogous stochastic mixtures

$$H_\alpha(\mathcal{E}) = \int_{\mathcal{X}} H_\alpha(\pi_x) \mathrm{d}P(x) \quad \text{and} \quad \tilde{H}_\alpha(\mathcal{E}) = \int_{\mathcal{X}} \tilde{H}_\alpha(\pi_x) \mathrm{d}P(x). \tag{4.7}$$

In what follows we investigate this relation.

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

$$\mathcal{H}^+_\alpha(e_B) = \max_{e_B(\mathcal{E})=e_B} H_\alpha(\mathcal{E}) \quad \text{and} \quad \mathcal{H}^-_\alpha(e_B) = \min_{e_B(\mathcal{E})=e_B} H_\alpha(\mathcal{E}), \tag{4.8}$$

by means of the auxiliary function

$$h(t) = -t\ln t - (1-t)\ln(1-t), \quad 0 \le t \le 1 \;\text{ where }\; 0\ln 0 = 0 \tag{4.9}$$

and the auxiliary constants

$$a_{\alpha,k} = \begin{cases} \frac{1-k^{1-\alpha}}{\alpha-1} & \text{if } \alpha \ne 1 \\ \lim_{\alpha\to 1} a_{\alpha,k} = \ln k & \text{if } \alpha = 1 \end{cases} \quad \text{and} \quad c_k = \frac{k-1}{k}, \quad 1 \le k \le n \tag{4.10}$$

as well as

$$b_{\alpha,k} = \frac{a_{\alpha,k+1} - a_{\alpha,k}}{c_{k+1} - c_k}, \quad 1 \le k \le n-1. \tag{4.11}$$

In (4.8) 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

$$0 \le e(\pi),\; e_B \le c_n. \tag{4.12}$$

In the proof of the next theorem are used the formulas

$$H^+_\alpha(e) = \frac{1-(n-1)^{1-\alpha}e^\alpha - (1-e)^\alpha}{\alpha-1}, \quad 0 \le e \le c_n \tag{4.13}$$

$$H^-_\alpha(e) = \frac{1-[1-k(1-e)]^\alpha - k(1-e)^\alpha}{\alpha-1}, \quad c_k \le e \le c_{k+1}, \quad 1 \le k \le n-1 \tag{4.14}$$

and their limits

$$H^+_1(e) = h(e) + e\ln(n-1), \quad 0 \le e \le c_n \tag{4.15}$$

$$H^-_1(e) = h(k(1-e)) + k(1-e)\ln k, \quad c_k \le e \le c_{k+1}, \quad 1 \le k \le n-1 \tag{4.16}$$

for the attainable upper and lower power entropy bounds

$$H^+_\alpha(e) = \max_{e(\pi)=e} H_\alpha(\pi) \quad \text{and} \quad H^-_\alpha(e) = \min_{e(\pi)=e} H_\alpha(\pi) \tag{4.17}$$

(for details about these bounds see Theorem 2 in Morales et al. (1996)).

Theorem 4.1. The power entropy bounds (4.8) are for every $0 \le e_B \le c_n$ explicitly given by

$$\mathcal{H}_\alpha^+(e_B) = \begin{cases} H_\alpha^+(e_B) = \frac{1}{\alpha-1}\left[1-(n-1)^{1-\alpha}e_B^\alpha - (1-e_B)^\alpha\right] & \text{if } \alpha \neq 1 \\ \\ H_1^+(e_B) = h(e_B) + e_B \ln(n-1) & \text{if } \alpha = 1 \end{cases} \tag{4.18}$$

(cf. (4.13), (4.16)) and

$$\mathcal{H}_\alpha^-(e_B) = \begin{cases} a_{\alpha,k} + b_{\alpha,k}(e_B - c_k) & \text{if } c_k \le e_B \le c_{k+1},\ 1 \le k \le n-1,\ 0 < \alpha < 2 \\ a_{\alpha,n} e_B / c_n, & \text{if } \alpha \ge 2. \end{cases} \tag{4.19}$$

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

$$\mathcal{H}_\alpha^+(0) = \mathcal{H}_\alpha^-(0) = 0 \quad \text{and } \mathcal{H}_\alpha^+(c_n) = \mathcal{H}_\alpha^-(c_n) = a_{\alpha,n} > 0. \tag{4.20}$$

Proof. 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 above by the function $\mathcal{H}_\alpha^+(tc + (1-t)\tilde{c})$ and bounded 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 (4.7), (4.8) and (4.17) 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 (4.18) follow from (4.13) and (4.15). On the other hand, H_α^- given by (4.14) and (4.16) 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

$$\Phi_\alpha(t) = a_{\alpha,k} + b_{\alpha,k}(t - c_k) \quad \text{for} \quad c_k \le t \le c_{k+1}, \quad 1 \le k \le n-1. \tag{4.21}$$

This function is convex (concave) if the sequence

$$\frac{\Phi_\alpha(c_k)}{c_k} = \frac{a_{\alpha,k}}{c_k} = \begin{cases} \frac{k\left(1-k^{1-\alpha}\right)}{(\alpha-1)(k-1)} & \text{if } \alpha \neq 1 \\ \\ \lim_{\alpha\to1} a_{\alpha,k} = \frac{k}{k-1}\ln k & \text{if } \alpha = 1 \end{cases}$$

is increasing (decreasing or constant) 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 $0 < \alpha < 2$ 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 $\alpha \ge 2$. This proves (4.19). The last assertion including relations (4.20) is clear from what has already been proved. ■

In Figures 4.1 and 4.2 are drawn 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 that the lower bounds $\mathcal{H}_\alpha^-(e_B)$ for $\alpha \ge 2$ are linear in the variable e_B.

Remark 4.1. Relation (4.15) is the well known Fano bound of information theory and (4.13) is its extension obtained previously in Vajda (1968) for $\alpha = 2$ and in Morales et al. (1996) and other references mentioned there for remaining $\alpha > 0$.

Remark 4.2. It is easy to verify that all power entropy bounds (4.13) - (4.19) 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

$$e_\alpha^{\mp}(H) = \max_{H_\alpha^{\pm}(e) \le H} e \quad \text{and} \quad e_{B,\alpha}^{\mp}(H) = \max_{H_\alpha^{\pm}(e_B) \le H} e_B \tag{4.22}$$

(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

$$e_\alpha^{-}(H) < e_\alpha^{+}(H) \ \text{ and } \ e_{B,\alpha}^{-}(H) < e_{B,\alpha}^{+}(H) \tag{4.23}$$

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 bounds

$$\tilde{\mathcal{H}}_\alpha^{+}(e_B) = \max_{e_B(\mathcal{E}) = e_B} \tilde{H}_\alpha(\mathcal{E}) \quad \text{and} \quad \tilde{\mathcal{H}}_\alpha^{-}(e_B) = \min_{e_B(\mathcal{E}) = e_B} \tilde{H}_\alpha(\mathcal{E}). \tag{4.24}$$

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

$$\tilde{a}_{\alpha,k} = \begin{cases} \frac{k-1}{\alpha-1}\left[1 - \left(\frac{k-1}{k}\right)^{\alpha-1}\right] & \text{if } \ \alpha \ne 1 \\ \lim_{\alpha\to 1} \tilde{a}_{\alpha,k} = (1-k)\ln\frac{k-1}{k} & \text{if } \ \alpha = 1 \end{cases} \quad \text{for } 1 \le k \le n \tag{4.25}$$

and

$$\tilde{b}_{\alpha,k} = \frac{\tilde{a}_{\alpha,k+1} - \tilde{a}_{\alpha,k}}{c_{k+1} - c_k}, \quad 1 \le k \le n-1 \tag{4.26}$$

where $0 \ln 0 = 0$ in (4.25).

Theorem 4.2. Let $\alpha > 0$ be arbitrary fixed. The alternative power entropy bounds (4.24) are for every $0 \le e_B \le c_n$ explicitly given by

$$\tilde{\mathcal{H}}_\alpha^{+}(e_B) = \begin{cases} \frac{1}{\alpha-1}\left[n - 1 - e_B^{\alpha} - (n-1)\left(1 - \frac{e_B}{n-1}\right)^{\alpha}\right] & \text{if } \alpha \ne 1 \\ \lim_{\alpha\to 1} \tilde{\mathcal{H}}_\alpha^{+}(e_B) = -e\ln e - (n-1-e)\ln\left(\frac{n-1-e}{n-1}\right) & \text{if } \alpha = 1 \end{cases} \tag{4.27}$$

and

$$\tilde{\mathcal{H}}_\alpha^-(e_B) = \begin{cases} \tilde{a}_{\alpha,k} + \tilde{b}_{\alpha,k}(e_B - c_k) & \text{if } c_k < e_B < c_{k+1}, 1 \le k \le n-1, \alpha > 2 \\ \\ \tilde{a}_{\alpha,n} e_B / c_n & \text{if } 0 < \alpha \le 2. \end{cases} \tag{4.28}$$

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

$$\tilde{\mathcal{H}}_\alpha^+(0) = \tilde{\mathcal{H}}_\alpha^-(0) = 0 \quad \text{and } \tilde{\mathcal{H}}_\alpha^+(c_n) = \tilde{\mathcal{H}}_\alpha^-(c_n) = \tilde{a}_{\alpha,n} > 0. \tag{4.29}$$

Proof. (I) By Theorem 1 in Vajda and Vašek (1985), for every $0 \le e \le c_n$

$$e(\pi) = e \quad \text{implies} \quad \tilde{H}_\alpha^-(e) \le \tilde{H}_\alpha(\pi) \le \tilde{H}_\alpha^+(e) \tag{4.30}$$

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

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

and

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

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

$$\tilde{H}_\alpha^+(e) = \tilde{H}_\alpha(\pi^+) = \frac{1}{\alpha-1}\left[n-1-e^\alpha-(n-1)\left(1-\frac{e}{n-1}\right)^\alpha\right] \tag{4.31}$$

and

$$\tilde{H}_\alpha^-(e) = \tilde{H}_\alpha(\pi^-) = \frac{k-ke^\alpha-k^\alpha(1-e)^\alpha}{\alpha-1} \tag{4.32}$$

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

$$\tilde{H}_1^+(e) = \tilde{H}_1(\pi^+) = \lim_{\alpha\to 1} \tilde{H}_\alpha^+(e) = -e\ln e - (n-1-e)\ln\left(\frac{n-1-e}{n-1}\right) \tag{4.33}$$

and

$$\tilde{H}_1^-(e) = \tilde{H}_1^-(\pi^-) = \lim_{\alpha\to 1} \tilde{H}_\alpha^-(e) = -ke - k(1-e)\ln[k(1-e)] \tag{4.34}$$

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 (4.27) follows from (4.31) and (4.33). On the other hand, $\tilde{H}^-_\alpha$ given by (4.32) and (4.34) 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 (4.28). The last assertion including the equations (4.29) follow from what was already proved above. ■

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

In Figures 4.3 and 4.4 are drawn the curves $\tilde{\mathcal{H}}^\pm_\alpha(e_B)$ 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 4.3, 4.4, and easily verified also formally, that all alternative power entropy bounds (4.27) - (4.34) 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

$$\tilde{e}^\mp_\alpha(\tilde{H}) = \max_{\tilde{H}^\pm_\alpha(e) \le \tilde{H}} e \quad \text{and} \quad \tilde{e}^\mp_{B,\alpha}(\tilde{H}) = \max_{\tilde{H}^\pm_\alpha(e_B) \le \tilde{H}} e_B \tag{4.35}$$

(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

$$\tilde{e}^-_\alpha(\tilde{H}) < \tilde{e}^+_\alpha(\tilde{H}) \;\; \text{and} \;\; \tilde{e}^-_{B,\alpha}(\tilde{H}) < \tilde{e}^+_{B,\alpha}(\tilde{H}) \tag{4.36}$$

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 (4.36) we obtain the attainable upper and lower estimates $\tilde{e}^\pm_\alpha(\tilde{H}_\alpha(\pi))$, $\tilde{e}^\pm_\alpha(\tilde{H}_\alpha(\pi_x))$ and $\tilde{e}^\pm_{B,\alpha}(\tilde{H}_\alpha(\mathcal{E}))$ 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

Previous section demonstrated that the Bayes decision errors

$$e \in \{e(\pi), e(\pi_x), e_B(\mathcal{E})\} \tag{5.1}$$

depend on the levels achieved by the respective information criteria (prior, posterior or conditional power entropies and alternative power entropies)

$$H_\alpha \in \{H_\alpha(\pi), H_\alpha(\pi_x), H_\alpha(\mathcal{E})\} \text{ and } \tilde{H}_\alpha \in \{\tilde{H}_\alpha(\pi), \tilde{H}_\alpha(\pi_x), \tilde{H}_\alpha(\mathcal{E})\} \tag{5.2}$$

and vice versa. We remind that the range of the errors e is the interval $[0, c_n]$ and the range of the power entropy H_α or the alternative power entropy $\tilde{H}_\alpha$ 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 (5.2) by means of the errors (5.1) and vice versa. For simplicity, we restrict ourselves to the posterior Bayes errors and posterior entropies

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

and related estimates

$$\mathcal{H}^-_\alpha(e_B) \leq \mathcal{H}^+_\alpha(e_B), \quad \tilde{\mathcal{H}}^-_\alpha(e_B) \leq \tilde{\mathcal{H}}^+_\alpha(e_B) \tag{5.3}$$

and

$$e^-_{B,\alpha}(H_\alpha) < e^+_{B,\alpha}(H_\alpha), \quad \tilde{e}^-_{B,\alpha}(\tilde{H}_\alpha) < \tilde{e}^+_{B,\alpha}(\tilde{H}_\alpha) \tag{5.4}$$

established by Theorems 4.1, 4.2 and their corollaries. Similar results for the prior Bayes errors and prior entropies

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

and related estimates $H^\pm_\alpha(e), \tilde{H}^\pm_\alpha(e)$ and $e^\pm_\alpha(H_\alpha), \tilde{e}^\pm_\alpha(\tilde{H}_\alpha)$ mentioned or established in previous section follow similarly as below.

By (5.3), under a given Bayes decision error e_B the corresponding conditional entropies 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 tight estimates in the sense that all their values are achievable by these entropies in the situations with the Bayes error e_B. 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*

$$AI_n(H_\alpha|e_B) = \frac{1}{c_n}\int_0^{c_n}[\mathcal{H}_\alpha^+(e) - \mathcal{H}_\alpha^-(e))]\mathrm{d}e \tag{5.5}$$

and

$$AI_n(\tilde{H}_\alpha|e_B) = \frac{1}{c_n}\int_0^{c_n}[\tilde{\mathcal{H}}_\alpha^+(e) - \tilde{\mathcal{H}}_\alpha^-(e)]\mathrm{d}e \tag{5.6}$$

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*

$$AI_{n,\alpha}(e_B|H_\alpha) = \frac{1}{a_{\alpha,n}}\int_0^{a_{\alpha,n}}[e_{B,\alpha}^+(H) - e_{B,\alpha}^-(H)]\mathrm{d}H \tag{5.7}$$

and

$$AI_{n,\alpha}(e_B|\tilde{H}_\alpha) = \frac{1}{\tilde{a}_{\alpha,n}}\int_0^{\tilde{a}_{\alpha,n}}[\tilde{e}_{B,\alpha}^+(\tilde{H}) - \tilde{e}_{B,\alpha}^-(\tilde{H})]\mathrm{d}\tilde{H} \tag{5.8}$$

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 entropy bounds $\mathcal{H}_\alpha^\pm(e_B)$ satisfy the integral formulas

$$\int_0^{c_n} \mathcal{H}_\alpha^+(e)\,\mathrm{d}e = \begin{cases} \frac{1}{\alpha-1}\left[\frac{n-1}{n} - \frac{n^\alpha+n-2}{(\alpha+1)n^\alpha}\right] & \text{if } \alpha \neq 1 \\ \frac{1}{2n}\left[n-1+(n-2)\ln n\right] & \text{if } \alpha = 1 \end{cases} \tag{5.9}$$

and

$$\int_0^{c_n} \mathcal{H}_\alpha^-(e)\,\mathrm{d}e = \begin{cases} \frac{1}{2(\alpha-1)}\sum_{k=1}^{n-1}\frac{2-k^{1-\alpha}-(k+1)^{1-\alpha}}{k(k+1)} & \text{if } 0<\alpha<2, \alpha\neq 1 \\ \frac{1}{2}\sum_{k=1}^{n-1}\frac{\ln[k(k+1)]}{k(k+1)} & \text{if } \alpha = 1 \\ \frac{(n-1)(1-n^{1-\alpha})}{2(\alpha-1)n} & \text{if } \alpha \geq 2. \end{cases} \tag{5.10}$$

Proof. For $\alpha \neq 1$ the result of (5.9) follows by a routine integration of the power functions of $e = e_B$ appearing in the formula (4.18) 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 (5.9) 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

$$\int x \ln x \mathrm{d}x = \frac{x^2}{2}\left(\ln x - \frac{1}{2}\right) \tag{5.11}$$

obtained by differentiating the function $x^2 \ln x$. The upper and lower formulas of (5.10) follow by a routine integration of the linear or piecewise linear functions of $e = e_B$ appearing in the formulas (4.19) for the lower bound $\mathcal{H}_\alpha^-(e) = \mathcal{H}_\alpha^-(e_B)$. The middle formula of (5.10) can be obtained similarly as above, by taking the limit for $\alpha \to 1$ in the already proved upper formula of (5.10). Alternatively, we can integrate the piecewise linear function $\mathcal{H}_1^-(e) = \mathcal{H}_1^-(e_B)$ of (4.16). Details can be found in Appendix 1. ■

Formula (5.9) was obtained previously by Vajda and Zvárová (2007). Formula (5.10) 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

$$\int_0^{c_n} \tilde{\mathcal{H}}_\alpha^+(e_B)\,\mathrm{d}e_B = \begin{cases} \frac{1}{\alpha-1}\left[\frac{(n-1)^2}{n} - \frac{(n-1)^2}{\alpha+1} + \frac{n(n-2)}{\alpha+1}\left(\frac{n-1}{n}\right)^{\alpha+1}\right] & \text{if } \alpha \neq 1 \\ \frac{(n-1)^2}{2n}\left[1 + (n-2)\ln\frac{n-1}{n}\right] & \text{if } \alpha = 1 \end{cases} \tag{5.12}$$

and

$$\int_0^{c_n} \tilde{\mathcal{H}}_\alpha^-(e_B)\,\mathrm{d}e_B = \begin{cases} \frac{(n-1)^2}{2n(\alpha-1)}\left[1 - \left(\frac{n-1}{n}\right)^{\alpha-1}\right] & \text{if } 0 < \alpha \leq 2, \alpha \neq 1 \\ \frac{(n-1)^2}{2n}\ln\frac{n}{n-1} & \text{if } \alpha = 1 \\ \frac{1}{2(\alpha-1)}\sum_{k=1}^{n-1}\frac{2k-1-(k-1)\left(\frac{k-1}{k}\right)^{\alpha-1} - k\left(\frac{k}{k+1}\right)^{\alpha-1}}{k(k+1)} & \text{if } \alpha > 2. \end{cases} \tag{5.13}$$

Proof. Similarly as in the previous proof, for $\alpha \neq 1$ the result of (5.12) follows by a routine integration of the power functions of $e = e_B$ appearing in the formula (4.27) 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 (5.12) 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 (5.11). The upper and lower formulas of (5.13) follow by a routine integration of the linear or piecewise linear functions of $e = e_B$ appearing in the formulas (4.28) for the lower bound $\tilde{\mathcal{H}}_\alpha^-(e) = \tilde{\mathcal{H}}_\alpha^-(e_B)$. The middle formula of (5.13) can be obtained by taking the limit for $\alpha \to 1$ in the already proved upper formula of (5.13) 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 1. ■

Theorem 5.1. The average inaccuracies $AI_n(H_\alpha|e_B)$ and $AI_n(\tilde{H}_\alpha|e_B)$ of estimation of the power entropies $H_\alpha = H_\alpha(\mathcal{E})$ 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

$$AI_n(H_\alpha|e_B) = \frac{1}{c_n}\Big(\int_0^{c_n} \mathcal{H}_\alpha^+(e)\mathrm{d}e - \int_0^{c_n} \mathcal{H}_\alpha^-(e))\mathrm{d}e\Big) \tag{5.14}$$

and

$$AI_n(\tilde{H}_\alpha|e_B) = \frac{1}{c_n}\Big(\int_0^{c_n} \tilde{\mathcal{H}}_\alpha^+(e)\mathrm{d}e - \int_0^{c_n} \tilde{\mathcal{H}}_\alpha^-(e)\mathrm{d}e\Big) \tag{5.15}$$

where the integrals are given by lemmas 5.1 and 5.2.

Proof. Clear from (5.5), (5.6) 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 entropies $H_\alpha = H_\alpha(\mathcal{E})$ and $\tilde{H}_\alpha = \tilde{H}_\alpha(\mathcal{E})$ are given by the formulas

$$AI_{n,\alpha}(e_B|H_\alpha) = \frac{1}{a_{\alpha,n}}\Big(\int_0^{c_n} \mathcal{H}_\alpha^+(e)\mathrm{d}e - \int_0^{c_n} \mathcal{H}_\alpha^-(e)\mathrm{d}e\Big) \tag{5.16}$$

and

$$AI_{n,\alpha}(e_B|\tilde{H}_\alpha) = \frac{1}{\tilde{a}_{\alpha,n}}\Big(\int_0^{c_n} \tilde{\mathcal{H}}_\alpha^+(e)\mathrm{d}e - \int_0^{c_n} \tilde{\mathcal{H}}_\alpha^-(e)\mathrm{d}e\Big) \tag{5.17}$$

where the integrals are given by lemmas 5.1 and 5.2.

Proof. By the definitions of inverse functions (5.1) and (5.3), 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 (5.7), (5.8). ■

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 5.1 - 5.4 and the numerical values for $2 \le n \le 1000$ are in Tables 5.1 - 5.4. We see from these results that the minima of $AI_{n,\alpha}(e_B|H_\alpha)$ and $AI_{n,\alpha}(e_B|\tilde{H}_\alpha)$ are achieved at $\alpha = 2$ for all $n \ge 2$. The minima of $AI_n(\tilde{H}_\alpha|e_B)$ are achieved there only for $n > 4$ and the remaining minima, 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 $\alpha = 2$ demonstrates 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 for this estimation in Vajda (1968).

Conclusion 5.2 By comparing Figures 5.3 and 5.4 or Tables 5.3 and 5.4 one can see that the inaccuracies $AI_{n,\alpha}(e_B|\tilde{H}_\alpha)$ are slightly less than $AI_{n,\alpha}(e_B|H_\alpha)$ for almost all powers α and the state space sizes n except the optimal power $\alpha = 2$ where they coincide 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 estimation of the Bayes decision errors and 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 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 (2.1) on a space Θ of size

$$n = |\Theta|$$

and with an experiment $\mathcal{E}$ (cf. (2.2)). The losses are characterized by the median loss and the loss range

$$\Lambda > 0, \quad \Delta \ge 0 \quad \text{cf. (3.5)} \tag{6.1}$$

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

$$L_B(\pi), \;\; L_B(\pi_x) \;\text{ and }\; R_B = R_B(\mathcal{E}) \quad \text{(cf. (2.5) and (2.9))} \tag{6.2}$$

respectively.

In the next theorem the knowledge about experiment $\mathcal{E}$ is represented by the prior, posterior and average power entropies

$$H_\alpha(\pi),\ \ H_\alpha(\pi_x)\ \text{ and }\ H_\alpha(\mathcal{E})\ \text{ for some } \alpha > 0 \text{ (cf. (4.3) and (4.5))} \tag{6.3}$$

respectively. We study the tight upper bounds

$$H_\alpha^+(L_B|\Lambda,\Delta) = \max_{L_B(\pi)=L_B} H_\alpha(\pi) \equiv \max_{L_B(\pi_x)=L_B} H_\alpha(\pi_x) \tag{6.4}$$

$$\mathcal{H}_\alpha^+(R_B|\Lambda,\Delta) = \max_{R_B(\mathcal{E})=R_B} H_\alpha(\mathcal{E}) \tag{6.5}$$

and the tight lower bounds

$$H_\alpha^-(L_B|\Lambda,\Delta) = \min_{L_B(\pi)=L_B} H_\alpha(\pi) \equiv \min_{L_B(\pi_x)=L_B} H_\alpha(\pi_x), \tag{6.6}$$

$$\mathcal{H}_\alpha^-(R_B|\Lambda,\Delta) = \min_{R_B(\mathcal{E})=R_B} H_\alpha(\mathcal{E}) \tag{6.7}$$

for these entropies at given values of the prior and posterior Bayes losses and the Bayes risk appearing in (6.2), respectively.

Theorem 6.1. The bounds (6.4)-(6.7) are given in the whole definition domains

$$0 \le L_B \le c_n(\Lambda + \Delta/2) \quad \text{and} \quad 0 \le R_B \le c_n(\Lambda + \Delta/2) \tag{6.8}$$

by the formulas

$$H_\alpha^\pm(L_B|\Lambda,\Delta) = H_\alpha^\pm\left(\frac{L_B}{1 \mp \Delta/2}\right) \quad \text{and} \quad \mathcal{H}_\alpha^\pm(R_B|\Lambda,\Delta) = \mathcal{H}_\alpha^\pm\left(\frac{R_B}{1 \mp \Delta/2}\right) \tag{6.9}$$

for $H_\alpha^\pm(\cdot)$, $\mathcal{H}_\alpha^\pm(\cdot)$ defined in the domain $[0, c_n]$ by (4.13)-(4.16) and extended to $t > c_n$ by

$$H_\alpha^+(t) = H_\alpha^+(c_n) \equiv a_{\alpha,n}, \quad \mathcal{H}_\alpha^+(t) = \mathcal{H}_\alpha^+(c_n) \equiv a_{\alpha,n} \tag{6.10}$$

where

$$c_n = \frac{n}{n-1} \quad \text{and} \quad a_{\alpha,n} = \begin{cases} \frac{n^{1-\alpha}-1}{1-\alpha} & \text{if } \alpha \ne 1 \\ \ln n & \text{if } \alpha = 1 \end{cases} \quad \text{(cf. (4.10))}. \tag{6.11}$$

Proof. 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

$$\frac{L_B(\pi)}{\Lambda + \Delta/2} \le e_B(\pi) \le \max\left\{\frac{L_B}{1-\Delta/2},\, c_n\right\} \tag{6.12}$$

$$\frac{L_B(\pi)}{\Lambda + \Delta/2} \le e_B(\pi_x) \le \max\left\{\frac{L_B}{1-\Delta/2},\, c_n\right\} \tag{6.13}$$

$$\frac{R_B(\mathcal{E})}{\Lambda + \Delta/2} \le e_B(\mathcal{E}) \le \max\left\{\frac{R_B(\mathcal{E})}{\Lambda+\Delta/2},\, c_n\right\} \tag{6.14}$$

for c_n given by (6.11) and these bounds are tight. Applying these bounds in the definitions (6.4)-(6.7) of $H_\alpha^\pm(L_B|\Lambda,\Delta)$ and $\mathcal{H}_\alpha^\pm(R_B|\Lambda,\Delta)$ and using the definitions (4.13)-(4.16) of $H_\alpha^\pm(e)$ and $\mathcal{H}_\alpha^\pm(e_B)$ we get the desired formulas (6.9). The new bounds (6.4)-(6.7) are attained because the initial bounds (4.13)-(4.16) were proved to be attained. ■

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

$$L_{B,\alpha}^+(H|\Lambda,\Delta) = \max_{H_\alpha(\pi)=H} L_B(\pi) = \max_{H_\alpha(\pi_x)=H} L_B(\pi_x) \tag{6.15}$$

$$R_{B,\alpha}^+(H|\Lambda,\Delta) = \max_{H_\alpha(\mathcal{E})=H} R_B(\mathcal{E}) \tag{6.16}$$

and the tight lower bounds

$$L_{B,\alpha}^-(H|\Lambda,\Delta) = \min_{H_\alpha(\pi)=H} L_B(\pi) = \min_{H_\alpha(\pi_x)=H} L_B(\pi_x) \tag{6.17}$$

$$R_{B,\alpha}^-(H|\Lambda,\Delta) = \min_{H_\alpha(\mathcal{E})=H} R_B(\mathcal{E}) \tag{6.18}$$

of the Bayes losses and risks (6.2) in the models with loss parameters Λ, Δ and given values of the power entropies (6.3).

Corollary 6.1. The tight upper and lower bounds (6.15) - (6.18) are given in the corresponding definition domains

$$0 \le H \le a_{\alpha,n}, \quad \alpha > 0 \tag{6.19}$$

of the power entropies (6.3) by the formulas

$$L_{B,\alpha}^\pm(H|\Lambda,\Delta) = e_\alpha^\pm(H)(\Lambda \pm \Delta/2), \quad R_{B,\alpha}^\pm(H|\Lambda,\Delta) = e_{B,\alpha}^\pm(H)(\Lambda \pm \Delta/2) \tag{6.20}$$

for $e_\alpha^\pm(H)$, $e_{B,\alpha}^\pm(H)$ defined by (4.22).

Now we deal with the situation where the knowledge about experiment $\mathcal{E}$ is represented by the prior, posterior and average alternative entropies power entropies

$$\tilde{H}_\alpha(\pi),\ \ \tilde{H}_\alpha(\pi_x)\ \text{ and }\ \tilde{H}_\alpha(\mathcal{E})\ \text{ for some } \alpha > 0 \text{ (cf. (4.5))} \tag{6.21}$$

respectively. We study the tight upper bounds

$$\tilde{H}_\alpha^+(L_B|\Lambda,\Delta) = \max_{L_B(\pi)=L_B} \tilde{H}_\alpha(\pi) \equiv \max_{L_B(\pi_x)=L_B} \tilde{H}_\alpha(\pi_x) \tag{6.22}$$

$$\tilde{\mathcal{H}}_\alpha^+(R_B|\Lambda,\Delta) = \max_{R_B(\mathcal{E})=R_B} \tilde{H}_\alpha(\mathcal{E}) \tag{6.23}$$

and the tight lower bounds

$$\tilde{H}_\alpha^-(L_B|\Lambda,\Delta) = \min_{L_B(\pi)=L_B} \tilde{H}_\alpha(\pi) \equiv \min_{L_B(\pi_x)=L_B} \tilde{H}_\alpha(\pi_x), \tag{6.24}$$

$$\tilde{\mathcal{H}}_\alpha^-(R_B|\Lambda,\Delta) = \min_{R_B(\mathcal{E})=R_B} \tilde{H}_\alpha(\mathcal{E}) \tag{6.25}$$

of these entropies for given values of the prior and posterior Bayes losses and the Bayes risk appearing in (6.2).

Theorem 6.2. The bounds (6.4)-(6.7) are given in the whole definition domains

$$0 \le L_B \le c_n(\Lambda + \Delta/2) \quad \text{and} \quad 0 \le R_B \le c_n(\Lambda + \Delta/2) \tag{6.26}$$

by the formulas

$$\tilde{H}^{\pm}_{\alpha}(L_B|\Lambda, \Delta) = \tilde{H}^{\pm}_{\alpha}\left(\frac{L_B}{1 \mp \Delta/2}\right) \quad \text{and} \quad \tilde{\mathcal{H}}^{\pm}_{\alpha}(R_B|\Lambda, \Delta) = \tilde{\mathcal{H}}^{\pm}_{\alpha}\left(\frac{R_B}{1 \mp \Delta/2}\right) \tag{6.27}$$

for $\tilde{H}^{\pm}_{\alpha}(\cdot)$, $\mathcal{H}^{\pm}_{\alpha}(\cdot)$ defined in the domain $[0, c_n]$ by (4.27), (4.28) and for $\tilde{H}^{+}_{\alpha}(\cdot)$, $\mathcal{H}^{+}_{\alpha}(\cdot)$ extended to $t > c_n$ by

$$\tilde{H}^{+}_{\alpha}(t) = \tilde{H}^{+}_{\alpha}(c_n) \equiv \tilde{a}_{\alpha,n}, \quad \tilde{\mathcal{H}}^{+}_{\alpha}(t) = \tilde{\mathcal{H}}^{+}_{\alpha}(c_n) \equiv \tilde{a}_{\alpha,n} \tag{6.28}$$

where

$$c_n = \frac{n}{n-1} \quad \text{and} \quad \tilde{a}_{\alpha,n} = \begin{cases} \frac{n-1}{1-\alpha}\left[\left(\frac{n}{n-1}\right)^{1-\alpha} - 1\right] & \text{if } \alpha \ne 1 \\ (n-1)\ln\frac{n}{n-1} & \text{if } \alpha = 1 \end{cases} \quad \text{(cf. (4.10))}. \tag{6.29}$$

Proof. 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 (6.12) - (6.14) for c_n given by (6.11) and these bounds are tight. Applying these bounds in the definitions (6.22)-(6.25) of $\tilde{H}^{\pm}_{\alpha}(L_B|\Lambda, \Delta)$ and $\tilde{\mathcal{H}}^{\pm}_{\alpha}(R_B|\Lambda, \Delta)$ and using the definitions (4.27), (4.28) of $\tilde{H}^{\pm}_{\alpha}(e)$ and $\tilde{\mathcal{H}}^{\pm}_{\alpha}(e_B)$ we get the desired formulas (6.27). The new bounds (6.27) are attained because the initial bounds (4.27), (4.28) were proved to be attained. ∎

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

$$L^{+}_{B,\alpha}(\tilde{H}|\Lambda, \Delta) = \max_{\tilde{H}_{\alpha}(\pi) = \tilde{H}} L_B(\pi) \equiv \max_{\tilde{H}_{\alpha}(\pi_x) = \tilde{H}} L_B(\pi_x) \tag{6.30}$$

$$R^{+}_{B,\alpha}(\tilde{H}|\Lambda, \Delta) = \max_{\tilde{H}_{\alpha}(\mathcal{E}) = \tilde{H}} R_B(\mathcal{E}) \tag{6.31}$$

and the tight lower bounds

$$L^{-}_{B,\alpha}(\tilde{H}|\Lambda, \Delta) = \min_{\tilde{H}_{\alpha}(\pi) = \tilde{H}} L_B(\pi) \equiv \min_{\tilde{H}_{\alpha}(\pi_x) = \tilde{H}} L_B(\pi_x) \tag{6.32}$$

$$R^{-}_{B,\alpha}(\tilde{H}|\Lambda, \Delta) = \min_{\tilde{H}_{\alpha}(\mathcal{E}) = \tilde{H}} R_B(\mathcal{E}) \tag{6.33}$$

of the Bayes losses and risks (6.2) in models with parameters Λ, Δ and given values of the power entropies (6.21).

Corollary 6.2. The attainable upper bounds (6.30) - (6.33) are given in the definitions domains

$$0 \le \tilde{H} \le \tilde{a}_{\alpha,n}, \quad \alpha > 0 \tag{6.34}$$

of the power entropies (6.21) by the formulas

$$L^{\pm}_{B,\alpha}(\tilde{H}|\Lambda, \Delta) = e^{\pm}_{\alpha}(\tilde{H})(\Lambda \pm \Delta/2), \quad R^{\pm}_{B,\alpha}(\tilde{H}|\Lambda, \Delta) = e^{\pm}_{B,\alpha}(\tilde{H})(\Lambda \pm \Delta/2) \tag{6.35}$$

for $e^{\pm}_{\alpha}(\tilde{H})$, $e^{\pm}_{B,\alpha}(\tilde{H})$ defined by (4.35).

Conclusion 6.1 Conclusion 5.1 implies 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 entropies $H_\alpha = H_\alpha(\mathcal{E})$ is minimized at the power $\alpha = 2$.

Conclusion 6.2 Conclusion 5.2 implies 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 power entropies $\tilde{H}_\alpha = \tilde{H}_\alpha(\mathcal{E})$ 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 6.1 and 6.2 illustrate the power entropy bounds $H^\pm_\alpha(L_B|\Lambda,\Delta)$ for the entropy 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 $\mathcal{H}^\pm_\alpha(R_B|\Lambda,\Delta)$ for the same entropy and loss function parameters are in Figures 6.3 and 6.4. Inverse functions to the bounds of Figures 6.1 - 6.4 illustrate the corresponding prior and Bayes risk bounds $L^\pm_{B,\alpha}(H|\Lambda,\Delta)$ and $R^\pm_{B,\alpha}(H|\Lambda,\Delta)$.

References

M. Ben Bassat (1978). f-entropies, probability of error, and feature selection. *Information and Control* **39**, 227-242.

M. Ben Bassat and J. Raviv (1978). Rényi's entropy and probability of error. *IEEE Transactions on Information Theory* **24,** 324-331.

Berger J. O. (1986). *Statistical Decision Theory and Bayesian Analysis.* 2-nd Ed., Springer, Berlin.

T. M.Cover and P. E. Hart (1967). Nearest neighbor pattern classification. *IEEE Transactions on Information Theory* **13,** 21-27.

P. Devijver and J. Kittler (1982). *Pattern Recognition. A Statistical Approach.* Prentice Hall, Englewood Cliffs, New Jersey.

L. Devroye, L. Györfi and G. Lugosi (1996). *A Probabilistic Theory of Pattern Recognition.* Springer, Berlin.

D. K. Faddeev (1957). Zum Begriff der Entropie einer endlichen Wahrscheinlichkeitsschemas. *Arbeits zur informationstheorie,* vol. I, Deutscher Verlag der Wissenschaften, Berlin.

M. Feder

and N. Merhav (1994). Relations between entropy and error probability. *IEEE Transactions on Information Theory* **40,** 259-266.

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

L. Kanal (1974). Patterns in pattern recognittion. *IEEE Transactions on Information Theory* **20,** 697-707.

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

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

A. Rényi (1961). On measures of entropy and information. In *Proceedings of 4-th Berkel;ey Symp. on Probab. Statist.* Univ. of California Press, Berkeley, California.

N. P. Salichov (1974). Confirmation of a hypothesis of I. Vajda (in Russian). *Problemy Peredachi Informacii* **10**, 114-115.

D. L. Tebbe and S. J. Dwyer III (1968). Uncertainty and probability of error. *IEEE Transactions on Information Theory*, **14,** 516-518.

G. T. Toussaint (1977). A generalization of Shannon's equivocation and the Fano bound. *IEEE Transactions on System, Man and Cybernetics* **7**, 300-302.

I. Vajda (1968). Bounds on the minimal error probability and checking a finite or countable number of hypotheses. *Information Transmission Problems* **4**, 9-17.

I. Vajda (1969). A contribution to informational analysis of patterns. In *Methodologies of Pattern Recognition* (Ed. M. S. Watanabe). Academic Press, New York.

I. Vajda and K. Vašek (1985). Majorization, concave entropies and comparison of experiments. *Problems of Control and Information Theory* **14**, 105-115.

I. Vajda and J. Zvárová (2007). On generalized entropies, Bayesian decisions and statistical diversity. *Kybernetika* **43**, 675-696.

Appendix 1: Proofs of Lemmas 5.1 and 5.2

Lemma 5.1, formulas of (5.9).

(i) If $\alpha > 0$, $\alpha \neq 1$ then using the upper formula of (4.18) 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 (4.18) 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

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

Lemma 5.1, formulas of (5.10).

(i) If $0 < \alpha < 2$ then using the upper formula of (4.19) we obtain

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

(ii) If $\alpha \geq 2$, then using the lower formula of (4.19) 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 (5.12).

(i) If $\alpha > 0, \alpha \neq 1$, then by the upper formula of (4.27)

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

(ii) If $\alpha = 1$, then by the lower formula of (5.12)

$$\begin{aligned}
&\int_0^{c_n} \tilde{\mathcal{H}}_1^+(e)\,\mathrm{d}e = \int_0^{c_n}(-e\ln e - (n-1-e)\ln(n-1-e) + (n-1-e)\ln(n-1))\,\mathrm{d}e\\
&= \frac{1}{2}\left[-e^2(\ln e - \frac{1}{2}) + (n-1-e)^2(\ln(n-1-e) - \frac{1}{2}) - (n-1-e)^2\ln(n-1)\right]_0^{\frac{n-1}{n}}\\
&= -\frac{1}{2}\Big(\frac{n-1}{n}\Big)^2\Big[\ln\frac{n-1}{n} - \frac{1}{2}\Big] + \frac{1}{2}\frac{(n-1)^4}{n^2}\Big[\ln\frac{(n-1)^2}{n} - \frac{1}{2}\Big] - \frac{1}{2}\frac{(n-1)^4}{n^2}\ln(n-1)\\
&+ 0 - \frac{(n-1)^2}{2}\Big[\ln(n-1) - \frac{1}{2}\Big] + \frac{(n-1)^2}{2}\ln(n-1)\\
&= -\frac{1}{2}\Big(\frac{n-1}{n}\Big)^2\ln(n-1) + \frac{1}{2}\Big(\frac{n-1}{n}\Big)^2\ln n + \frac{1}{4}\Big(\frac{n-1}{n}\Big)^2 + \frac{(n-1)^4}{n^2}\ln(n-1)\\
&- \frac{1}{2}\frac{(n-1)^4}{n^2}\ln n - \frac{1}{4}\frac{(n-1)^4}{n^2} - \frac{1}{2}\frac{(n-1)^4}{n^2}\ln(n-1) + \frac{(n-1)^2}{4}\\
&= \frac{1}{2}\Big(\frac{n-1}{n}\Big)^2 n(n-2)\ln(n-1) - \frac{1}{2}\Big(\frac{n-1}{n}\Big)^2 n(n-2)\ln n + \frac{1}{4}\Big(\frac{n-1}{n}\Big)^2 2n\\
&= \frac{(n-1)^2}{2n}\left[1 + (n-2)\ln\frac{n-1}{n}\right].
\end{aligned}$$

Lemma 5.2, formulas (5.13).

(i) If $0 < \alpha < 2, \alpha \neq 1$ then by the lower formula of (4.28) and by the definition of $\tilde{a}_{\alpha,n}$ in (4.25) 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 (4.28) and by the definition of $\tilde{a}_{\alpha,n}$ in (4.25) 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 (4.28)

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

By (4.25)

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

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,

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

Appendix 2: Tables

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 5.1. 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 5.2. 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 5.3. 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 5.4. Alternative average inaccuracies $AI_{n,\alpha}(e_B|\tilde{H}_\alpha)$ for selected α and n.

Appendix 3: Figures

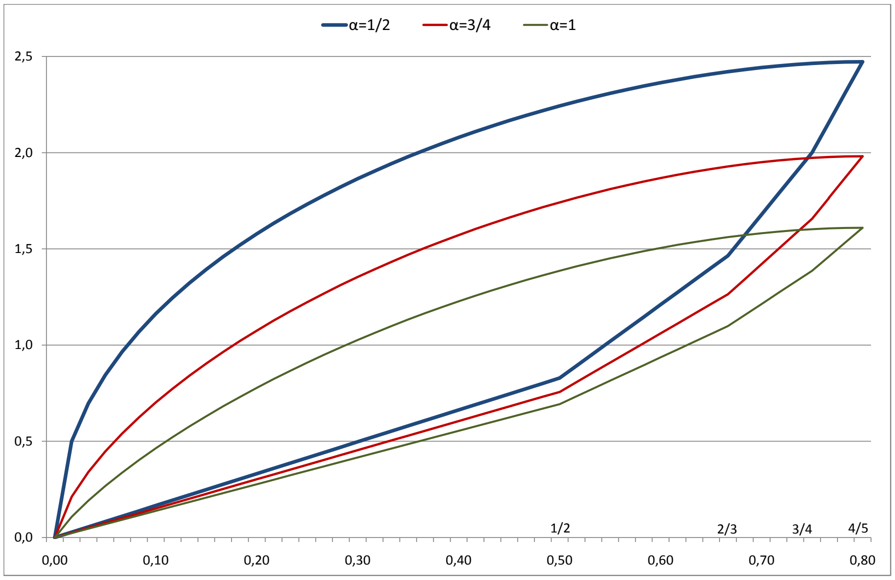

Figure 4.1: $\mathcal{H}^{\pm}_{\alpha}(e_B)$ as functions of variable e_B for $\alpha = 1/2, 3/4, 1$.

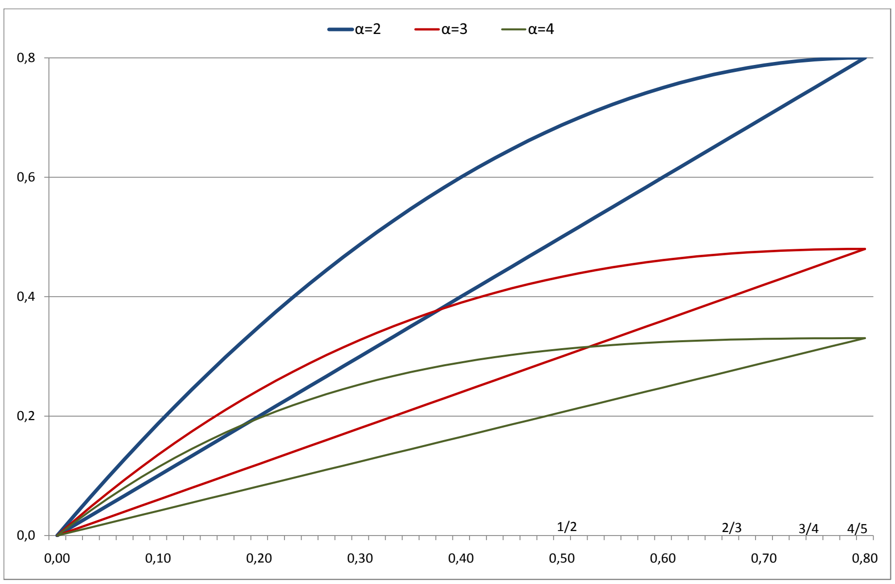

Figure 4.2: $\mathcal{H}^{\pm}_{\alpha}(e_B)$ as functions of variable e_B for $\alpha = 2, 3, 4$.

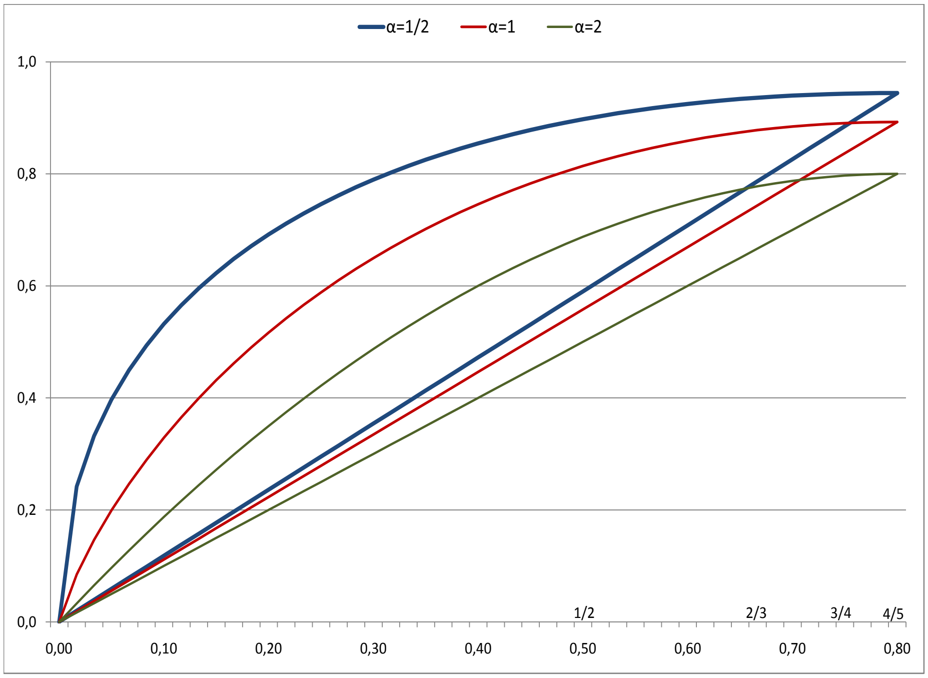

Figure 4.3: $\tilde{\mathcal{H}}_{\alpha}^{\pm}(e_B)$ as functions of variable e_B for $\alpha = 1/2, 1, 2$.

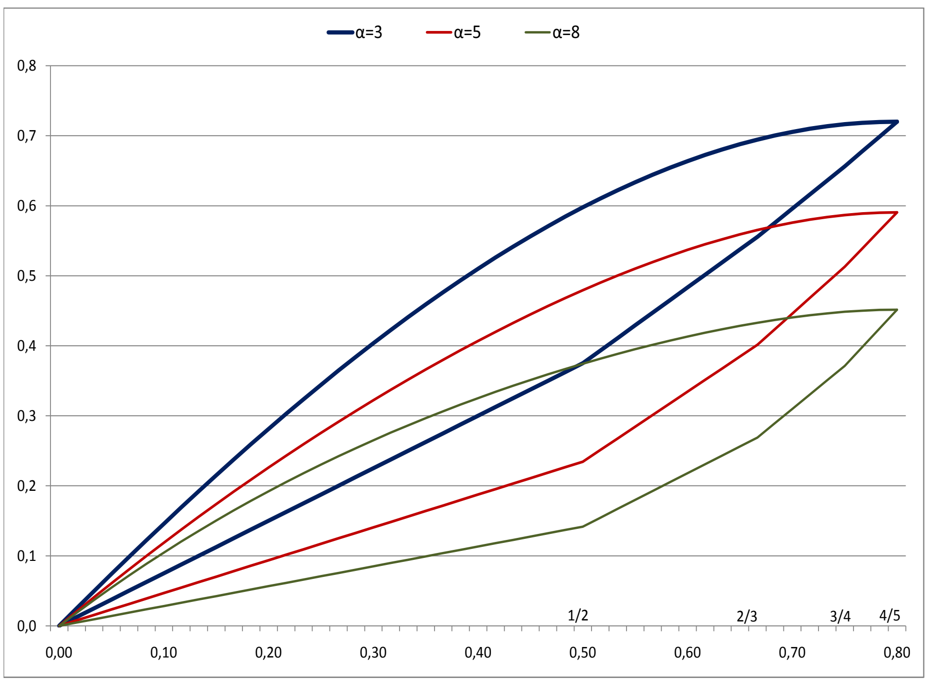

Figure 4.4: $\tilde{\mathcal{H}}_{\alpha}^{\pm}(e_B)$ as functions of variable e_B for $\alpha = 3, 5, 8$.

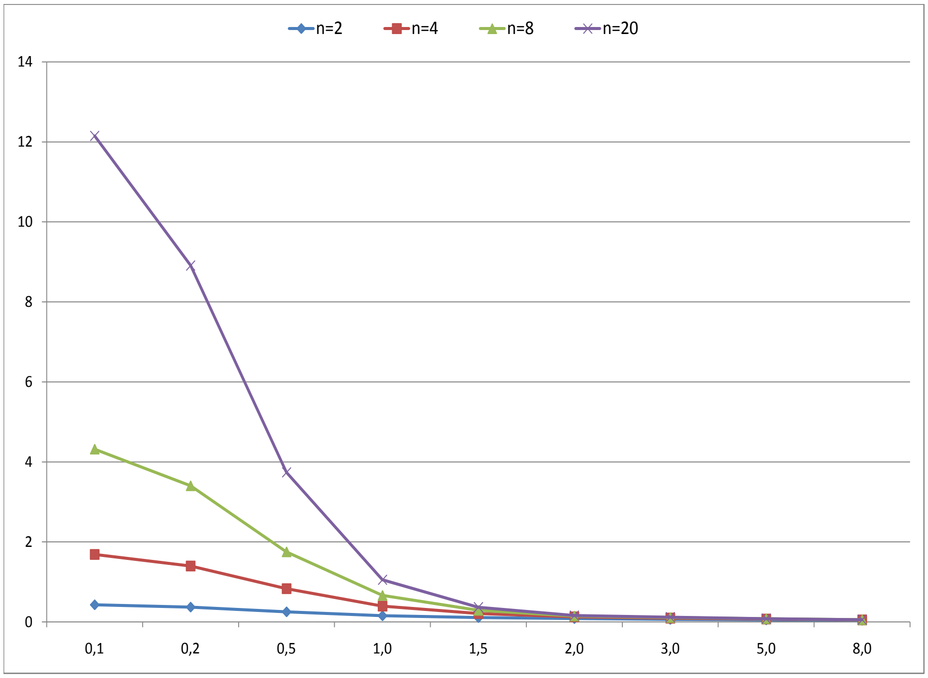

Figure 5.1: Average inaccuracies $AI_n(H_\alpha|e_B)$ for selected n as function of α.

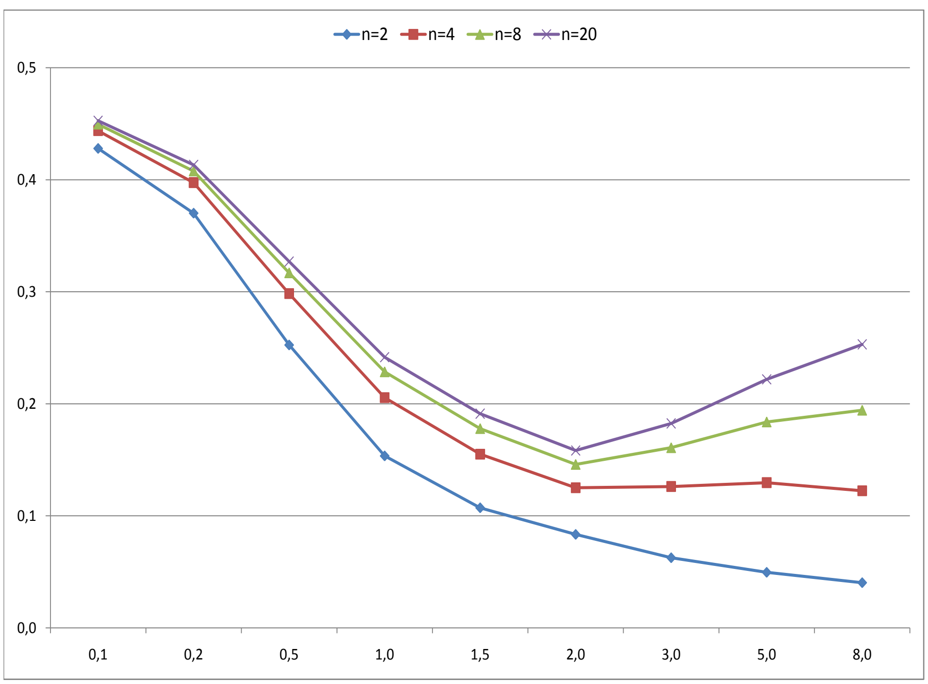

Figure 5.2: Alternative average inaccuracies $AI_n(\tilde{H}_\alpha|e_B)$ for selected n as function of α.

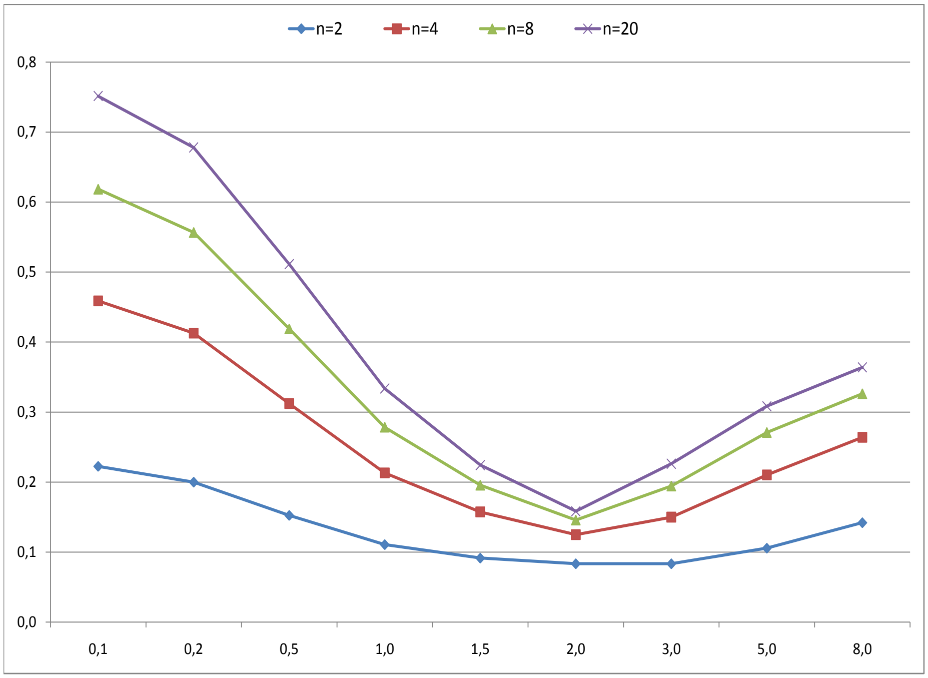

Figure 5.3 Average inaccuracies $AI_{n,\alpha}(e_B|H_\alpha)$ for selected n as function of α.

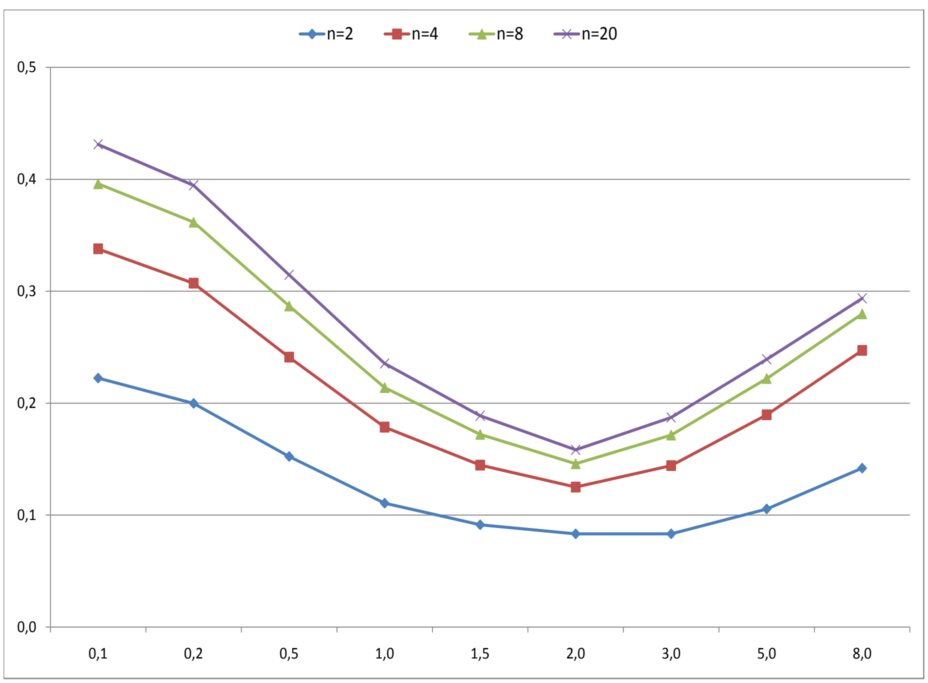

Figure 5.4: Alternative average inaccuracies $AI_{n,\alpha}(e_B|\tilde{H}_\alpha)$ for selected n as function of α.

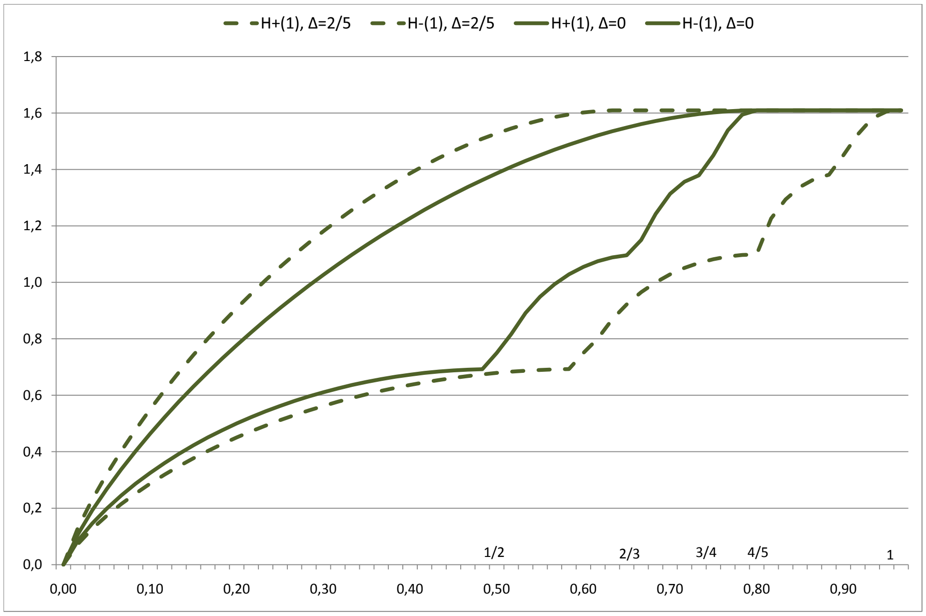

Figure 6.1: Entropy bounds $H_1^{\pm}(L_B|\Lambda, \Delta)$ for $\Lambda = 1$ and $\Delta = 0$ (full line) or $\Delta = 2/5$ (interrupted line).

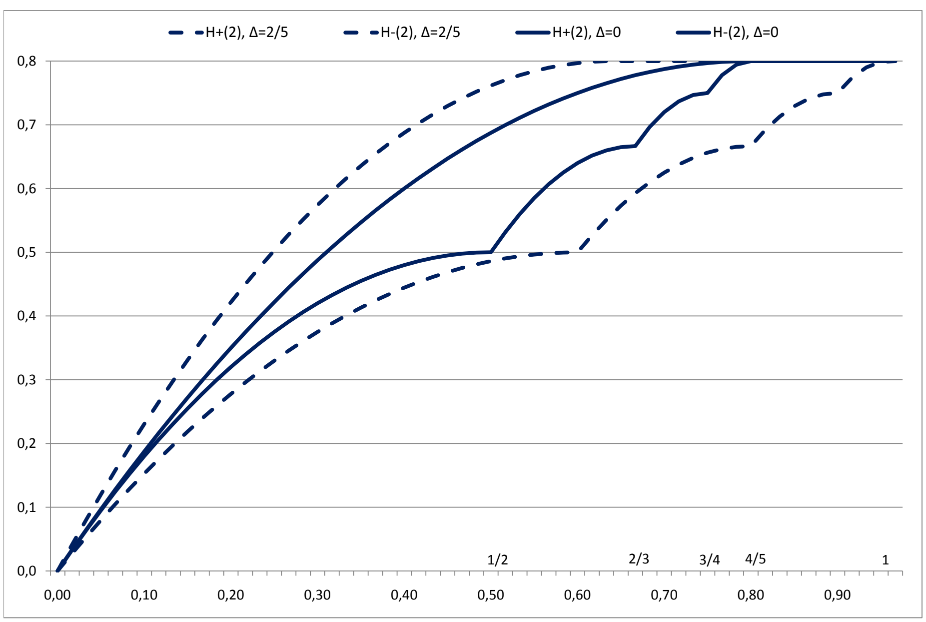

Figure 6.2: Entropy bounds $H_2^{\pm}(L_B|\Lambda, \Delta)$ for $\Lambda = 1$ and $\Delta = 0$ (full line) or $\Delta = 2/5$ (interrupted line).

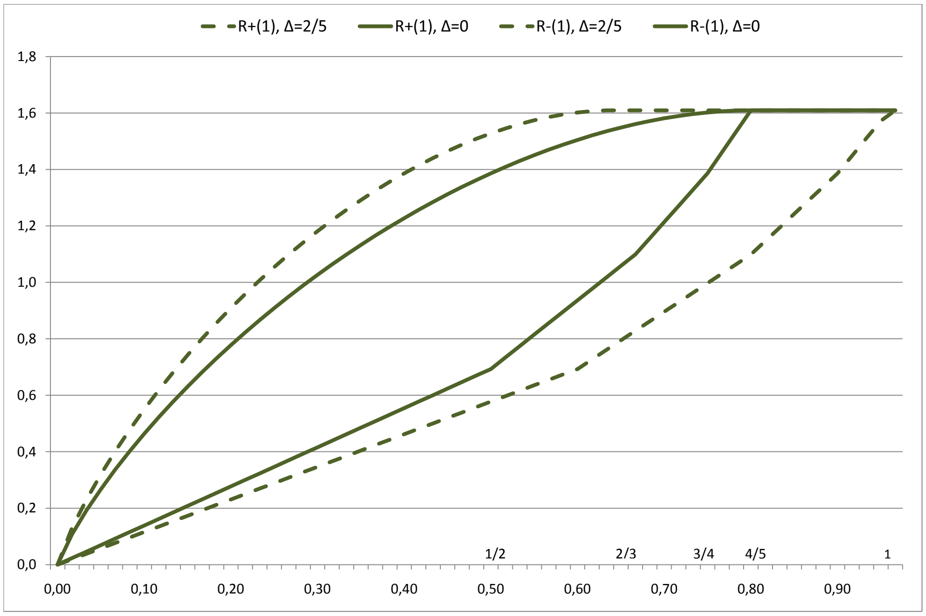

Figure 6.3: Entropy bounds $\mathcal{H}_1^{\pm}(R_B|\Lambda, \Delta)$ for $\Lambda = 1$ and $\Delta = 0$ (full line) or $\Delta = 2/5$ (interrupted line).

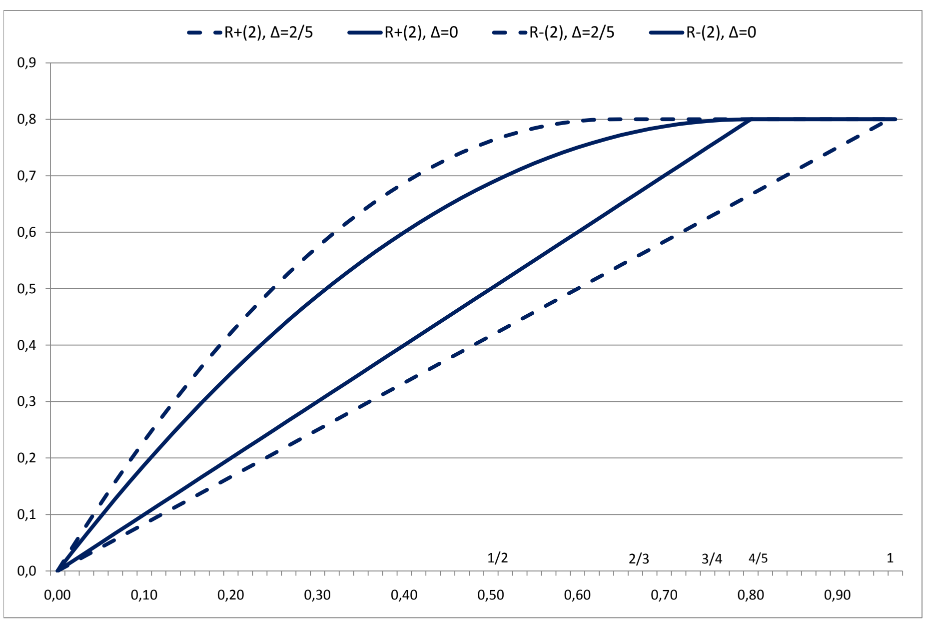

Figure 6.4: Entropy bounds $\mathcal{H}_2^{\pm}(R_B|\Lambda, \Delta)$ for $\Lambda = 1$ and $\Delta = 0$ (full line) or $\Delta = 2/5$ (interrupted line).