*Journal of Statistical Physics, Vol. 11, No. 6, 1974*

# Letter to the Editor

# A Comment on the Paper "A Comparison of the Shannon and Kullback Information Measures"[1]

Michael Batty[2]

*Received August 7, 1974*

The article by Hobson and Cheng, in which they derive Shannon's measure for uncertainty from Kullback's "information for discrimination" statistic demonstrates that the field of information theory is rich in different interpretations. In this spirit, readers may be interested in a related but somewhat oblique comparison between the Shannon and Kullback formulas. First, consider Shannon's formula for continuous entropy $S(x)$

$$S(x) = -\int \rho(x) \ln \rho(x)\, dx \tag{1}$$

[1] The paper "A Comparison of the Shannon and Kullback Information Measures" by Arthur Hobson and Bin-Kang Cheng appeared in *J. Stat. Phys.* 7(4):301 (1973).

[2] Department of Geography, University of Reading, Whiteknights, Reading, England.

where $\rho(x)$ is the probability density function. A discrete approximation to Eq. (1) can be derived as

$$S = -\sum p_i \ln (p_i/\Delta x_i) \tag{2}$$

where $\Delta x_i$ represents the interval width over which the probability $p_i$ is measured.[1,2] That

$$S(x) = \lim_{\Delta x_i \to 0} S$$

can be easily demonstrated by noting that $p_i = p(x_i)\Delta x_i$, substituting for $p_i$ in Eq. (2), and passing to the limit. Equation (2) represents a modification of the formula for discrete Shannon entropy in which the uncertainty has been renormalized by "throwing out an infinite contribution."

As Hobson and Cheng point out, Eqs. (1) and (2) are dimensionally incorrect. One way of accounting for this is to deduct $S$ from its maximum value, thus forming a relative index. The maximum value of Eq. (2) is $\ln X$ where $X = \sum \Delta x_i$ as can be checked by maximizing (2) subject to its appropriate constraints. Then, the new statistic is given as

$$I = \ln X - S \tag{3}$$

and it is clear that (3) can be rewritten as

$$\begin{aligned} I &= \sum p_i \ln \frac{p_i}{\Delta x_i/X} \\ &= \sum p_i \ln \frac{p_i}{p_i^0} \end{aligned} \tag{4}$$

The term $p_i^0$ is now the prior probability based on the interval size which implies a uniform prior probability distribution. Equation (4) is obviously a special case of Kullback's statistic; it is particularly useful in problems where the form of the coordinate system on which probabilities are measured is of prime interest. For example, it can lead to new insights in the analysis of geographical and other spatial relationships using information theory, in contrast to previous work where the form of the system has always been implicit.[1,3]

## REFERENCES

1. M. Batty, *Geographical Analysis* **6**:1 (1974).
2. E. T. Jaynes, in *Statistical Physics*, K. W. Ford, ed., Benjamin, New York (1963).
3. A. G. Wilson, *Entropy in Urban and Regional Modelling*, Pion, London (1970).