Information Diagrams: Entropy, Index of Coincidence and Probability of Error

Peter Harremoës
Aurehøj Amtsgymnasium,
Gentofte, Denmark
e-mail:moes@post7.tele.dk

Flemming Topsøe
Department of Mathematics,
University of Copenhagen
e-mail: topsoe@math.ku.dk

Abstract — **The range of the map $P \curvearrowright (IC(P), H(P))$ is determined. Here, P denotes a distribution, $IC(P)$ its index of coincidence and $H(P)$ its entropy.**

Let $M_+^1(n)$ be the set of probability distributions over an n-letter alphabet. Uniform distributions over k-subsets are denoted U_k. We consider *entropy* $H(P)$ and *divergence* $D(P\|Q)$ as well as the *index of coincidence,* $IC(P) = \sum p_i^2$ (known from cryptoanalysis). The *measure of roughness* is $MR_n(P) = IC(P) - 1/n$ and the *relative measure of roughness* is

$$\overline{MR}^n(P) = \frac{MR_n(P)}{MR_n(U_1)} = \frac{IC(P) - \frac{1}{n}}{1 - \frac{1}{n}}.$$

Qualitatively, $1 - \overline{MR}^n(P)$ behaves like a kind of entropy.

We investigate the relationship between $D(P\|U_n)$ and $MR_n(P)$. As $MR_n(P) = IC(P) - \frac{1}{n}$ and $D(P\|U_n) = \ln n - H(P)$, we have decided to work mainly with the map $\vec{\varphi} : P \curvearrowright (IC(P), H(P))$. The range of this map, which we call the *IC/H-diagram*, is denoted by Δ_n.

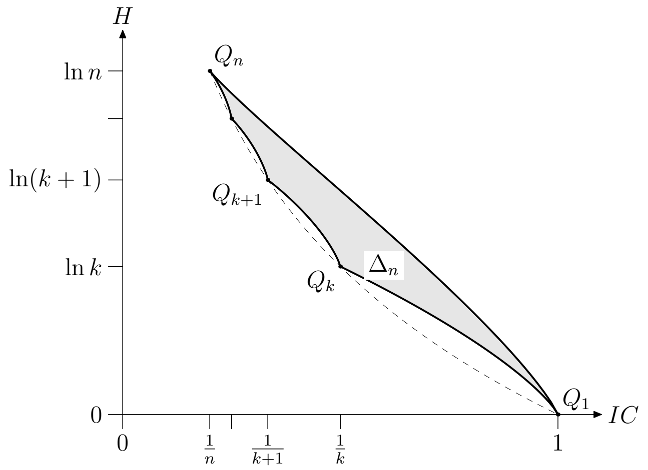

Fig. 1: The IC/H-diagram $\Delta_n (n = 5,\ k = 2)$

The *IC/H-diagram*, shown above (for $n = 5, k = 2$), contains the points Q_k corresponding to uniform distributions. These points all lie on the smooth curve $y = -\ln x,\ 0 < x \le 1$. The arcs joining the points are denoted $\frown Q_nQ_{n-1}, \cdots, \frown Q_2Q_1$ and then $\frown Q_1Q_n$ for the "upper arc". The arc $\frown Q_{k+1}Q_k$ has the parametrization $s \curvearrowright \vec{\varphi}\left((1-s)U_{k+1} + sU_k\right),\quad 0 \le s \le 1$. The curve $J_n = \frown Q_nQ_{n-1} + \frown Q_{n-1}Q_{n-2} + \cdots + \frown Q_2Q_1 + \frown Q_1Q_n$ plays a key role. Main results are:

Theorem 1. *For $n \ge 3$, J_n is a positively oriented Jordan curve in the plane, and the bounded region which it determines (including J_n itself) coincides with the IC/H-diagram Δ_n.*

[1]Research supported by grants from the COWI Foundation (FT) and from the Danish Natural Science Research Council (PH+FT).

A novelty is the proof which involves topological methods.

Theorem 2. *For a discrete distribution P and any $k \ge 1$,*

$$H(P) \ge \alpha_k - \beta_k IC(P)$$

with α_k and β_k defined via the constants $e_k = \left(1 + k^{-1}\right)^k$ by $\alpha_k = \ln(k+1) + \ln e_k\,, \beta_k = (k+1)\ln e_k$.

Theorem 3. *There exists an increasing sequence $(\gamma_n)_{n\ge 2}$ of constants with $\gamma_2 = (2\ln 2)^{-1} \approx 0.7213$ and $\lim_{n\to\infty} \gamma_n = 1$ such that the inequalities*

$$H(P) \le \ln n \cdot \left(1 - \overline{MR}^n(P)\right)^{\gamma_n} \le \ln n \left(1 - \gamma_n \overline{MR}^n(P)\right)$$

hold for $n \ge 2$ and all $P \in M_+^1(n)$. For divergence and χ^2-distance this implies that

$$D(P\|U_n) \ge \frac{\gamma_n \ln n}{n-1} \chi^2(P, U_n).$$

The inequalities in Theorem 2 have direct applications to prediction in Bernoulli sources (to be published) and to rate distortion theory, cf. György and Linder, [1].

An equivalent form of Theorem 1 is obtained by replacing $IC(P)$ by the *Rényi entropy* $H_2(P)$ of *order* 2 given by $H_2(P) = -\ln IC(P)$. Then one obtains the the H_2/H– *diagram* (not shown here).

Generalizations include the consideration of other powers than 2. This leads to diagrams involving general Rényi entropies and, as a limiting case, the error of probability.

Related diagrams were first discovered by Kovalevskij [2], but later, independently, taken up by several others (Tebbe and Dwyer, Ben-Bassat, Feder and Merhav) and, very recently, by György and Linder, [1].

A full discussion of the above results covers other diagrams, universality of the constants of Theorem 2, the role of $IC(P)$ and natural extensions, scope of the topological method, study of other types of divergences, etc.

Acknowledgments

The authors have had helpful discussions with Boris Ryabko, Igor Vajda, Rasmus B. Hansen and András György.

References

[1] A. György and T. Linder, "Optimal Entropy-Constrained Scalar Quantization of a Uniform Source," *IEEE Trans. Inform. Theory*, vol. 46, pp. 2704–2711, 2000.

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

[3] P. Harremoës and F. Topsøe, "Inequalities between Entropy and Index of Coincidence derived from Information Diagrams," (submitted for publication).