Entropy in Models for Estimation

J Rissanen, T, Roos, P. Myllymaki

Helsinki Institute for Information Technology

In this talk I'll present recent and new ideas in estimation, which is regarded as a main topic in a theory of statistics. No one in statistics knows how to obtain an optimal way to estimate the distribution as a *model* to explain the statistical regular features in a string $x = x^n = x_1, \ldots, x_n$, where the numbers x_i are taken as real. This means that there is no theory in statistics to be taken seriously. After all, if we do not know how to handle this serious problem how can we consider the more important problems of estimating the random behavior in statistics.

We begin with Shannon's notion of entropy of a distribution $P = \{P(i),\, i = 1, 2, \ldots, m\}$, which is

$$H(P) = \sum P(i) \log 1/P(i),$$

which lies on the foundation of information theory. As stated it has no real application in estimation nor statistics, which concerns a set of parametric models $M = \{f(x; \theta_i)\}$, where θ_i denotes a k-dimensional parameter of some number k, also to be estimated. The obvious reason is that no single distribution P is in the family.

We redefine entropy as the length of the shortest mean code for x, which in Shannon's case is the same entropy $H(P)$ by the beautiful theorem of MacMillan and Doob, and nothing is lost of Shannon's information theory.