Journal of Statistical Physics, Vol. 3, No. 4, 1971

Jaynes's Maximum Entropy Prescription and Probability Theory

Kenneth Friedman[1] and Abner Shimony[2]

Received March 9, 1971

Jaynes's prescription of maximizing the information-theoretic entropy is applied in a special situation to determine a certain set of posterior probabilities (when evidence fixing the expected value of a dynamical variable is given) and also the corresponding set of prior probabilities (when this evidence is not given). It is shown that the resulting values of these probabilities are inconsistent with the principles of probability theory. Three possible ways of avoiding this inconsistency are briefly discussed.

KEY WORDS: Information theory; probability; entropy; Jaynes.

1. In a series of publications[(1)] beginning in 1957, Jaynes has developed a new formulation of statistical mechanics, based upon two main theses. (I) The concept of probability in statistical mechanics is best understood not in the sense of relative frequency (which is the common interpretation), but in the sense of reasonable degree of belief[3] (which is the central concept in the probability theories of Laplace, Keynes, Jeffreys, and Carnap).[(2)] (II) The classic difficulty in theories of reasonable degree of belief, namely the problem of specifying probabilities when little information is available, can be resolved unambiguously by using the prescription of maximizing the information-theoretic entropy subject to constraints imposed by the available information.

Jaynes's program is controversial, but it has recently received some approbation,[4]

[1] Department of Philosophy, State University of New York, Oswego, New York.

[2] Departments of Philosophy and Physics, Boston University, Boston, Massachusetts.

[3] Note that Jaynes calls the concept under discussion "degree of reasonable belief" (Ref. 1a, p. 208), thereby following Keynes's terminology, "degree of rational belief" (Ref. 2b, p. 4). Their terminology is misleading, since presumably a rational man seeks to be as reasonable as possible in holding appropriate partial degree of belief. We follow the more accurate terminology of Jeffreys (Ref. 2c, p. 15) and Carnap.[(2e)]

[4] Several important recent expositions[(4)] of statistical mechanics are in close agreement with Jaynes's point of view.

such as the following[3]: "This formulation... has not been widely adopted, probably because it relinquishes the goal of exhibiting *thermodynamics* as a branch of mechanics. . . . On the other hand, Jaynes's approach does possess the merit of being self-contained, intuitively pleasing, deductively simple, and, by its very nature, minimally biased."

The purpose of this note is to question the adequacy of Jaynes's program on grounds quite different from those previously adduced. We have found, on a straightforward reading of his proposals, that there is an inconsistency between thesis I and thesis II. Although we can envisage the possibility of construing his proposals so as to avoid outright inconsistency, to do so would surely require thoroughgoing conceptual clarifications.

2. We shall use the notation $P(h \mid e) = r$ to signify "The reasonable degree of belief in proposition h, if the total body of evidence is e, is (the real number) r." In actual situations, there is usually a body of background information b which can be readily distinguished from evidence gathered in the course of experimentation. We shall call $P(h \mid b)$ the "prior probability of h," though it evidently is relative to b. If d is additional evidence, we shall call $P(h \mid b \,\&\, d)$ the "posterior probability of h" and let the evidential proposition be understood from context.

Now suppose that $\{d_i\}$ is a set of N propositions which are mutually exclusive and exhaustive given b, i.e., b logically implies $d_1 \vee \cdots \vee d_N$ and also $\sim(d_i \,\&\, d_j)$ for $i \neq j$. Then in all standard formulations of the theory of reasonable degree of belief, it is an elementary result that

$$P(h \mid b) = \sum_{i=1}^{N} P(h \mid b \,\&\, d_i)\, P(d_i \mid b) \tag{1}$$

Furthermore, if the theory is suitably extended to deal with continua, then an analogous theorem can be stated about a nondenumerable family of propositions $\{d_\beta\}$, with a real number index β, which are mutually exclusive and exhaustive given b, namely (e.g., Ref. 5)

$$P(h \mid b) = \int_{-\infty}^{\infty} P(h \mid b \,\&\, d_\beta)\, dF(d_\beta \mid b) \tag{2}$$

Here, $F(d_\beta \mid b)$ is the probability (in the sense of reasonable degree of belief), given b as the total body of evidence, that a member of the family with index equal to or less than β is true. Thesis I implies that Jaynes accepts Eq. (1), and his employment of the probability concept on continua indicates that Eq. (2) is also acceptable to him.

Now we shall consider the implications of Jaynes's thesis II in a somewhat special case. Suppose a system of interest has n distinct states, and let h_i be the hypothesis that it is in the ith state. If b contains no information about the system other than its structure (which determines the set of its possible states), then the prior probabilities $p_i \equiv P(h_i \mid b)$ are uniquely determined by Jaynes's prescription: maximize $-\sum_{i=1}^{n} p_i \ln p_i$ with no constraints on the p_i other than $\sum_{i=1}^{n} p_i = 1$. The result is

$$P(h_i \mid b) = 1/n, \qquad i = 1,\ldots, n \tag{3}$$

Suppose also that there is a dynamical variable E which has the value E_i in the ith state, and suppose there is a state, the mth, such that $E_m = \sum_{i=1}^{n} E_i/n$. Let $\hat{d}_\epsilon$ be evidence that the posterior expected value of E is ϵ, i.e.,

$$\sum_{i=1}^{n} E_i P(h_i \mid b \ \& \ \hat{d}_\epsilon) = \epsilon \tag{4}$$

Maximizing $-\sum_{i=1}^{n} p_i' \ln p_i'$, where $p_i' \equiv P(h_i \mid b \ \& \ \hat{d}_\epsilon)$, with $\sum_{i=1}^{n} p_i' = 1$ and Eq. (4) as constraints, yields

$$P(h_i \mid b \ \& \ \hat{d}_\epsilon) = Z^{-1} e^{-\beta E_i} \tag{5}$$

where β is a monotonically decreasing function of ϵ and $Z = \sum_{i=1}^{n} e^{-\beta E_i}$. For later convenience, we shall redesignate the proposition $\hat{d}_\epsilon$ by d_β. Then,

$$P(h_i \mid b \ \& \ d_\beta) = [e^{\beta(E_i - E_1)} + \cdots + e^{\beta(E_i - E_n)}]^{-1} \tag{6}$$

The background information b does not in general imply a definite value of ϵ, but one expects that in a theory of reasonable degree of belief, it is meaningful to consider the probability $F(\hat{d}_\epsilon \mid b)$. This is the prior probability that the evidence at a certain stage of experimentation will imply a posterior expected value of E equal to or less than ϵ. By the foregoing redesignation, $F(\hat{d}_\epsilon \mid b) = 1 - F(d_\beta \mid b)$. Now, if h_i is substituted for h in Eq. (2), and Eqs. (3) and (6) are used to express the prior and posterior probabilities of h_i, we obtain

$$1/n = \int_{-\infty}^{\infty} [e^{\beta(E_i - E_1)} + \cdots + e^{\beta(E_i - E_n)}]^{-1} \, dF(d_\beta \mid b) \tag{7}$$

for $i = 1,..., n$. One can easily check that for $i = m$, the integrand of Eq. (7) has maximum value $1/n$ at $\beta = 0$ and is less than $1/n$ for all other values of β. Hence, for this value of i, Eq. (7) can be satisfied only if $F(d_\beta \mid b)$ is the stepfunction which is 0 for $\beta < 0$ and 1 for $\beta \geqslant 0$, or less formally, only if

$$P(d_\beta \mid b) = \delta(\beta) \tag{8}$$

which in turn implies

$$P(\hat{d}_\epsilon \mid b) = \delta(\epsilon - E_m) \tag{9}$$

In other words, in this situation, a necessary condition for the consistency of theses I and II is the inferrability with certainty from b that evidence will be forthcoming which will fix the posterior expected value of E to be E_m, the same as the prior expected value. Since one of the primary motivations of Jaynes's program is to find a probability assignment which "honestly describes what we know" (Ref. 1c, p. 186), he surely would not find this necessary condition acceptable.

3. Three possible ways of escaping the inconsistency have occurred to us.

(a) The first is to deny that the proposition $\hat{d}_\epsilon$ can be well-defined. Indeed, it is hard to know what evidence would enable one to infer with certainty that the posterior expected value of E is ϵ. Clearly, any finite sample of replicas of the system of interest in similar environments would provide at best a distribution over ϵ having finite dispersion. Perhaps the requisite evidence would involve an infinite sample. Whatever the answer may be, Jaynes could hardly adopt the position that $\hat{d}_\epsilon$ is incapable of being well-defined, since imposing the expected value of E as a constraint is essential to his prescription, and presumably the knowledge of this expected value must be empirical.

(b) The second is to deny that the probabilities $F(\hat{d}_\epsilon \mid b)$ are capable of being well-defined, even though each $\hat{d}_\epsilon$ is well-defined. A defense along these lines seems promising to us. However, to make it convincing, one would need criteria for deciding when a proposition can and when it cannot be assigned a reasonable degree of belief on given evidence, which in turn presuppose a deep and systematic analysis of the concept of reasonable degree of belief.

(c) The third is to claim that theorems (1) and (2) are of limited validity, not holding for all possible sets of mutually exclusive and exhaustive propositions $\{d_i\}$ and $\{d_\beta\}$. One would say that there are certain kinds of evidence which so completely restructure a problem that it is unreasonable to relate prior and posterior probabilities in accordance with the standard theorems of conditional probability.[5] Again, however, a deep analysis of the concept of reasonable degree of belief would be needed in order to defend this general claim, and in particular to show that the propositions $\hat{d}_\epsilon$ indeed require such restructuring. Furthermore, such a limitation upon the use of the theorems of conditional probability might be crippling to Jaynes's treatment of the statistical mechanics of time-dependent processes.

REFERENCES

1. (a) E. T. Jaynes, *Phys. Rev.* **106**:620 (1957); (b) *Phys. Rev.* **108**:171 (1957); (c) Information theory and statistical mechanics, in *Statistical Physics* (Brandeis University Summer Institute Lectures in Theoretical Physics, 1962, Vol. 3), W. K. Ford, ed., Benjamin, New York, 1963, p. 181; (d) Foundations of probability theory and statistical mechanics, in *Delaware Seminar in the Foundations of Physics*, M. Bunge, ed., Springer, New York, 1967, p. 77.
2. (a) Pierre Simon de Laplace, *A Philosophical Essay on Probabilities*, Dover, New York, 1951 [transl. from *Essai philosophique sur les probabilités* (1796) by F. W. Truscott and F. L. Emory]; (b) J. M. Keynes, *A Treatise on Probability*, Macmillan, London and New York, 1921; (c) Harold Jeffreys, *Theory of Probability*, 3rd Ed., Oxford Univ. Press, London, 1961; (d) Rudolf Carnap, *Logical Foundations of Probability*, 2nd Ed., Univ. of Chicago Press, Chicago, 1962; (e) Rudolf Carnap, *The Philosophy of Rudolf Carnap*, P. A. Schilpp, ed., Open Court, Lasalle, 1963, pp. 971–972.
3. H. Reiss, *J. Stat. Phys.* **1**:107 (1969).
4. R. S. Imgarden and K. Urbanik, *Acta Phys. Polon.* **21**:281 (1962): A. Katz, *Principles of Statistical Mechanics*, Freeman, San Francisco and London, 1967; R. Baierlein, *Atoms and Information Theory: An Introduction to Statistical Mechanics*, Freeman, San Francisco and London, 1970.
5. Yu. V. Prohorov and Yu. A. Rozanov, *Probability Theory*, Springer, New York, 1969, p. 110.
6. A. Shimony, Scientific inference, in *The Nature and Function of Scientific Theories*, R. G. Colodny, ed., Univ. of Pittsburgh Press, Pittsburgh, 1971.

[5] There is a suggestion of this kind in Ref. 6, especially pp. 103–104.