

Hugh Everett, III
Institute for Defense Analyses
The Pentagon
Washington, D. C.
June 11, 1957

Dr. E. T. Jaynes
Department of Physics
Stanford University
Stanford, California

Dear Dr. Jaynes:

I am writing with respect to your article, "Information Theory and Statistical Mechanics" in the May 15 Physical Review.

While I fully sympathize with the "subjectivist" view of statistical mechanics that you express, I must point out a rather fundamental and inescapable difficulty with the principle of maximum entropy as you have stated it. It occurred to me also several years ago that one might take such an approach-- that one might be able to circumvent the reliance on dynamical laws by basing deductions on a minimum information principle, with the subjectivist interpretation of probabilities as the justification. This is indeed an appealing idea.

The difficulty is that one is seduced by this method into believing, as you apparently do and I did, that one has circumvented the problem of assigning a priori probabilities when one has in fact done no such thing, but has tacitly admitted a particular a priori distribution.

Briefly, the trouble lies in using the expression $\sum p_i \ln p_i$ to measure information. I shall demonstrate shortly that this choice is highly prejudicial and is equivalent to merely assuming equal a priori probabilities for each state.

It has occurred independently to several people, myself included, that the proper definition of information is a relative one, the information of a probability distribution relative to some underlying (basic) measure (or probability distribution) already given.

Dr. E. T. Jaynes
 June 11, 1957
 Page 2

Thus for a discrete set of states $\{S_i\}$, with basic measure $\{\mu_i\}$ (simply a set of weights in the discrete case), the relative information of a distribution $\{P_i\}$ over the states is defined as:

$$\text{Def } I = \sum_i P_i \ln \frac{P_i}{\mu_i}$$

The reasons for this more general definition will become clear shortly. First, it enables one to define information for any probability distribution, i.e., for arbitrary probability measures over completely arbitrary sets of unrestricted cardinality. (Judging by your footnote you are aware of the difficulties of even the continuous case for the ordinary definition.) This general definition comes about as follows: Consider an arbitrary set X , with probability measure P and underlying (I call it an information measure) measure μ . Now consider a finite partition \mathcal{P} of X into subsets X_i . We then have a finite distribution over these sets and an information (in the relative sense) defined:

$$\text{Def } I^{\mathcal{P}} = \sum_i P(X_i) \ln \frac{P(X_i)}{\mu(X_i)}$$

Now it is an easily proved theorem that any refinement of \mathcal{P} will never decrease the information (i.e., \mathcal{P}' is a refinement of \mathcal{P} implies that $I^{\mathcal{P}'} \geq I^{\mathcal{P}}$). Hence $I^{\mathcal{P}}$ is a monotone function on the directed set of all finite partitions, and always has a limit, which we define as the information of P relative to μ . Thus this relative definition generalizes, while the usual one doesn't, as you know.

A second advantage lies in the application to stochastic processes. If one defines the information of a distribution relative to a stationary distribution of the process, one can prove the theorem that information never increases with time (entropy never decreases).

Example: Two state Markov process with transition probabilities $T_{i,j}$:

$$\begin{matrix} & & 1 & 2 \\ & & \begin{pmatrix} 1/3 & 2/3 \\ 1/3 & 2/3 \end{pmatrix} \\ 1 & & & \\ 2 & & & \end{matrix}$$

Dr. E. T. Jaynes
June 11, 1957
Page 3

has the stationary distribution $P_1^* = \frac{1}{3}$, $P_2^* = \frac{2}{3}$. For this process the information of a distribution P_1, P_2 should be:

$$I = \sum P_i \ln \frac{P_i}{P_i^*} = P_1 \ln \frac{P_1}{1/3} + P_2 \ln \frac{P_2}{2/3}$$

It is only for doubly-stochastic processes (where $\sum_i T_{ij} = 1$ as well as $\sum_j T_{ij} = 1$) that the stationary measure is uniform, and one can get away with the old definition $\sum P_i \ln P_i$.

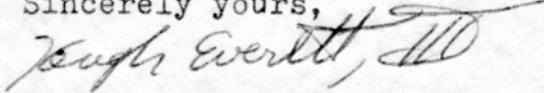
But notice that to determine the stationary measure, one must know the dynamics of the system. If you try to make predictions about this example using a minimum $\sum P_i \ln P_i$ you will make worse predictions than I, who use $\sum P_i \ln P_i / P_i^*$, since I take into account the known fact that this system is not equally likely to be in any of its states.

Similarly in the case of statistical mechanics of gases. Only after one has established that the measure one is using is stationary (Liouville's Theorem) is one justified in using it. The central problem is, as always, discovering the basic measure (or a priori probabilities, if you will). It is just fortuitous 1) that Lebesgue measure is the proper one for phase space so that $\int \rho \ln \rho d\omega$, which is really information relative to Lebesgue measure, is correct, and 2) that for doubly-stochastic processes, the type almost always encountered in physics, the uniform measure is stationary and hence $\sum P_i \ln P_i$ correct. These two circumstances are, I believe, what cause people to be seduced into believing a special case can be regarded as a general principle.

I really have a lot more to say on the subject, but time doesn't permit. I hope this rather hastily written letter conveys adequately the nature of my objection that you have not really sidestepped the fundamental problem of assigning a priori probabilities--which does depend on dynamical laws, ergodic properties, etc.,--but have only camouflaged it.

Nevertheless, I sympathize with your viewpoint, and believe a lot can be done in this line. I hope you will continue, and that my remarks on the more general definition of information may be of help to you.

Sincerely yours,


Hugh Everett, III

Dr. E. T. Jaynes
June 11, 1957
Page 4

P.S. Re the axioms which "uniquely" determine the definition of information (or entropy):

1. This general definition satisfies continuity in p_i 's.
2. It satisfies composition law, when basic measure composed also same way as probabilities.
3. It satisfies condition that $\min I$ is monotone decreasing with the number of allowed states (increasing uncertainty). This is analogous, and more natural, than the requirement $I(\frac{1}{n}, \dots, \frac{1}{n})$ is monotone decreasing.

It is, then, on quite as firm ground as the more restricted form, and your statement, "Therefore one expects that deductions made from any other information measure, if carried far enough, will eventually lead to contradictions," is a bit too strong.

HE:ne