Entropy, some new lower bounds

Flemming Topsøe *
University of Copenhagen

Abstract

We derive second order lower bounds for the entropy function expressed in terms of the index of coincidence. The constants found either explicitly or implicitly are best possible in a natural sense.

Keywords Entropy, index of coincidence, measure of roughness.

1 Background, introduction

We study probability distributions over the natural numbers. The set of all such distributions is denoted $M_+^1(\mathbb{N})$ and the set of $P \in M_+^1(\mathbb{N})$ which is supported by $\{1, 2, \cdots, n\}$ is denoted $M_+^1(n)$.

We use U_k to denote a generic uniform distribution over a k-element set, and if also U_{k+1} $(U_{k+2}, \cdots)$ is considered, it is assumed that the supports are increasing. By H and by IC we denote *entropy* and *index of coincidence*, respectively, i.e.

$$H(P) = -\sum_{k=1}^{\infty} p_k \ ln \, p_k \, ,$$

$$IC(P) = \sum_{k=1}^{\infty} p_k^2 \, .$$

In Harremoës and Topsøe [1] the exact range of the map $P \curvearrowright (IC(P), H(P))$ with P varying over either $M_+^1(n)$ or $M_+^1(\mathbb{N})$ was determined. The ranges in question, termed *IC/H-diagrams*, were denoted Δ, respectively Δ_n:

$$\Delta = \{(IC(P), H(P)) \mid P \in M_+^1(\mathbb{N})\} \, ,$$
$$\Delta_n = \{(IC(P), H(P)) \mid P \in M_+^1(n)\} \, .$$

*This work was supported by the Danish Natural Science Research Council and by INTAS, project 00-738.

By Jensen's inequality we find that $H(P) \geq -\ ln\, IC(P)$, thus the logarithmic curve $t \curvearrowright (t, -\ ln\, t)$; $0 < t \leq 1$ is a lower bounding curve for the IC/H-diagrams. Further, we note that the points $Q_k = \left(\frac{1}{k},\ ln\, k\right)$; $k \geq 1$ which correspond to the uniform distributions U_k lie on this curve. No other points in the diagram Δ lie on the curve, in fact, Q_k; $k \geq 1$ are extremal points of Δ in the sense that the convex hull they determine contains Δ. No smaller set has this property.

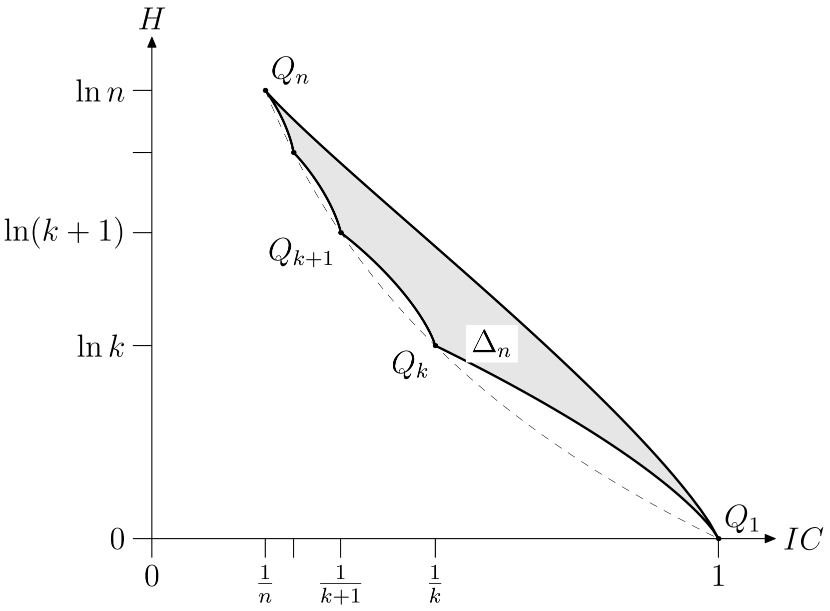

Fig.1. The restricted IC/H-diagram $\Delta_n (n = 5,\ k = 2)$ (from [1]).

Figure 1 illustrates the situation for the restricted diagrams Δ_n. The key result of [1] states that Δ_n is the bounded region determinated by a certain Jordan curve determined by n smooth arcs, the "upper arc" connecting Q_1 and Q_n and then $n-1$ "lower arcs" connecting Q_n with Q_{n-1}, Q_{n-1} with Q_{n-2} etc. until Q_2 which is connected with Q_1.

In [1] the main result was used to develop concrete upper bounds for the entropy function. Here our concern will be lower bounds. The study depends crucially on the nature of the lower arcs. In [1] these arcs were identified. Indeed, the arc connecting Q_{k+1} with Q_k is the curve which may be parametrized as follows:

$$s \curvearrowright \vec{\varphi}((1-s)U_{k+1} + s\, U_k)$$

with s running through the unit interval and with $\vec{\varphi}$ denoting the *IC/H-map* given by $\vec{\varphi}(P) = (IC(P), H(P))$; $P \in M_+^1(\mathbb{N})$.

The distributions in $M_+^1(\mathbb{N})$ fall in *IC-complexity classes*. The k'th class consists of all $P \in M_+^1(\mathbb{N})$ with $IC(U_{k+1}) < IC(P) \leq IC(U_k)$. In order to determine good lower bounds for the entropy of a distribution P, one first determines the IC-complexity class k of P. One may then determine that value of $s \in]0,1]$ for which $IC(P_s) = IC(P)$ with $P_s = (1-s)U_{k+1} + s\,U_k$. Then $H(P) \geq H(P_s)$ is the theoretically best lower bound of $H(P)$ in terms of $IC(P)$.

In order to write the sought lower bounds for $H(P)$ in a convenient form, we introduce the k'th *relative measure of roughness* by

$$\overline{MR}_k(P) = \frac{IC(P) - IC(U_{k+1})}{IC(U_k) - IC(U_{k+1})} = k(k+1)\left(IC(P) - \frac{1}{k+1}\right). \quad (1)$$

This definition applies to any $P \in M_+^1(\mathbb{N})$ but really, only distributions of IC-complexity class k will be of relevance to us. Clearly, $\overline{MR}_k(U_{k+1}) = 1$, $\overline{MR}_k(U_k) = 0$ and for any distribution of IC-complexity class k, $0 \leq \overline{MR}_k(P) \leq 1$. For a distribution on the lower arc connecting Q_{k+1} with Q_k one finds (cf. Lemma 1 of [1]) that

$$\overline{MR}_k((1-s)U_{k+1} + s\,U_k) = s^2\,. \quad (2)$$

In view of the above said, it follows that for any distribution P of IC-complexity class k, the theoretically best lower bound for $H(P)$ in terms of $IC(P)$ is given by the inequality

$$H(P) \geq H((1-x)U_{k+1} + x\,U_k) \quad (3)$$

where x is determined so that P and $(1-x)U_{k+1} + x\,U_k$ have the same index of coincidence, i.e.

$$x^2 = \overline{MR}_k(P)\,. \quad (4)$$

By writing out the right-hand-side of (3) we then obtain the best lower bound of the type discussed. Doing so one obtains a quantity of mixed type, involving logarithmic and rational functions. It is desirable to search for structurally simpler bounds, getting rid of logarithmic terms. The simplest and possibly most useful bound of this type is the linear bound

$$H(P) \geq H(U_k)\overline{MR}_k(P) + H(U_{k+1})(1 - \overline{MR}_k(P)) \quad (5)$$

which expresses the fact mentioned above regarding the extremal position of the points Q_k in relation to the set Δ. Note that (5) is the best linear lower bound as equality holds for $P = U_{k+1}$ as well as for $P = U_k$. Another comment is that though (5) was developed with a view to distributions of IC-complexity class k, the inequality holds for all $P \in M_+^1(\mathbb{N})$ (but is weaker than the trivial bond $H \geq -\ln IC$ for distributions of other IC-complexity classes).

Writing (5) directly in terms of $IC(P)$ we obtain the inequalities

$$H(P) \geq \alpha_k - \beta_k \, IC(P)\,;\ k \geq 1 \tag{6}$$

with the constants α_k and β_k given by

$$\alpha_k = \ln(k+1) + k \ln\left(1+\frac{1}{k}\right), \tag{7}$$
$$\beta_k = (k+1)k \ln\left(1+\frac{1}{k}\right).$$

In the present paper we shall develop sharper inequalitites than those above by adding a second order term. More precisely, for $k \geq 1$, we denote by γ_k the largest constant such that the inequality

$$H \geq \frac{1}{k}\overline{MR}_k + \frac{1}{k+1}(1-\overline{MR}_k) + \frac{\gamma_k}{2k}\overline{MR}_k(1-\overline{MR}_k) \tag{8}$$

holds for all $P \in M_+^1(\mathbb{N})$. Here, $H = H(P)$ and $\overline{MR}_k = \overline{MR}_k(P)$. Expressed directly in terms of $IC = IC(P)$, (8) states that

$$H \geq \alpha_k - \beta_k \, IC + \frac{\gamma_k}{2}k(k+1)^2\left(IC - \frac{1}{k+1}\right)\left(\frac{1}{k} - IC\right) \tag{9}$$

for $P \in M_+^1(\mathbb{N})$.

The basic results of our paper may be summarized as follows.

Theorem 1. *The constants $(\gamma_k)_{k\geq 1}$ are increasing with $\gamma_1 = \ln 4 - 1 \approx 0.3863$ and with limit value $\gamma \approx 0.9640$.*

More substance will be given to this result by developing rather narrow bounds for the γ_k's in terms of γ and by other means. In particular, we mention that γ is the minimum of the function

$$f(x) = \frac{2(-x - \ln(1-x))}{x^2(1+x)}\,;\ 0 < x < 1 \tag{10}$$

and that the inequalities

$$\left(2k \ln(1+\frac{1}{k}) - 1\right)\gamma \leq \gamma_k \leq \frac{k}{k+1}\left(\gamma - (2k+1)\ln(1+\frac{1}{k}) + 2\right). \tag{11}$$

k	γ_k	lower bound	upper bound
1	0.3863	0.3724	0.4423
2	0.6071	0.5995	0.6245
3	0.7039	0.7000	0.7127
4	0.7593	0.7569	0.7646
5	0.7952	0.7936	0.7987
6	0.8204	0.8192	0.8229
7	0.8390	0.8381	0.8409
8	0.8534	0.8527	0.8548
9	0.8648	0.8642	0.8659
10	0.8740	0.8736	0.8750
20	0.9175	0.9174	0.9177
50	0.9450	0.9450	0.9450
100	0.9544	0.9544	0.9544
10.000	0.9639	0.9639	0.9639
1.000.000	0.9640	0.9640	0.9640

Table 1:

hold.

This leeds to rather narrow bounds for the constants γ_k, cf. Table 1.

The motivation to develop the refined second order inequalities lies in applications to problems of exact prediction in Bernoulli models. The author plans to take this up in a separate publication. However, in view of the significance of entropy and of index of coincidence – closely related to the popular chi-square distance – it is found that the results are of interest in their own right. In addition, the tools applied, though simple in principle, result in some delicate estimates and involve relatively sophisticated classical techniques which may have a bearing on related studies. For instance, some auxiliary elementary inequalities for the logarithmic function in terms of rational functions may be of wider applicability.

Earlier related work includes Kovalevskij [2], Tebbe and Dwyer [3], Ben-Bassat [4], Golic [5], Feder and Merhav [6] and the already cited work by Harremoës and Topsøe [1].

References

[1] P. Harremoës and F. Topsøe, "Inequalities between entropy and index of coincidence derived from information diagrams," *IEEE trans.Inform.*

Theory, vol. 47, pp. 2944–2960, Nov. 2001. Accepted for publication in IEEE Trans. Inform. Theory.

[2] V. A. Kovalevskij, *The problem of character recognition from the point of view of mathematical statistics*, pp. 3–30. New York: Spartan, 1967.

[3] D. Tebbe and S. Dwyer, "Uncertainty and the probability of error," *IEEE Trans. Inform. Theory*, vol. 14, pp. 516–518, 1968.

[4] M. Ben-Bassat, "f-entropies, probability of error, and feature selection," *Information and Control*, vol. 39, pp. 227–242, 1978.

[5] J. D. Golic, "On the relationship between the information measures and the bayes probability of error," *EEE Trans. Inform. Theory*, vol. 33, pp. 681–693, 1987.

[6] M. Feder and N. Merhav, "Relations between entropy and error probability," *IEEE Trans. Inform. Theory*, vol. 40, pp. 259–266, 1994.