Journal of Statistical Physics, Vol. 9, No. 3, 1973

Replies to Tribus and Motroni and to Gage and Hestenes

Kenneth Friedman[1]

Received March 30, 1973

An earlier criticism of Jaynes's maximum-entropy prescription is vindicated with respect to two recent replies to that criticism.

KEY WORDS: Information theory; Jaynes; probability; statistical mechanics.

Both Tribus and Motroni[1] and Gage and Hestenes[2] defend Jaynes's maximum-entropy prescription against a criticism by Friedman and Shimony.[3] This criticism purports to find a difficulty verging on inconsistency in Jaynes's prescription and argues as follows: Consider a system about which the only information b is our background knowledge including the values of the energy states of the system. Then Jaynes's prescription assigns probabilities to the energy states in accordance with

$$P(h_i \mid b) = 1/n \tag{1}$$

If one obtains the additional information $\hat{d}_\epsilon$ that the posterior expected value of the energy E is ϵ, Jaynes's prescription yields

$$P(h_i \mid b \,\&\, \hat{d}_\epsilon) = e^{-\beta E_i} \times \left(\sum e^{-\beta E_i}\right)^{-1} \tag{2}$$

where β is a monotonically decreasing function of ϵ.

Equations (1) and (2) are consistent only under the assumption that

$$P(\hat{d}_\epsilon \mid b) = \delta\left(\epsilon - (1/n) \sum E_i\right)$$

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

Thus Jaynes's prescription is consistent only if one has prior certainty that the posterior expected energy of the system will equal its mean energy (i.e., that the temperature of the reservoir with which the system is in equilibrium is infinite).

I consider the comments of Tribus and Motroni first. They write (Ref. 1, p. 228)

> "It is obvious that ... the statement 'The expected value of E is $\frac{1}{2}(n + 1)$' [which Friedman and Shimony designate $\hat{d}_{\epsilon=(n+1)/2}$] leads to precisely the same distribution as the statement 'I have no reason to choose one value of i over another.' [which Friedman and Shimony designate b]. The information theory interpretation of this result is quite simple: Entropy measures what is unknown. Many statements in the English language lead to the same entropy because they say the same thing."

Two comments appear in order. First, the claim that the two cited statements say the same thing begs the question. The former statement plus Jaynes's prescription implies the latter; the latter plus Jaynes's prescription implies the former. Yet the acceptance of Jaynes's prescription is just the issue in question. Moreover, as the first statement is used by Jaynes and others, it should be amended to read, "We have positive information about the posterior expected value of E, and enough to fix that value as $(1/n) \sum E_i$." This is the point of the identification of the measured energy with the expected energy. The second statement should read, "We have no information about the system except for the values of the possible energy states." On these readings the two statements are clearly not equivalent.

Second, the difficulty with Jaynes's prescription does not stem from the fact that the two conditioning statements cited by Tribus and Motroni lead to the same probability distribution. Rather, it stems from the fact, illustrated by their second graph (Ref. 1, p. 228), that given the nature of the information we are considering, the posterior probability of the mean energy state (in this case $E = 4$) can never be revised upward, but must be revised downward for almost any value of ϵ. According to a Bayesian account that a prior probability cannot be revised upward (given the type of information being considered), but will with probability p be revised downward, implies that unless $p = 0$ the prior probability must be too high.

The comment by Gage and Hestenes finds two inaccuracies in the paper of Friedman and Shimony. We consider the latter inaccuracy first. They write (Ref. 2, p. 90), "Second, just following their Eq. (6) Friedman and Shimony state that the background information b does not in general imply a definite value of ϵ." Gage and Hestenes contest this point. From (1) they derive

$$\epsilon = \sum E_i P(h_i \mid b) = (1/n) \sum E_i \tag{3}$$

This is unexceptionable. They continue, however, "This can be expressed alternatively by writing

$$P(\hat{d}_\epsilon \mid b) = \delta \left(\epsilon - \frac{1}{n} \sum E_i\right)\text{''} \tag{4}$$

This may be the basis of Tribus and Motroni's claim that the two statements they consider say the same thing. However, Eq. (4) is *not* equivalent to Eq. (3). Equation (3) just states what the expected energy is according to our prior distribution. Given the tentative nature of that prior distribution, the probability that our prior estimate of the expected energy is correct is surely less than unity. Yet Eq. (4) denies this. Indeed, if Eq. (4) were correct, it is not clear how a Bayesian could establish that evidence can ever fix the posterior expected value of any variable at any value different from its prior expected value. It is not clear how one could justify Jaynes's identification of the measured energy with the average energy (at least when the measured energy differs from $(1/n) \sum E_i$).

Perhaps the motivation for the transition from (3) to (4) is the existence of limit theorems to the effect that as the length of a sequence of independent trials, each yielding a value E_i ($1 \leqslant i \leqslant n$), increases, the fraction of E_i's approaches $P(E_i)$. Thus if the probability distribution over the n energy states is a uniform one, the central limit theorem yields

$$P \left\{ \left| \epsilon - (1/n) \sum E_i \right| < \alpha\sigma(m)^{-1/2} \right\} \to (2\pi)^{-1/2} \int_{-\alpha}^{\alpha} \exp -\tfrac{1}{2}x^2 \, dx \tag{5}$$

For any α, and hence for any value of the right-hand side of (5) less than unity, there is a value of m such that the left-hand side of (5) is satisfied. Thus for any positive ϵ_1 and ϵ_2, as the length of the sequence increases indefinitely, the probability of the average energy lying within ϵ_1 of $(1/n) \sum E_i$ is greater than $1 - \epsilon_2$. This result can be derived in other ways, e.g., from a generalization of the Laplace–DeMoivre limit theorem to the multinomial distribution.[4]

The independence of the trials is a necessary assumption in all of these derivations. Yet this assumption is not warranted in the situation we are considering. If for a very large value of m the first m trials strongly favor a proper subset of the energy spectrum, it is very likely that subsequent trials will do so as well. Thus in general the probability that the $(m + 1)$th trial will yield a particular energy state depends on whether it is conditioned by just our background knowledge, or by our background knowledge and also our knowledge of the first m trials. Hence the trials are not statistically independent.

(The reply that the sequence of energy states generated by a thermodynamic system in thermal equilibrium with a reservoir is a statistically independent sequence is unsatisfactory. The probabilities we are considering either measure a reasonable degree of belief in the occurrence of various energy states at a given time or measure the physical propensity for a thermodynamic system to be in certain energy states. If we adopt the former interpretation, it is legitimate to postulate a prior uniform distribution, but incorrect to suppose that the energy states in the sequence are statistically independent. On the latter interpretation it is legitimate to suppose that the energy states in the sequence are statistically independent, but it is unclear that one can talk meaningfully about a prior distribution over the energy states.)

The first alleged inaccuracy cited by Gage and Hestenes seems unrelated to this argument, but it introduces an interesting ramification. They write (Ref. 2, p. 90),

> "First, just before their Eq. (4) they introduce $\hat{d}_\epsilon$ as the 'evidence that the posterior expected value of (a dynamical variable) E is ϵ.' Later they admit that this definition of $\hat{d}_\epsilon$ is less than clear. However, their use of $\hat{d}_\epsilon$ in connection with Jaynes's algorithm *demands* that it is equivalent to the well-defined proposition 'the expected value of E is ϵ.'"

This "inaccuracy" is not an inaccuracy in the interpretation of Jaynes, but rather reveals another difficulty with Jaynes's algorithm. The reason for the problem in interpreting $\hat{d}_\epsilon$ as fixing the expected value of the energy at a single point is that no finite amount of evidence can imply that the expected value of a statistical variable is a definite value. Indeed, if we make the not altogether unreasonable assumption that $P(\hat{d}_\epsilon \mid b)$ is a uniformly continuous function of ϵ, then for any finite amount of evidence e and for any value of ϵ, $P(\hat{d}_\epsilon \mid b \,\&\, e) = 0$.

It may be more reasonable to think of $\hat{d}_\epsilon$ as "evidence that the posterior expected value of the energy lies between certain limits $\underline{E}$ and $\bar{E}$ ($\underline{E} < \bar{E}$)." However, even this fails to render Eqs. (1) and (2) mutually compatible under reasonable assumptions.

Presumably, Jaynes's prescription now chooses that probability distribution with the greatest entropy out of all those distributions whose expected value of E lies in the interval $[\underline{E}, \bar{E}]$. However, if $\underline{E} \leqslant 1/n \sum E_i \leqslant \bar{E}$, this distribution will just be the uniform distribution, the same distribution we would obtain had we interpreted $\hat{d}_\epsilon$ as "evidence that the posterior expected value of E is exactly $(1/n) \sum E_i$." If $\underline{E} < (1/n) \sum E_i$ ($\bar{E} > (1/n) \sum E_i$), the probability distribution with the greatest entropy is one whose expectation value of the energy is $\underline{E}$ ($\bar{E}$). This is the same distribution we would have obtained had we interpreted $\hat{d}_\epsilon$ as "evidence that the posterior expected value of the energy is exactly $\underline{E}$ ($\bar{E}$)."

Thus the incompatibility between Eq. (1) and (2) can be avoided by this reinterpretation of $\hat{d}_\epsilon$ only if

$$P(\underline{E} \leqslant (1/n) \sum E_i \leqslant \bar{E}) = 1 \tag{6}$$

However, (a) this constraint does not seem reasonable, and (b) carrying out Jaynes's prescription subject to this constraint will always yield the uniform distribution as the posterior distribution.

REFERENCES

1. M. Tribus and H. Motroni, *J. Stat. Phys.* **4**:227 (1972).
2. D. Gage and D. Hestenes, *J. Stat. Phys.* **7**:89 (1973).
3. K. Friedman and A. Shimony, *J. Stat. Phys.* **3**:381 (1971).
4. B. V. Gnedenko, *The Theory of Probability*, Chelsea Publishing Co., New York (1963).