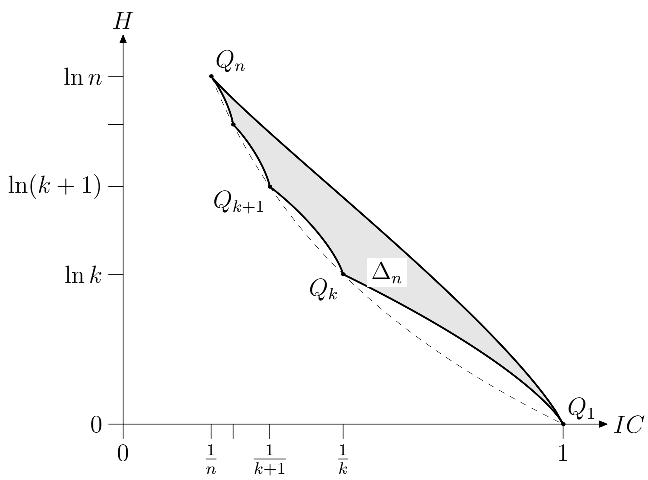
Entropy, some new lower bounds

Flemming Topsøe[1]
Department of Mathematics
University of Copenhagen
Denmark.
e-mail: topsoe@math.ku.dk

Let $M_+^1(\mathbb{N})$ be the set of probability distributions over $\mathbb{N}$ and denote by U_k a generic uniform distribution over a k-element set. By H and IC we denote, respectively, *entropy* and *index of coincidence*, i.e. $H(P) = -\sum p_k \ln p_k$ and $IC(P) = \sum p_k^2$.



In [4] the range of $P \curvearrowright (IC(P), H(P))$ with $P \in M_+^1(\mathbb{N})$ was determined. The resulting *information diagrams* Δ and Δ_n (with Δ_n obtained by restriction to distributions P with $p_k = 0$ for $k > n$) are indicated in the figure, copied from [4], which shows Δ_n for $n = 5$.

By Jensen's inequality, $H(P) \geq -\ln IC(P)$, thus $t \curvearrowright (t, -\ln t)$ is a lower bounding curve for the IC/H-diagrams (the dashed curve in the figure). The points $Q_k = \left(\frac{1}{k}, \ln k\right)$ which correspond to the U_k's lie on this curve. No other points in Δ does so, in fact, the Q_k's are extremal points of Δ.

In [4], the Jordan curve determining Δ_n was determined. The "lower arcs" connect, for each k, the point Q_{k+1} with Q_k and correspond to the mixtures $(1-s)U_{k+1} + s\,U_k$. By convexity of these arcs one obtains the linear inequalities $H(P) \geq \alpha_k - \beta_k\, IC(P)$ where

$$\alpha_k = \ln(k+1) + u_k\ ,\ \beta_k = (k+1)u_k$$

with $u_k = k\ln(1 + \frac{1}{k})$.

We shall add a quadratic term to this inequality. By definition, γ_k is the largest constant such that, for any $P \in M_+^1(\mathbb{N})$,

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

The basic results 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 is added to this result by these facts:

$$\gamma = \min_{0<x<1} \frac{2(-x - \ln(1-x))}{x^2(1+x)} = \min_{0<x<1} \frac{2}{1+x} \sum_{n=0}^{\infty} \frac{x^n}{n+2}\,,$$

$$(2u_k - 1)\gamma \leq \gamma_k \leq \frac{k}{k+1}(\gamma + 2) - \frac{2k+1}{k+1} u_k\,.$$

This leeds to quite narrow bounds for the γ_k's (as $u_k \uparrow 1$).

The motivation to develop lower bounds as here presented lies in applications to certain problems of exact prediction in Bernoulli models.

On the technical side we point to the socalled *lemma of replacement* from [4] and to certain elementary more specific auxiliary inequalities which compares the logarithmic function with certain rational functions. These results, especially the former, are believed to be of independent interest.

Earlier related work starts with Kovalevskij [5] and continues with [6], [1], [3], [2] and [4].

References

[1] M. Ben-Bassat, "f-Entropies, Probability of Error, and Feature Selection," *Information and Control*, vol. 39, pp. 227–242, 1978.

[2] M. Feder and N. Merhav, "Relations Between Entropy and Error Probability," *IEEE Trans. Inform. Theory*, vol. 40, pp. 259–266, 1994.

[3] J. Dj. Golic, "On the relationship between the information measures and the Bayes probability of error," *IEEE Trans. Inform. Theory*, vo. 33, pp. 681–693, 1987.

[4] Peter Harremoës and Flemming Topsøe, "Inequalities between Entropy and Index of Coincidence derived from Information Diagrams," *IEEE Trans. Inform. Theory*, vol. 27, pp. 2944–2960, 2001.

[5] V. A. Kovalevskij, "The problem of character recognition from the point of view of mathematical statistics," in *Character Readers and Pattern Recognition*, V. A. Kovalevskij (ed.), pp. 3–30, Spartan, New York, 1967 (russian original 1965).

[6] D. L. Tebbe and S. J. Dwyer, "Uncertainty and the Probability of Error," *IEEE Trans. Inform. Theory*, vol. 14, pp. 516–518, 1968.

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