Journal of Statistical Physics, Vol. 13, No. 3, 1975

Note on the Paper "Gibbs vs. Shannon Entropies" by Richard L. Liboff

Helmut Hirsch[1]

Received April 17, 1975

In addition to the comparison of the Gibbs negentropy $\eta_G = \int \Pi \ln \Pi$ to the Shannon negentropy $\eta_S = \sum E_i \Pi_i \ln(E_i \Pi_i)$, discussed in the cited paper, a comparison of the former to $\eta_S' = \eta_S - \sum E_i \ln E_i$, the decrease in uncertainty, is of general interest from an information-theoretic standpoint. It is shown that η_G and η_S' are similar in the situation in question. The additivity properties of η_G are briefly discussed.

KEY WORDS: Information gain (directed divergence); decrease in uncertainty; additivity of measures of information.

In the paper in question,[1] the Gibbs negentropy

$$\eta_G = \int_{\mathbf{E}} \Pi \ln \Pi \, d\mathbf{E} = \sum \Pi_i E_i \ln \Pi_i \tag{1}$$

where $\mathbf{E}$ is the energy shell, Π the density of a coarse-grained distribution, and E_i the measure of the ith cell of $\mathbf{E}$ (whose measure is normalized to one), is compared with the Shannon negentropy

$$\eta_S = \sum \Pi_i E_i \ln(\Pi_i E_i) \tag{2}$$

Interesting results are given for various fine-grained distributions D for successively refined partitions. It is pointed out that η_G and η_S show a rather

[1] I. Physikalisches Institut der Universität Wien, Vienna, Austria.

dissimilar behavior, the former being consistently a measure of the information contained in the coarse-grained distribution Π, the latter the negative uncertainty related to an experiment on the state of the system.

However, it appears to the author that a comparison of the two concepts should be based to a larger extent on their probability-theoretic significance. Equation (1) may readily be rewritten as

$$\eta_G = \sum \Pi_i E_i \ln(\Pi_i E_i / E_i) \tag{3}$$

thus revealing its nature more clearly. η_G is the information gain yielded by taking the probability distribution $\{\Pi_i E_i\}$ instead of the distribution $\{E_i\}$,[2] also called the directed divergence.[3]

A different explicatum for the information gain in this substitution of distributions, involving the Shannon entropy, is

$$\eta_S' = -\sum E_i \ln E_i + \sum \Pi_i E_i \ln(\Pi_i E_i) \tag{4}$$

the decrease in uncertainty resulting from the substitution.

In information theory, both η_S' and η_G have been considered as measures of amount of information. The choice between them is complicated by the fact that both have a serious shortcoming: η_S' is capable of assuming negative values as well as positive ones, which in many cases may appear unacceptable; η_G is not additive for successive changes of the probability distribution, i.e., we have

$$\int f_2 \ln(f_2/f_1)\, dx + \int f_3 \ln(f_3/f_2)\, dx \neq \int f_3 \ln(f_3/f_1)\, dx \tag{5}$$

where f_1, f_2, and f_3 are arbitrary, not almost everywhere identical probability density functions. It has to be decided in each separate case which is the more suitable.

Considering this, it is clear that a comparison of η_G and η_S', rather than η_G and η_S, is of more general interest.

The prior distribution $\{E_i\}$ can be interpreted as representing a situation in which no information about D is available at all; in this case $\Pi = 1$ everywhere, as the measure of $\mathbf{E}$ is one. A straightforward calculation now yields

$$\Delta = \eta_S' - \eta_G = \sum (\Pi_i - 1) E_i \ln E_i \tag{6}$$

If all elements of the partition have equal measure, $\Delta = 0$, as the term $\ln E_i$ may then be taken out of the sum on the right-hand side of (6) and the expectation of Π clearly is one. In this case a plot of η_S' vs. η_G would simply show a straight line at an angle of 45° for every D. If the partition consists of unequal elements, it is easy to show that both negative and positive values for Δ are possible. In the limit, the fine-grained Gibbs negentropy

$$\eta_D = \int_{\mathbf{E}} D \ln D \, d\mathbf{E} \tag{7}$$

is equal to the one-dimensional Shannon negentropy of the distribution D. This is due to the fact that, again interpreting η_D as information gain in the change from uniform prior distribution to D, the density of the prior distribution is one everywhere and thus does not appear explicitly in (7). If the underlying measure space had an infinite measure (such as, e.g., the Lebesgue measure on the Borel sets of real numbers), the one-dimensional Shannon negentropy cannot be interpreted as information gain, as the Lebesgue measure (being the analog of a uniform distribution in this case) is not a probability measure.

Another interpretation is possible for η_G as defined in (1). Consider a sequence of successively refined partitions. Let $\{\Pi_i^n\}$ denote the density of the nth coarse-grained distribution and E_i^n the measure of the ith element of the corresponding partition. Then

$$\begin{aligned} &\sum \Pi_i^n E_i^n \ln(\Pi_i^n E_i^n / \Pi_i^{n-1} E_i^n) + \sum \Pi_i^{n-1} E_i^n \ln(\Pi_i^{n-1} E_i^n / \Pi_i^{n-2} E_i^n) \\ &\quad + \cdots + \sum \Pi_i^1 E_i^n \ln(\Pi_i^1 E_i^n / E_i^n) = \sum \Pi_i^n E_i^n \ln(\Pi_i^n) = \eta_G \end{aligned} \tag{8}$$

where the sum extends over the same number of summands in each term (Π_i^j taking the same values several times for $j < n$). Thus, η_G may also be interpreted as the sum of all information gains successively obtained in the stepwise changes to finer distributions; in other words, η_G is additive in our case (although this property does not hold in general, as was pointed out above). The additivity of η_S' is trivial.

REFERENCES

1. R. L. Liboff, Gibbs vs. Shannon Entropies, *J. Stat. Phys.* **11**(4):343 (1974).
2. A. Rényi, *Wahrscheinlichkeitsrechnung mit einem Anhang über Informationstheorie*, VEB Deutscher Verlag der Wissenschaften, Berlin (1966).
3. S. Kullback, *Information Theory and Statistics*, Dover, New York (1968).