Generalized Informations and Bayesian errors*

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.

Abstract

Key words: Power informations, Rényi informations, Bayesian errors

1. Introduction

We consider decisions about a random variable $X \sim (\{1, 2, \dots, n\}, P_X)$ on the basis of a random observation $Y \sim (\mathcal{Y}, \mathcal{B}, P_Y)$. By $P_{X|y}$ we denote the conditional probability distribution of X given $Y = y$ for $y \in \mathcal{Y}$. Let

$$e(X) = e(P_X) = 1 - \max_{1 \le i \le n} P_X(i) \tag{1.1}$$

be the error of the a priori Bayesian decision about X. We are interested in the error

$$e(X|Y) = \int_{\mathcal{Y}} e(P_{X|y}) dP_Y(y) \tag{1.2}$$

of the a posteriori Bayesian decision about X based on the observation Y.

We shall estimate the Bayesian decision errors $e(X|Y)$ by means of the conditional entropies

$$H_\alpha(X|Y) = \int_{\mathcal{Y}} H_\alpha(P_{X|y}) dP_Y(y) \tag{1.3}$$

of order $\alpha > 0$ where the unconditional (a priori) entropies

$$H_\alpha(P_X) \equiv H_\alpha(X), \quad \alpha > 0.$$

* Supported by grants BMF 2003-04820 and MŠMT 1M0572

are of two possible types, namely the *power entropies*

$$H_\alpha^I(P_X) \equiv H_\alpha^I(X) = \frac{1}{1-\alpha}\left(\sum_{i=1}^n P_X(i)^\alpha - 1\right), \quad \alpha \neq 1 \tag{1.4}$$

or the *Rényi entropies*

$$H_\alpha^{II}(P_X) \equiv H_\alpha^{II}(X) = \frac{1}{1-\alpha}\ln\sum_{i=1}^n P_X(i)^\alpha, \quad \alpha \neq 1 \tag{1.5}$$

with common limit

$$H_1^I(X) \equiv H_1^{II}(X) = H_1(X) = -\sum_{i=1}^n P_X(i)\ln P_X(i) \tag{1.6}$$

known as the *Shannon entropy.* In (1.3) the particular power conditional entropies will be derived by $H^I(X|Y)$ and the conditional Rényi entropies by $H^{II}(X|Y)$. As it is well known, the differences

$$I_\alpha^I(X,Y) = H_\alpha^I(X) - H_\alpha^I(X|Y) \quad \text{and} \quad I_\alpha^{II}(X,Y) = H_\alpha^{II}(X) - H_\alpha^{II}(X|Y)$$

are generalized informations, namely the power informations and Rényi informations of orders $\alpha > 0$, see Morales et al. (1996). For the particular order these generalized informations reduce to the classical Shannon information

$$I(X,Y) = I_1^I(X,Y) = I_1^{II}(X,Y).$$

Thus in fact, our paper studies reations between the Bayes errors $e(X|Y)$ (Bayes risks for the $0-1$ loss functions) and the generalized informations of two types, the power informations $I_\alpha^I(X,Y)$ of orders $\alpha > 0$ and the Rényi informations $I_\alpha^{II}(X,Y)$ of orders $\alpha > 0$.

The Bayesian errors $e(X)$ and $e(X|Y)$ take on values in the interval

$$0 \leq e(X),\, e(X|Y) \leq e_n \tag{1.7}$$

where here and in the sequel

$$e_n = \frac{n-1}{n}, \quad \text{for} \quad n = 1, 2, \ldots \tag{1.8}$$

By Theorem 2 in Morales et al. (1996),

$$e(X) = e \quad \text{implies} \quad \underline{H}_\alpha(e) \leq H_\alpha(X) \leq \overline{H}_\alpha(e) \tag{1.9}$$

where if $\alpha \neq 1$ then the (estimable) bounds are

$$\overline{H}_\alpha^I(e) = \frac{(n-1)^{1-\alpha}e^\alpha + (1-e)^\alpha - 1}{1-\alpha}, \tag{1.10}$$

$$\underline{H}^I_\alpha(e) = \sum_{k=1}^{n-1} \frac{(1-k(1-e))^\alpha + k(1-e)^\alpha - 1}{1-\alpha} I(e_k < e \leq e_{k+1}) \tag{1.11}$$

and

$$\overline{H}^{II}_\alpha(e) = \frac{1}{\alpha} \ln\left((n-1)^{1-\alpha} e^\alpha + (1-e)^\alpha\right), \tag{1.12}$$

$$\underline{H}^{II}_\alpha(e) = \frac{1}{1-\alpha} \sum_{k=1}^{n-1} \ln\left((1-k(1-e)^\alpha + k(1-e)^\alpha\right) I(e_k < e \leq e_{k+1}) \tag{1.13}$$

For $\alpha = 1$ the bounds are common limits of (1.10), (1.11) and (1.12), (1.13) for $\alpha \to 1$, namely

$$\overline{H}_1(e) = h(e) + e\ln(n-1), \tag{1.14}$$

$$\underline{H}_1(e) = -\sum_{k=1}^{n-1} \{[1-k(1-e)]\ln[1-k(1-e)] + k(1-e)\ln(1-e)\}\, I(e_k < e \leq e_{k+1}) \tag{1.15}$$

where

$$h(e) = -e\ln e - (1-e)\ln(1-e). \tag{1.16}$$

Further, by Theorem 4 in Morales et al. (1996)

$$e(X|Y) = e \quad \text{implies} \quad \underline{\underline{H}}_\alpha(e) \leq H_\alpha(X|Y) \leq \overline{H}_\alpha(e) \tag{1.17}$$

where the (attainable) upper bounds $\overline{H}_\alpha(e)$ are given by (1.10), (1.12) and (1.14) above. The (attainable) lower bounds are for $\alpha \neq 1$.

$$\underline{\underline{H}}_\alpha(e) = \frac{1}{1-\alpha} \sum_{k=1}^{n-1} \left(a_k + k(k+1)b_k(e-e_k)\right) I(e_k < e \leq e_{k+1}) \tag{1.18}$$

where

$$a^I_k = k^{1-\alpha} - 1, \quad b^I_k = (k+1)^{1-\alpha} - k^{1-\alpha} \tag{1.19}$$

and

$$a^{II}_k = (1-\alpha)\ln k, \quad b^{II}_k = (1-\alpha)\ln\frac{k+1}{k} \tag{1.20}$$

while the bound

$$\underline{\underline{H}}_1(e) = \sum_{k=1}^{n-1} \left(\ln k + \ln k(k+1)(e-e_k)\ln\frac{k+1}{k}\right) I(e_k < e \leq e_{k+1}) \tag{1.21}$$

is the common limit of $\underline{\underline{H}}^I_\alpha(e)$ and $\underline{\underline{H}}^{II}_\alpha(e)$ for $\alpha \to 1$.

Problem: To investigate how tight are attainable bounds for $e(X|Y)$ obtained from $H^I_\alpha(X|Y)$ and $H^{II}_\alpha(X|Y)$ as functions of $\alpha > 0$.

We shall reduce this problem to the evaluation of integrals

$$\tau_\alpha = \frac{1}{H_{\alpha,n}} \int_0^{e_n} \left(\overline{H}_\alpha(e) - \underline{\underline{H}}_\alpha(e) \right) de \tag{1.22}$$

where

$$H_{\alpha,n} = H_{\alpha,max}(X), \quad \text{for } n\text{-valued } X. \tag{1.23}$$

In the next two sections are presented the results of numerical evaluation of these integrals for power informations (τ_α^I) and Rényi informations (τ_α^{II}) for selected values of n (rows) and α (columns). The corresponding analytical formulas and conclusions are under the current research. However, an immediate preliminary conclusion which can be even at this stage drawn from the Tables 2.1 and 3.1 is that the quadratic information (power information of the order $\alpha = 2$) is much more closely related to the Bayes decision error than the classical Shannon information.

2. Power informations and Bayesian errors

n	0.125	0.25	0.5	1	1.5	2	3	4
2	0.2163	0.1900	0.1524	0.1107	0.0914	0.0833	0.0833	0.0929
3	0.3619	0.3183	0.2538	0.2816	0.1345	0.1111	0.0938	0.0958
4	0.4465	0.3927	0.3121	0.3460	0.1574	0.1250	0.0981	0.0969
5	0.5034	0.4426	0.3511	0.3804	0.1720	0.1333	0.1004	0.0974
6	0.5448	0.4790	0.3795	0.4019	0.1821	0.1389	0.1017	0.0976
7	0.5767	0.5070	0.4014	0.4167	0.1897	0.1429	0.1025	0.0977
8	0.6021	0.5294	0.4190	0.4275	0.1956	0.1458	0.1031	0.0978
9	0.6230	0.5479	0.4335	0.4357	0.2004	0.1481	0.1035	0.0978
10	0.6404	0.5634	0.4457	0.4422	0.2043	0.1500	0.1038	0.0979
20	0.7319	0.6452	0.5113	0.4705	0.2242	0.1583	0.1047	0.0979
30	0.7699	0.6800	0.5400	0.4796	0.2322	0.1611	0.1049	0.0979
40	0.7916	0.7001	0.5571	0.4842	0.2367	0.1625	0.1050	0.0979
50	0.8058	0.7134	0.5688	0.4869	0.2397	0.1633	0.1050	0.0979
100	3.3185	2.5337	1.5735	0.9678	0.4113	0.2500	0.1051	0.0479
200	3.6874	2.7323	1.6311	0.9918	0.4118	0.2500	0.1051	0.0479
300	0.8668	0.7737	0.6268	0.4963	0.2536	0.1661	0.1051	0.0979
400	0.8711	0.7784	0.6322	0.4968	0.2547	0.1663	0.1051	0.0979
500	0.8740	0.7815	0.6358	0.4971	0.2555	0.1663	0.1051	0.0979
600	0.8759	0.7837	0.6385	0.4974	0.2561	0.1664	0.1051	0.0979
700	0.8774	0.7854	0.6406	0.4975	0.2566	0.1664	0.1051	0.0979
800	0.8785	0.7867	0.6423	0.4976	0.2569	0.1665	0.1051	0.0979
900	0.8794	0.7877	0.6437	0.4977	0.2572	0.1665	0.1051	0.0979
1000	0.8802	0.7886	0.6449	0.4978	0.2575	0.1665	0.1051	0.0979
1500	0.8826	0.7914	0.6489	0.4981	0.2583	0.1666	0.1051	0.0979
2000	0.8839	0.7930	0.6513	0.4982	0.2588	0.1666	0.1051	0.0979
2500	0.8847	0.7940	0.6529	0.4983	0.2592	0.1666	0.1051	0.0979
3000	0.8853	0.7948	0.6541	0.4984	0.2594	0.1666	0.1051	0.0979
3500	0.8857	0.7953	0.6550	0.4984	0.2596	0.1666	0.1051	0.0979
4000	0.8860	0.7957	0.6558	0.4985	0.2598	0.1666	0.1051	0.0979
4500	0.8863	0.7961	0.6564	0.4985	0.2599	0.1666	0.1051	0.0979
5000	0.8865	0.7964	0.6569	0.4985	0.2600	0.1666	0.1051	0.0979
5500	0.8867	0.7966	0.6574	0.4986	0.2601	0.1666	0.1051	0.0979
6000	0.8868	0.7968	0.6578	0.4986	0.2602	0.1666	0.1051	0.0979
6500	0.8869	0.7970	0.6581	0.4986	0.2603	0.1666	0.1051	0.0979
7000	0.8871	0.7971	0.6584	0.4986	0.2603	0.1666	0.1051	0.0979
7500	0.8872	0.7973	0.6587	0.4986	0.2604	0.1666	0.1051	0.0979
8000	0.8873	0.7974	0.6590	0.4987	0.2604	0.1666	0.1051	0.0979
8500	0.8873	0.7975	0.6592	0.4987	0.2605	0.1666	0.1051	0.0979
9000	0.8874	0.7976	0.6594	0.4987	0.2605	0.1666	0.1051	0.0979
9500	0.8875	0.7977	0.6596	0.4987	0.2606	0.1666	0.1051	0.0979
10000	0.8875	0.7978	0.6598	0.4987	0.2606	0.1667	0.1051	0.0979

Table ??.1. Average inaccuracies $\tau^I_{\alpha.n}$ for selected α and n.

3. Rényi informations and Bayesian errors

n	0.125	0.25	0.5	1	1.5	2	3	4
2	0.2235	0.2001	0.1617	0.1107	0.0798	0.0596	0.0352	0.0212
3	0.3524	0.3208	0.2659	0.2816	0.1433	0.1134	0.0782	0.0585
4	0.4320	0.3963	0.3319	0.3460	0.1826	0.1466	0.1049	0.0821
5	0.4863	0.4483	0.3776	0.3804	0.2087	0.1682	0.1221	0.0974
6	0.5260	0.4867	0.4114	0.4019	0.2270	0.1830	0.1337	0.1077
7	0.5566	0.5163	0.4376	0.4167	0.2404	0.1937	0.1419	0.1149
8	0.5809	0.5401	0.4587	0.4275	0.2507	0.2017	0.1479	0.1202
9	0.6008	0.5596	0.4761	0.4357	0.2587	0.2077	0.1523	0.1241
10	0.6175	0.5760	0.4908	0.4422	0.2651	0.2124	0.1557	0.1270
20	0.7047	0.6633	0.5698	0.4705	0.2923	0.2297	0.1662	0.1360
30	0.7416	0.7012	0.6051	0.4796	0.2992	0.2316	0.1656	0.1353
40	0.7631	0.7236	0.6266	0.4842	0.3012	0.2305	0.1634	0.1332
50	0.7776	0.7390	0.6416	0.4869	0.3015	0.2286	0.1608	0.1310
100	0.8134	0.7777	0.6810	0.9678	0.2975	0.2189	0.1506	0.1220
200	0.8393	0.8066	0.7126	0.9918	0.2888	0.2061	0.1388	0.1119
300	0.8513	0.8202	0.7284	0.4963	0.2826	0.1982	0.1320	0.1061
400	0.8588	0.8288	0.7387	0.4968	0.2778	0.1925	0.1273	0.1022
500	0.8640	0.8349	0.7462	0.4971	0.2740	0.1881	0.1237	0.0992
600	0.8680	0.8396	0.7521	0.4974	0.2708	0.1846	0.1209	0.0969
700	0.8712	0.8433	0.7568	0.4975	0.2681	0.1817	0.1186	0.0949
800	0.8739	0.8464	0.7608	0.4976	0.2658	0.1792	0.1166	0.0933
900	0.8761	0.8491	0.7643	0.4977	0.2637	0.1770	0.1149	0.0919
1000	0.8780	0.8514	0.7673	0.4978	0.2618	0.1750	0.1134	0.0907
1500	0.8849	0.8595	0.7782	0.4981	0.2546	0.1677	0.1079	0.0862
2000	0.8893	0.8648	0.7854	0.4982	0.2496	0.1628	0.1042	0.0832
2500	0.8925	0.8686	0.7908	0.4983	0.2457	0.1591	0.1015	0.0810
3000	0.8950	0.8716	0.7950	0.4984	0.2425	0.1561	0.0994	0.0792
3500	0.8970	0.8740	0.7984	0.4984	0.2399	0.1537	0.0977	0.0778
4000	0.8986	0.8760	0.8013	0.4985	0.2376	0.1516	0.0962	0.0766
4500	0.9001	0.8778	0.8038	0.4985	0.2356	0.1498	0.0949	0.0756
5000	0.9013	0.8793	0.8060	0.4985	0.2338	0.1483	0.0938	0.0747
5500	0.9024	0.8806	0.8079	0.4986	0.2322	0.1469	0.0929	0.0739
6000	0.9034	0.8818	0.8096	0.4986	0.2308	0.1456	0.0920	0.0732
6500	0.9043	0.8828	0.8112	0.4986	0.2294	0.1444	0.0912	0.0726
7000	0.9051	0.8838	0.8126	0.4986	0.2282	0.1434	0.0905	0.0720
8000	0.9065	0.8855	0.8152	0.4986	0.2260	0.1415	0.0892	0.0709
8500	0.9071	0.8863	0.8163	0.4987	0.2251	0.1407	0.0886	0.0705
9000	0.9077	0.8870	0.8174	0.4987	0.2241	0.1399	0.0881	0.0700
9500	0.9082	0.8877	0.8184	0.4987	0.2233	0.1392	0.0876	0.0696
10000	0.9087	0.8883	0.8193	0.4987	0.2225	0.1385	0.0871	0.0693

Table 3.1. Average inaccuracies $\tau^{II}_{\alpha.n}$ for selected α and n.

References

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.