View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by Elsevier - Publisher Connector

ADVANCES IN APPLIED MATHEMATICS 2, 172–211 (1981)

A Critique of Jaynes' Maximum Entropy Principle*

PENHA MARIA CARDOSO DIAS

Department of Physics, Boston University, Boston, Massachusetts 02215

AND

ABNER SHIMONY

Departments of Philosophy and Physics, Boston University, Boston, Massachusetts 02215

Friedman and Shimony exhibited an anomaly in Jaynes' maximum entropy prescription: that if a certain unknown parameter is assumed to be characterized a priori by a normalizable probability measure, then the prior and posterior probabilities computed by means of the prescription are consistent with probability theory only if this measure assigns probability 1 to a single value of the parameter and probability 0 to the entire range of other values. We strengthen this result by deriving the same conclusion using only the assumption that the probability measure is σ-finite. We also show that when the hypothesis and evidence to which the prescription is applied are expressed in certain rather simple languages, then the maximum entropy prescription yields probability evaluation in agreement with one of Carnap's λ-continuum of inductive methods, namely $\lambda = \infty$. We conclude that the maximum entropy prescription is correct only under special circumstances, which are essentially those in which it is appropriate to use $\lambda = \infty$.

I. INTRODUCTION

In the past two decades Jaynes [12–17] has developed an influential program concerning inductive inference and the foundations of statistical mechanics, by proceeding from two premisses.

1. The concept of probability which is deployed in those two disciplines should usually be interpreted in the "logical" sense. That is to say, a

*This paper is based upon the Ph.D. thesis submitted by one of us (P. D.) to the Graduate School of Boston University in 1979. The research of one of us (A. S.) was supported in part by the National Science Foundation, and of the other (P. D.) by the Organization of American States.

0196-8858/81/020172-40$05.00/0
Copyright © 1981 by Academic Press, Inc.
All rights of reproduction in any form reserved.

sentence involving probability, such as "the probability of hypothesis h upon data d is the real number r"—which may be abbreviated as "$P(h|d) = r$"—should be interpreted as "the reasonable degree of belief in h, if d is the total available evidence, is r." (Jaynes does not disallow the possibility of other interpretations, such as the frequency interpretation, which has probably been favored by the overwhelming majority of statistical physicists in the past, but he claims that there are great advantages in clarifying and justifying the procedures both of inductive logic and of statistical mechanics if the logical interpretation is employed systematically.)

2. In order to evaluate $P(h|d)$ quantitatively, one should apply the *maximum entropy principle*, whenever d has a form which permits this principle to be employed. The concept of entropy involved in this principle is that of information theory [25]: that is, if one is considering a set of n mutually exclusive and exhaustive hypotheses $\{h_i\}$ and if d is the total body of data, then the entropy in this situation is defined as

$$S(p_1,\ldots,p_n) = -C\sum_i^n p_i \ln p_i, \tag{1.1}$$

where $p_i = P(h_i|d)$ and C is an arbitrary positive real number. Suppose now that the data consist of numerical values of the averages $\epsilon_1,\ldots,\epsilon_r$ of r quantities $E^1,\ldots,E^r$, where E^j has a definite value $E_i^j \equiv E^j(h_i)$ in case the hypothesis h_i is true; here r is a nonnegative integer, possibly 0, the latter case corresponding to null evidence. These data presumably are factual in character, but the way one obtains them is irrelevant at the present abstract level of treatment so long as it is legitimate to assert the equalities

$$\sum_i E_i^j p_i = \epsilon_j. \tag{1.2}$$

With these preliminaries, the maximum entropy principle can be explicitly formulated: the numerical values $P(h_i|d) \equiv p_i$ must be such as to maximize the entropy $S(p_1,\ldots,p_n)$, subject to the constraints of Eq. (1.2) and of $\Sigma p_i = 1$.

Parts of Jaynes' program can be found in the work of earlier authors. The logical interpretation of probability was advocated by Keynes [19], Jeffreys [18], Carnap [2] and perhaps even by Laplace [20]. Shannon [25] defined the concept of information-theoretical entropy and showed how it must be expressed in terms of probabilities if certain reasonable desiderata are to be satisfied. However, the combination of ideas in Jaynes' program is original. He uses information theoretical entropy as a fundamental concept in statistical mechanics, and under appropriate conditions (primarily that the maximum entropy principle be used to calculate the probabilities p_i) he even identifies this kind of entropy with that of thermodynamics [13, pp.

196–197]. And he proposes that the perennial difficulty of the logical probability theorists—the problem of providing a systematic and objective procedure for numerically evaluating probabilities—can be solved in large part by using the maximum entropy principle (e.g., [15, pp. 228ff]). Jaynes' striking achievement, which has attracted a considerable following to his ideas, is the derivation of the generalized Boltzmann distribution from his small set of premisses, in remarkably few steps and without recourse to ergodicity or other special physical assumptions. The maximization of S in Eq. (1.1), subject to the constraints of Eq. (1.2) and the normalization condition $\Sigma p_i = 1$, can be carried out by the usual method of Lagrange multipliers. The result is

$$p_i = \frac{\exp\left(-\sum_{j=1}^{r} \beta_j E_i^j\right)}{Z(\beta_1,\ldots,\beta_r)}, \tag{1.3}$$

where

$$Z(\beta_1,\ldots,\beta_r) = \sum_{i=1}^{r} \exp\left(-\sum_{j=1}^{r} \beta_j E_i^j\right). \tag{1.4}$$

In the special case where r is 1 and E^1 is taken to be the energy E, there is only a single parameter $\beta_1 \equiv \beta$, and this can be identified with $1/(kT)$, where T is the absolute temperature and k is the Boltzmann constant. One then has the usual Boltzmann distribution for a system with n states, in contact with a heat bath at temperature T:

$$p_i = \frac{\exp(-E_i/kT)}{\sum_l \exp(-E_l/kT)}. \tag{1.5}$$

Despite the ease of Jaynes's derivation of the Boltzmann distribution, many scientists have remained skeptical, mainly because they were unsure of the rationale for the maximum entropy principle (e.g., [31, p. 248; 22, p. 44]). Jaynes himself offers an epistemological justification of the probability distribution obtained by means of the maximum entropy principle: it "is uniquely determined as the one which is maximally noncommittal with regard to missing information" [12, I, p. 623], and it "provides the most honest description of what we know" [16, p. 97]. Between this quite plausible justification and the residual suspicion that somehow too much knowledge has been extracted from ignorance there may appear to be a deadlock.

An additional consideration against the maximum entropy principle was presented by Friedman and Shimony [9], to be abbreviated henceforth by "FS." They showed that in certain circumstances, the posterior probabilities, calculated by means of the maximum entropy principle when one constraint of the form (1.2) is given, and the prior probabilities, calculated with no constraints other than $\Sigma p_i = 1$, cannot be made consistent both with each other and with the general rules of probability theory unless a special and highly implausible condition is satisfied. Specifically, they consider background information b which specifies

(i) that there are n mutually exclusive and exhaustive hypotheses $h_1,\ldots,h_n$,

(ii) that the quantity E has the value E_i if h_i is true,

(iii) that not all the $E_i,\ldots,E_n$ are equal, and

(iv) that one of these values—say, the mth—is the average of all of them, i.e.,

$$E_m = (1/n)\sum E_i. \tag{1.6}$$

The prior probability of h_i is $p_i^0 \equiv P(h_i|b)$. If the data $\hat{d}_\epsilon$ assert that the average of E is ϵ, then the posterior probability of h_i relative to $\hat{d}_\epsilon$ (which will sometimes be called simply "the posterior probability", when it is clear from context what body of data has been assumed) is $p_i \equiv P(h_i|b\,\&\,\hat{d}_\epsilon)$. Since the average of E is a monotonically decreasing function of the parameter β when the Boltzmann distribution (1.5) is used, there is a one-one mapping of the range of ϵ onto the range of β, and therefore the data $\hat{d}_\epsilon$ can equally well be expressed as the data d_β. Consequently, p_i can be expressed as $P(h_i|b\,\&\,d_\beta)$. The background information does not specify the value of ϵ and hence not of β, but from the point of view of the logical interpretation of probability it makes sense to speak of the probability distributions of ϵ and of β, given b. Hence FS assumed that there are well defined distribution functions $F(\beta|b)$ and $\hat{F}(\epsilon|b)$ such that

$F(\beta|b) =$ the probability that the parameter has a value equal to or less than β, given b,

$\hat{F}(\epsilon|b) =$ the probability that the average energy has a value equal to or less than ϵ, given b.

For F to be a distribution function means that it is monotonically nondecreasing and continuous on the right and that

$$\lim_{\beta\to-\infty} F(\beta|b) = 0 \qquad \text{and} \qquad \lim_{\beta\to\infty} F(\beta|b) = 1;$$

and likewise for $\hat{F}$. The existence of the distribution function F is, of course, equivalent to the existence of a probability measure μ such that for every Borel subset S of R,

$$\mu(S) = \int_S dF(\beta|b), \tag{1.7}$$

and

$$\mu(R) = 1. \tag{1.8}$$

FS then pointed out that according to the general rules of probability theory the prior and posterior probabilities are connected by the equation

$$P(h_i|b) = \int P(h_i|b \& d_\beta) dF(\beta|b), \tag{1.9}$$

for each $i = 1, \ldots, n$. If the prior and posterior probabilities are computed in accordance with the maximum entropy principle and inserted into Eq. (1.9), one obtains

$$\frac{1}{n} = \int \frac{e^{-\beta E_i}}{\sum_{k=1}^{n} e^{-\beta E_k}} dF(\beta|b). \tag{1.10}$$

They then inquired what restrictions are imposed upon the distribution function F by Eq. (1.10) and proved that F is determined uniquely

$$\begin{aligned} F(\beta|b) &= 0 \qquad \beta < 0, \\ F(\beta|b) &= 1 \qquad \beta \geq 0, \end{aligned} \tag{1.11}$$

or equivalently,

$$\mu(\{0\}) = 1, \qquad \mu(R - \{0\}) = 0. \tag{1.12}$$

A proof of this result is given in Appendix A. The result is dangerous to Jaynes' program, for it shows that the repeated application of the maximum entropy principle is compatible with the general rules of probability theory only if there is probability 1 upon the background information b (which is quite meagre information) that the parameter β will be found to be 0. But to assert this much seems to be inconsistent with Jaynes' cardinal epistemological maxim that one should be honest about the extent of one's ignorance. Seidenfeld [24] states some important related criticisms.

Several attempts have been made to refute FS on the grounds that they have not properly understood the logical concept of probability [10, 11], but these refutations seem to us unconvincing [7, 27]. It was suggested by FS that either the domain of definition of the logical concept of probability might be restricted, or the domain of applicability of the maximum entropy principle could be limited; but to our knowledge no one has worked out a limited version of Jaynes' program along either of these lines.

There is another possible defense against the difficulty raised by FS, which seems not to have been mentioned previously in the literature. That is the possibility that they made too restrictive an assumption about the distribution function F and the related probability measure μ. Expectation values and other probabilistic quantities of interest can be computed even when the probability distribution is not normalizable—i.e., when

$$\lim_{\beta \to \infty} F(\beta | b) = \infty, \tag{1.13}$$

or equivalently,

$$\mu(R) = \infty. \tag{1.14}$$

It is important to explore whether the prior and the posterior applications of the maximum entropy principle are consistent with each other, with probability theory, and with the general maxims of Jaynes if appropriate nonnormalizable functions of β are used. Sections II and III of this paper are devoted to an examination of this possible line of defense, and the results indicate that it will not succeed. Specifically, in Section II the assumption that $F(\beta | b)$ is normalizable, or equivalently that $\mu(R) = 1$, is replaced by the assumption that μ is a σ-finite measure, i.e., that $\mu(S)$ is finite for any bounded Borel subset of R; all the other assumptions of FS are retained. It is shown that even with this relaxation of an assumption, the implausible Eqs. (1.11) and (1.12) can be derived. In Section III we investigate a distribution function which is not associated with a σ-finite measure. It turns out that there are ambiguities in the expressions for expectation values in terms of this distribution function; but when the ambiguities are removed in such a way as to obtain Eq. (1.10), then one again obtains the result that the a priori probability of finding $\beta = 0$ is unity.

These difficulties indicate that Jaynes' program should be curtailed. Nevertheless, it seems that some part of it should be rescued, since there is something attractive about the maximum entropy principle and since the principle does yield some important results, notably the Boltzmann distribution, which we believe to be correct for independent reasons. Section IV is devoted to an investigation of the circumstances under which it is legitimate to apply the maximum entropy principle. The procedure of this Section is to compare Jaynes' probability evaluations with those given by each of Carnap's "continuum of inductive methods" [3], which is a family of methods parametrized by a single variable λ such that $0 \leq \lambda \leq \infty$. We show that Jaynes' method agrees with the method corresponding to $\lambda = \infty$ and disagrees with all others. Since $\lambda = \infty$ is the method which assigns equal probabilities to all states compatible with the evidence, we conclude that Jaynes' method is legitimate in circumstances when that assignment is justifiable, but not otherwise.

II. σ-Finite Measures on the Parameter Space

In this Section we shall consider probability distributions which need not be normalizable, but which have all the other properties assumed in Section I. That is, $F(\beta|b)$ is real, monotonically nondecreasing, and continuous on the right, with $\lim F(\beta|b) = 0$ as β goes to $-\infty$; but we permit $\lim F(\beta|b)$ as β goes to ∞ to be either a positive real number or infinite. It follows that $\mu(S) = \int_S dF(\beta|b)$ is finite for every bounded Borel subset S of R, and hence is a σ-finite measure. The expectation value of a function $A(\beta)$ relative to F can be expressed as follows:

$$\langle A \rangle = \lim_{K\to\infty} \left[\frac{\int_{-K}^{K} A(\beta)dF(\beta|b)}{\int_{-K}^{K} dF(\beta|b)} \right] = \lim_{K\to\infty} \left[\frac{\int_{[-K,K]} A(\beta)\,d\mu}{\mu([-K,K])} \right]. \tag{2.1}$$

Similarly, Eq. (1.10), which expressed the relation imposed by probability theory upon the prior and posterior probabilities resulting from the maximum entropy principle, must be generalized:

$$\frac{1}{n} = \lim_{K\to\infty} \left[\frac{\int_{[-K,K]} \frac{e^{-\beta E_i}}{\Sigma e^{-\beta E_i}}\,d\mu}{\mu([-K,K])} \right]. \tag{2.2}$$

For every positive real number K which is large enough that $\mu([-K,K]) > 0$ we can define a probability measure

$$\mu_K(S) \equiv \frac{\mu(S \cap [-K,K])}{\mu([-K,K])}, \tag{2.3}$$

with the obvious property that

$$\mu_K(R) = 1. \tag{2.4}$$

Then Eq. (2.2) can be restated as

$$\frac{1}{n} = \lim_{K\to\infty} \int \frac{e^{-\beta E_i}}{\sum e^{-\beta E_l}} d\mu_K. \tag{2.5}$$

The remainder of Section II will be devoted to proving the following

THEOREM. *Equation* (2.2) (*or equivalently Eq.* (2.5)) *implies that* $F(\beta|b) = 0$ *for* $\beta < 0$ *and* $F(\beta|b) = c$ *for* $\beta \geq 0$ (*or equivalently, that* $\mu(\{0\}) = \mu(R) = c$), *where c is a finite positive real number.*

The background information b mentioned in this theorem is the same as in Section I. In the first of two lemmas used to prove the theorem, we use m to designate the integer mentioned in one of the conditions contained in b, such that

$$E_m = 1/n \sum E_i. \tag{1.6}$$

LEMMA 1. *If we define*

$$\xi_i(\beta) = \frac{e^{-\beta E_i}}{\sum e^{-\beta E_l}}, \qquad i = 1,\dots,n,$$

then $\xi_m(\beta)$ *is strictly increasing for* $\beta < 0$ *and strictly decreasing for* $\beta > 0$, *and therefore it has an absolute maximum at* $\beta = 0$.

Proof. If the expectation value ϵ of the quantity E is expressed as a function

$$\epsilon(\beta) \equiv \frac{\sum E_l e^{-\beta E_l}}{\sum e^{-\beta E_l}}, \tag{2.6}$$

then it is well known that $\epsilon' < 0$, where a prime denotes differentiation by β. From $\epsilon(0) = E_m$ it follows that $\epsilon(\beta) > E_m$ if $\beta < 0$ and $\epsilon(\beta) < E_m$ if $\beta > 0$. From $\xi_m'(\beta) = \xi_m(\beta)\left[\epsilon(\beta) - E_m\right]$ it follows that $\xi_m'(0) = 0$; and $\beta = 0$ must be an absolute maximum, since $\beta < 0$ implies $\xi_m'(\beta) > 0$ while $\beta > 0$ implies $\xi_m'(\beta) < 0$. Q.E.D.

From Eq. (2.5) and the fact that i takes on only a finite number n of values it follows that for any positive integer M there is an integer $N(M)$ (which can be selected so that $M' > M$ implies $N(M') > N(M)$) such that

$$\left| \frac{1}{n} - \int_{-\infty}^{\infty} \xi_i(\beta)\, d\mu_{N(M)} \right| < M^{-3}, \tag{2.7}$$

for all $i = 1,\ldots,n$. The following notation will be convenient:

$$\mu_M^0 = \int_{-1/2M}^{1/2M} d\mu_{N(M)}, \tag{2.8a}$$

$$\mu_M^+ = \int_{1/2M}^{\infty} d\mu_{N(M)}, \tag{2.8b}$$

$$\mu_M^- = \int_{-\infty}^{-1/2M} d\mu_{N(M)}, \tag{2.8c}$$

and

$$p_i^M = \int_{-\infty}^{\infty} \xi_i(\beta)\, d\mu_{N(M)}. \tag{2.9}$$

Clearly,

$$\mu_M^0 + \mu_M^+ + \mu_M^- = 1, \tag{2.10}$$

and

$$\left|\frac{1}{n} - p_i^M\right| < M^{-3}. \tag{2.11}$$

We can now prove the following

LEMMA 2. $\lim_{M\to\infty}(1 - \mu_M^0) = 0$.

Proof. By Lemma 1, $\beta < -1/2M$ implies

$$\int_{-\infty}^{-1/2M} \xi_m(\beta)\, d\mu_{N(M)} \leq \mu_M^- \xi_m(-1/2M),$$

$\beta > 1/2M$ implies

$$\int_{1/2M}^{\infty} \xi_m(\beta)\, d\mu_{N(M)} \leq \mu_M^+ \xi_m(1/2M),$$

and $-1/2M \leq \beta \leq 1/2M$ implies

$$\int_{-1/2M}^{1/2M} \xi_m(\beta)\, d\mu_{N(M)} \leq \mu_M^0 \cdot 1/n.$$

Hence, by Eqs. (2.8a), (2.8b), (2.8c), (2.9), (2.10), and (2.11),

$$p_m^M \leq 1/n,$$

and

$$0 \leq 1/n\left(1 - \mu_M^0\right) - \mu_M^+ \xi_m(1/2M) - \mu_M^- \xi_m(-1/2M) \leq \frac{1}{n} - p_m^M < M^{-3}, \tag{2.12}$$

whence

$$0 \le (1-\mu_M^0)\left[\frac{1}{n}-\xi_m(-1/2M)\right]-\mu_M^+[\xi_m(1/2M)-\xi_m(-1/2M)]$$
$$< M^{-3}. \tag{2.13}$$

By Taylor's theorem,

$$\xi_m(\pm 1/2M)=\frac{1}{n}-\frac{1}{2!}(\pm 1/2M)^2|\xi_m''(0)|+R_2(\pm 1/2M),$$

where the formula for the remainder is

$$R_n(h)=\frac{1}{(n+1)!}h^{n+1}\left[\frac{d^{n+1}}{d\beta^{n+1}}\xi_m(\beta)\right]_{\beta=\Theta h}, \qquad 0<\Theta<1.$$

Hence Ineq. (2.13) becomes

$$0 \le (1-\mu_M^0)\left[\tfrac{1}{2}(1/2M)^2|\xi_m''(0)|-\tfrac{1}{6}(1/2M)^3\xi_m'''(\tilde{\Theta}/2M)\right]$$
$$-\mu_M^+\left[\xi_m'''(\tilde{\Theta}/2M)+\xi_m''(-\Theta/2M)\right]\tfrac{1}{6}(1/2M)^3<M^{-3}.$$

We have used Θ for the case of $-1/2M$ and $\tilde{\Theta}$ for the case of $1/2M$ in the remainder formula. Since $|\xi_m''(0)|>0$, we can multiply by $8M^2/|\xi_m''(0)|$ and rearrange terms, to obtain

$$0 \le (1-\mu_M^0)<\frac{1}{M|\xi_m''(0)|}\left\{8-\tfrac{1}{6}(1-\mu_M^0)\xi_m'''(-\Theta/2M)\right.$$
$$\left.+\tfrac{1}{6}\mu_M^+\left[\xi_m'''(\tilde{\Theta}/2M)+\xi_m'''(-\Theta/2M)\right]\right\}. \tag{2.14}$$

Since $\mu_M^+ \le 1$, $(1-\mu_M^0)\le 1$, and $\xi_m'''(\beta)$ is continuous and therefore bounded in the interval $[-1,1]$, it follows that the expression in curly brackets is bounded by a number independent of M, and hence $\lim_{M\to\infty}(1-\mu_M^0)=0$. Q.E.D.

We can now complete the proof of the theorem by showing that $0<\mu(\{0\})<\infty$ and that $\mu(R-\{0\})=0$. Suppose first that $\mu(R)=\infty$. Then for any divergent monotonically increasing sequence $\{t_i\}$ of positive real numbers there is a sequence of integers $\{K_i\}$, which can be chosen increasing without loss of generality, such that

$$\mu([-K_i, K_i])>2t_i.$$

Since $\lim_{M\to\infty}(1-\mu_M^0)=0$, and since $N(M)$ as defined just before Eq. (2.7) is a monotonically increasing function of M, one can pick out a monotonically increasing sequence of positive integers $\{M_i\}$ such that

$$N(M_i) > K_i \qquad \text{for each } i$$

and

$$1 - \mu^0_{M_i} < \tfrac{1}{2} \qquad \text{for each } i.$$

Hence

$$\mu^0_{M_i} \equiv \frac{\mu([-1/2M_i, 1/2M_i])}{\mu([-N(M_i), N(M_i)]} > \frac{1}{2}.$$

Therefore,

$$\mu([-1/2M_i, 1/2M_i]) > \tfrac{1}{2}\mu([-N(M_i), N(M_i)]) > \tfrac{1}{2}\mu([-K_i, K_i]) > t_i.$$

Since the interval $[-1, 1]$ contains the interval $[-1/2M_i, 1/2M_i]$ for each i, we obtain

$$\mu([-1, 1]) > t_i, \qquad \text{for all } i,$$

contrary to the hypothesis that μ is a σ-finite measure. Therefore the supposition that $\mu(R)$ is infinite has led to a contradiction. But if $\mu(R)$ is finite, then μ is normalizable, and we can define

$$\hat{\mu}(S) \equiv \frac{\mu(S)}{\mu(R)}, \qquad \hat{\mu}(R) = 1.$$

The theorem of FS then states that $\hat{\mu}(\{0\}) = 1$ and $\mu(R - \{0\}) = 0$. Hence $0 < \mu(\{0\}) = \mu(R)$ and $\mu(R - \{0\}) = 0$. Equivalently, $F(\beta|b) = 0$ for $\beta < 0$, while $F(\beta|b) =$ a positive finite constant for $\beta \geq 0$. Q.E.D.

Once one has shown that $\mu(R)$ is finite, an alternative procedure is possible which dispenses with the theorem of FS: essentially, it consists of using Lemma 2 together with the measure theoretical proposition that

$$\mu(\{0\}) = \lim_{M \to \infty} \mu([-1/2M, 1/2M]).$$

III. The Probability Density $1/\beta$

We shall now consider whether it is possible to escape from the difficulty posed by the theorem of FS by assuming

$$dF(\beta|b) = d\beta/\beta, \qquad \beta > 0. \qquad (3.1)$$

The density $1/\beta$ is appealing, since it has certain invariance properties for a random variable which is known a priori (i.e., because of the background information b) to lie in the interval $[0, \infty]$ (see [18]). If the background information b does not preclude negative values of β, one may wish to consider an extension of Eq. (3.1):

$$dF(\beta|b) = d\beta/|\beta|, \qquad \beta \neq 0. \tag{3.1'}$$

The analysis of the present section goes through in essentially the same way and with essentially the same conclusions whether one uses Eq. (3.1) or Eq. (3.1′), but the calculations are somewhat simpler in the former case, and therefore we shall assume Eq. (3.1). In any case, our main interest in studying the distribution given by Eq. (3.1) is that it is not associated with a σ-finite measure and hence is not covered by the theorem of Section II. We wish to exhibit some of the difficulties which arise when one attempts to go beyond the confines of σ-finiteness, even though we are not able to treat them comprehensively.

Since the F of Eq. (3.1) is not normalizable, the nearest one can come to Eq. (1.10) using F is some equation of the following form:

$$\frac{1}{n} = \lim\left[\frac{\int_0^\infty \left(e^{-\beta E_i}/\Sigma e^{-\beta E_i}\right)\frac{1}{\beta}\,d\beta}{\int_0^\infty \frac{1}{\beta}\,d\beta}\right], \tag{3.2}$$

where the limiting procedure and the meaning of the integrals remain to be specified. In order to simplify the problem we shall take $n = 3$ and $E_i = i$. We shall examine the three quantities

$$P(h_1|b) = \lim\left[\frac{\int_{\beta_0}^{\beta_1} \frac{e^{-\beta}}{[e^{-\beta}+e^{-2\beta}+e^{-3\beta}]}\frac{d\beta}{\beta}}{\int_{\beta_0}^{\beta_1}(d\beta/\beta)}\right] = \lim\frac{I_1(\beta_0,\beta_1)}{N(\beta_0,\beta_1)}, \tag{3.3a}$$

$$P(h_2|b) = \lim\left[\frac{\int_{\beta_0}^{\beta_1} \frac{e^{-2\beta}}{[e^{-\beta}+e^{-2\beta}+e^{-3\beta}]}\frac{d\beta}{\beta}}{\int_{\beta_0}^{\beta_1}(d\beta/\beta)}\right] = \lim\frac{I_2(\beta_0,\beta_1)}{N(\beta_0,\beta_1)}, \tag{3.3b}$$

$$P(h_3|b) = \lim \left[\frac{\displaystyle\int_{\beta_0}^{\beta_1} \frac{e^{-3\beta}}{[e^{-\beta} + e^{-2\beta} + e^{-3\beta}]} \frac{d\beta}{\beta}}{\displaystyle\int_{\beta_0}^{\beta_1} (d\beta/\beta)} \right] = \lim \frac{I_3(\beta_0,\beta_1)}{N(\beta_0,\beta_1)}, \tag{3.3c}$$

where we use the notation

$$I_1(\beta_0,\beta_1) = \int_{\beta_0}^{\beta_1} \frac{d\beta}{[1 + e^{-\beta} + e^{-2\beta}]\beta}, \tag{3.4a}$$

$$I_2(\beta_0,\beta_1) = \int_{\beta_0}^{\beta_1} \frac{d\beta}{[e^{-\beta} + 1 + e^{\beta}]\beta}, \tag{3.4b}$$

$$I_3(\beta_0,\beta_1) = \int_{\beta_0}^{\beta_1} \frac{d\beta}{[1 + e^{\beta} + e^{2\beta}]\beta}, \tag{3.4c}$$

$$N(\beta_0,\beta_1) = \int_{\beta_0}^{\beta_1} \frac{d\beta}{\beta} = \ln\beta_1 - \ln\beta_0. \tag{3.4d}$$

Three different ways of allowing β_0 to approach 0 and β_1 to approach ∞ will be considered, and it will be seen that only the first and a special case of the third will yield

$$P(h_1|b) = P(h_2|b) = P(h_3|b) = \tfrac{1}{3}, \tag{3.5}$$

as required by the schematic Eq. (3.2).

1. First let $\beta_0 \to 0^+$, then let $\beta_1 \to \infty$.

For fixed β_1 satisfying $0 < \beta_1 < \infty$ it is obvious that $I_1(\beta_0,\beta_1)$, $I_2(\beta_0,\beta_1)$, $I_3(\beta_0,\beta_1)$, and $N(\beta_0,\beta_1)$ all diverge as $\beta_0 \to 0^+$. Hence, we use l'Hôpital's rule to calculate the ratios in Eqs. (3.3a), (3.3b), (3.3c).

$$\begin{aligned} \lim_{\beta_0 \to 0^+} \frac{I_1(\beta_0,\beta_1)}{N(\beta_0,\beta_1)} &= \lim_{\beta_0 \to 0^+} \frac{\partial I_1/\partial\beta_0}{\partial N/\partial\beta_0} \\ &= \lim_{\beta_0 \to 0^+} \left[\frac{-\dfrac{1}{[1 + e^{-\beta_0} + e^{-2\beta_0}]} \dfrac{1}{\beta_0}}{-1/\beta_0} \right] = \frac{1}{3}, \end{aligned} \tag{3.6a}$$

and similarly

$$\lim_{\beta_0 \to 0^+} \frac{I_2(\beta_0,\beta_1)}{N(\beta_0,\beta_1)} = \frac{1}{3}, \tag{3.6b}$$

$$\lim_{\beta_0 \to 0^+} \frac{I_3(\beta_0,\beta_1)}{N(\beta_0,\beta_1)} = \frac{1}{3}. \tag{3.6c}$$

The limit of each of these expressions as $\beta_1 \to \infty$ is obviously $\frac{1}{3}$. Hence by Eqs. (3.3a), (3.3b), (3.3c) we obtain the desired Eq. (3.5).

2. First let $\beta_1 \to \infty$, then let $\beta_0 \to 0^+$.

For fixed β_0 satisfying $0 < \beta_0 < \infty$ we have

$$\lim_{\beta_1 \to \infty} \frac{I_1(\beta_0,\beta_1)}{N(\beta_0,\beta_1)} = \lim_{\beta_1 \to \infty} \frac{\partial I(\beta_0,\beta_1)/\partial\beta_1}{\partial N(\beta_0,\beta_1)/\partial\beta_1}$$

$$= \lim_{\beta_1 \to \infty} \frac{\dfrac{1}{[e^{-2\beta_1} + e^{-\beta_1} + 1]} \dfrac{1}{\beta_1}}{1/\beta_1} = 1, \tag{3.7a}$$

again using l'Hôpital's rule, while

$$\lim_{\beta_1 \to \infty} \frac{I_2(\beta_0,\beta_1)}{N(\beta_0,\beta_1)} = \lim_{\beta_1 \to \infty} \frac{\partial I_2(\beta_0,\beta_1)/\partial\beta_1}{\partial N(\beta_0,\beta_1)/\partial\beta_1}$$

$$= \frac{\dfrac{1}{[e^{-\beta_1} + 1 + e^{\beta_1}]} \dfrac{1}{\beta_1}}{1/\beta_1} = 0, \tag{3.7b}$$

and similarly

$$\lim_{\beta_1 \to \infty} \frac{I_3(\beta_0,\beta_1)}{N(\beta_0,\beta_1)} = 0. \tag{3.7c}$$

Now taking limits as $\beta_0 \to 0^+$ and substituting in Eqs. (3.3a), (3.3b), (3.3c), we have

$$P(h_1|b) = 1, \qquad P(h_2|b) = P(h_3|b) = 0,$$

in disagreement with Eq. (3.5).

3. The limits are taken in tandem, by defining $\beta_1 = f(\beta_0)$, where f has a continuous first derivative for $0 \le \beta_0 < \infty$ and $\lim_{\beta_0 \to 0^+} f(\beta_0) = \infty$. Then

$$\lim_{\beta_0 \to 0^+} \frac{I_1(\beta_0, f(\beta_0))}{N(\beta_0, f(\beta_0))} = \lim_{\beta_0 \to 0^+} \frac{dI_1(\beta_0, f(\beta_0))/d\beta_0}{dN(\beta_0, f(\beta_0)/d\beta_0}$$

$$= \lim_{\beta_0 \to 0^+} \left\{ \frac{[1 + e^{-f(\beta_0)} + e^{-2f(\beta_0)}]^{-1} \dfrac{df/d\beta_0}{f(\beta_0)} - [1 + e^{-\beta_0} + e^{-2\beta_0}] \dfrac{1}{\beta_0}}{\dfrac{df/d\beta_0}{f(\beta_0)} - \dfrac{1}{\beta_0}} \right\}$$

$$= \lim_{\beta_0 \to 0^+} \frac{\dfrac{1}{3} - \dfrac{\beta_0}{f(\beta_0)} \dfrac{df}{d\beta_0}}{1 - \dfrac{\beta_0}{f(\beta_0)} \dfrac{df}{d\beta_0}}. \tag{3.8}$$

A necessary and sufficient condition for this limit to equal $\frac{1}{3}$, and also for the corresponding limits to equal $\frac{1}{3}$ when I_1 is replaced by I_2 or I_3, is

$$\lim_{\beta_0 \to 0^+} \frac{\beta_0}{f(\beta_0)} \frac{df}{d\beta_0} = 0. \tag{3.9}$$

In other words, Eq. (3.9) is a necessary and sufficient condition that Eq. (3.5) be satisfied if the third limiting procedure is used.

We shall now show that the first limiting procedure and the special case (3.9) of the third procedure imply effectively the existence of a probability measure μ for β such that $\mu(\{0\}) = 1$, $\mu(R - \{0\}) = 0$. Because the density $1/\beta$ diverges as $\beta \to 0$, we cannot initially assume that a probability measure is associated with β. However, we can define functions $\hat{\mu}_1, \hat{\mu}_3$ of open intervals (A, B) by means of our first and third limiting procedures, and then we can extend $\hat{\mu}_1$ and $\hat{\mu}_3$ to probability measures. If $0 < A < B$, we define

$$\hat{\mu}_1((A,B)) = \lim_{\beta_1 \to \infty} \lim_{\beta_0 \to 0^+} \frac{\int_A^B d\beta/\beta}{\int_{\beta_0}^{\beta_1} d\beta/\beta} = \lim_{\beta_1 \to \infty} \lim_{\beta_0 \to 0^+} \frac{\ln B - \ln A}{\ln \beta_1 - \ln \beta_0} = 0. \tag{3.10a}$$

If $A = 0$, we define

$$\hat{\mu}_1((0,B)) = \lim_{\beta_1 \to \infty} \lim_{\beta_0 \to 0^+} \frac{\ln B - \ln \beta_0}{\ln \beta_1 - \ln \beta_0} = 1. \tag{3.10b}$$

Similarly, if $0 < A < B$, we define

$$\hat{\mu}_3(A,B)) = \lim_{\beta_0 \to 0^+} \frac{\int_A^B d\beta/\beta}{\int_{\beta_0}^{f(\beta_0)} d\beta/\beta} = \lim_{\beta_0 \to 0^+} \frac{\ln B - \ln A}{\ln f(\beta_0) - \ln \beta_0}$$

$$= \lim_{\beta_0 \to 0^+} \frac{(\ln \beta_0)^{-1}(\ln B - \ln A)}{\dfrac{\beta_0}{f(\beta_0)} \dfrac{df}{d\beta_0} - 1} = 0, \tag{3.11a}$$

where we have used

$$\lim_{\beta_0 \to 0^+} \frac{\ln f(\beta_0)}{\ln \beta_0} = \lim_{\beta_0 \to 0^+} \frac{\beta_0}{f(\beta_0)} \frac{df}{d\beta_0}$$

(from l'Hôpital's rule) and also Eq. (3.9). If $A = 0$, we define

$$\hat{\mu}_3((0,B)) = \lim_{\beta_0 \to 0^+} \frac{\int_A^B d\beta/\beta}{\int_{\beta_0}^{\beta_1} d\beta/\beta} = \lim_{\beta_0 \to 0^+} \frac{\ln B - \ln \beta_0}{\ln f(\beta_0) - \ln \beta_0}$$

$$= \lim_{\beta_0 \to 0^+} \frac{(\ln \beta_0)^{-1} \ln B - 1}{\dfrac{\ln f(\beta_0)}{\ln \beta_0} - 1} = \lim_{\beta_0 \to 0^+} \frac{(\ln \beta_0)^{-1} \ln B - 1}{\dfrac{\beta_0}{f(\beta_0)} \dfrac{df}{d\beta_0} - 1} = 1. \tag{3.11b}$$

We can extend $\hat{\mu}_1$ and $\hat{\mu}_3$ to a probability measure μ (the same for both), by defining

$$\mu((A,B)) \equiv \hat{\mu}_1((A,B)) = \hat{\mu}_3((A,B)) \tag{3.12}$$

for all $0 \leq A < B < \infty$ and by requiring that μ be a σ-additive set function and that $\mu((-\infty, 0)) = 0$. The extension is unique, and one has

$$\mu(\{0\}) = 1, \qquad \mu(R - \{0\}) = 0, \tag{3.13}$$

or equivalently,

$$\mu(S) = 1 \qquad \text{if } 0 \in S, \tag{3.14a}$$

$$\mu(S) = 0 \qquad \text{if } 0 \notin S. \tag{3.14b}$$

We conclude that the schematic Eq. (3.2) can be satisfied only if calculations involving the density $1/\beta$ are performed in a way that is equivalent to assuming a probability measure on μ with total weight concentrated at $\beta = 0$.

IV. Relation of Jaynes' Method to Carnap's Continuum of Inductive Methods

The results of FS and of Sections II and III show that Jaynes' program cannot be maintained in the full generality which he intends. The problem remains to determine what part of the program is defensible, in particular to find out under what circumstances the maximum entropy principle is legitimate. Our strategy will be to compare Jaynes' method of evaluating probabilities with a family of methods systematically investigated by Carnap [3, 4] which he calls "the continuum of inductive methods" or "the λ-continuum." We shall set up a problem which is amenable both to Jaynes' and to Carnap's treatments, and we then find that precisely one among Carnap's continuum—namely, the one characterized by the value $\lambda = \infty$—yields probability evaluations in agreement with those of Jaynes, while all other methods in the continuum disagree with his evaluations. This result suggests that the circumstances under which it is proper to adopt the method with $\lambda = \infty$ are also those under which Jaynes' program is acceptable.

Carnap takes the arguments of his probability functions (which he calls "confirmation functions," in order to emphasize that he is dealing with the logical sense of probability) to be sentences in a definite language $\mathfrak{L}$. A confirmation function $\underline{c}$ is a real-valued function of ordered pairs $(\underline{h}, \underline{e})$ of sentences in $\mathfrak{L}$, the second member of a pair being required to be noncontradictory. A simple language $\mathfrak{L}$ which Carnap has studied, and which will be convenient for our purposes, has N individual names $\underline{a}_1, \ldots, \underline{a}_N$, a family of n predicates $\underline{P}_1, \ldots, \underline{P}_n$, and the logical connectives "$\sim$" (negation), "$\wedge$" (conjunction), and "$\vee$" (disjunction). A sentence of the form $\underline{P}_j \underline{a}_i$ is an atomic sentence of the language, and it asserts that the individual named by $\underline{a}_i$ has the property designated by $\underline{P}_j$. (Note that we shall use underlined letters to designate linguistic entities—names, predicates, sentences—or else, in the case of $\underline{c}$, a function of sentences.) By calling $\{\underline{P}_j\}$ a "family of predicates" we mean that $\underline{P}_j \underline{a}_i \wedge \underline{P}_k \underline{a}_i$ is logically false for any i and for $j \neq k$, and that $\underline{P}_1 \underline{a}_i \vee \underline{P}_2 \underline{a}_i \vee \cdots \vee \underline{P}_n \underline{a}_i$ is logically true for any i. The function $\underline{c}$ is required to satisfy a number of fairly intuitive conditions, which include the standard axioms of probability theory and also certain symmetry conditions, e.g.,

$\underline{c}(\underline{h} | \underline{e})$ is invariant under permutation of individual names

(Axiom of Symmetry with respect to Individuals), and

$\underline{c}(\underline{h}|\underline{e})$ is invariant under permutation of the predicates
(Axiom of Symmetry with respect to Predicates).

The axioms do not suffice to fix $\underline{c}(\underline{h}|\underline{e})$ for each ordered pair $(\underline{h}, \underline{e})$ of sentences in $\mathfrak{L}$, but they do imply that $\underline{c}$ is one of a family of functions $\{\underline{c}_\lambda\}$, parametrized by a single real variable λ, $0 \le \lambda \le \infty$. The significance of λ is exhibited by considering a sentence $\underline{e}$ which asserts only that in a sample of M individuals, M_j of them have the property designated by P_j. Let $\underline{a}_i$ name an individual which is not in the sample. Then it can be shown from the general conditions on $\underline{c}$-functions that

$$\underline{c}_\lambda(\underline{P}_j\underline{a}_i|\underline{e}) = \frac{M_j + \lambda/n}{N + \lambda} \tag{4.1a}$$

for $0 \le \lambda < \infty$. If we take the limit of Eq. (4.1a) as $\lambda \to \infty$, we obtain

$$\underline{c}_\infty(\underline{P}_j\underline{a}_i|\underline{e}) = 1/n, \tag{4.1b}$$

which is the inverse of the number of predicates in the family $\{\underline{P}_j\}$. An important feature of Eq. (4.1b) is that the probability evaluation is independent of the content of the evidential sentence $\underline{e}$, so long as the individual name $\underline{a}_i$ does not occur in $\underline{e}$. Equations (4.1a), (4.1b) show that λ is an index of the weight placed upon logical considerations (specifically, the number of mutually exclusive and exhaustive predicates in the family) as opposed to empirical ones.

The only background information which will be assumed in this Section is that which is implicit in the rules of the language $\mathfrak{L}$, i.e., the number of individuals and the number of mutually exclusive and exhaustive predicates. We shall follow Carnap's notation by writing prior probability statements (which are statements in which the only evidence is the background information) as $\underline{c}(\underline{h}|\underline{t})$, where $\underline{t}$ is any tautology in $\mathfrak{L}$, and by introducing a prior probability measure function $\underline{m}(\underline{h})$, which is defined as

$$\underline{m}(\underline{h}) = c(\underline{h}|\underline{t}). \tag{4.2}$$

We shall be primarily interested in the prior probability measure functions associated with those confirmation functions which belong to the λ-continuum:

$$\underline{m}_\lambda(\underline{h}) = \underline{c}_\lambda(\underline{h}|\underline{t}). \tag{4.2'}$$

A state description $\underline{D}_N$ is the conjunction of N atomic sentences, one concerning each of the N individuals, so that $\underline{D}_N$ has the form

$$\underline{P}_{j_1}\underline{a}_1 \wedge \underline{P}_{j_2}\underline{a}_2 \cdots \wedge \underline{P}_{j_N}\underline{a}_N.$$

The n-tuple $\{N_1,\ldots,N_n\}$ of a state description $\underline{D}_N$ is the ordered set of n integers such that N_k is the number of times the predicate $\underline{P}_k$ occurs in $\underline{D}_N$. The structure description $\underline{S}_N$ corresponding to a state description $\underline{D}_N$ is the disjunction of all state descriptions having the same n-tuple as $\underline{D}_N$. From combinatorial considerations, the number of state descriptions which are disjoined in $\underline{S}_N$ is

$$\frac{N!}{N_1!\cdots N_2!}.$$

Because of the Axiom of Symmetry with respect to Individuals, together with the standard probability axioms, we have

$$\underline{m}_\lambda(\underline{S}_N) = \frac{N!}{N_1!\cdots N_n!}\underline{m}_\lambda(\underline{D}_N). \tag{4.3}$$

A slight modification of an argument of Carnap [3, p. 31] yields

$$\underline{m}_\lambda(\underline{D}_N) = \frac{\prod_{j=1}^{n}\left\{\frac{\lambda}{n}\left(\frac{\lambda}{n}+1\right)\cdots\left(\frac{\lambda}{n}+N_j-1\right)\right\}}{\lambda(\lambda+1)\cdots(\lambda+N-1)}, \tag{4.4a}$$

where the expression in curly brackets is taken to be 1 if $N_j = 0$. Equivalently,

$$\underline{m}_\lambda(\underline{D}_N) = \frac{\Gamma(\lambda)}{\Gamma(\lambda+N)}\prod_{j=1}^{n}\left\{\frac{\Gamma(\lambda/n+N_j)}{\Gamma(\lambda/n)}\right\}, \tag{4.4b}$$

where Γ is the gamma function.

In order to set up a situation to which both Jaynes' and Carnap's methods are applicable, let us interpret the predicates $\underline{P}_1,\underline{P}_2,\ldots,\underline{P}_n$ as meaning "having the value E_1 of some quantity E," "having the value E_2 of some quantity E,"..., "having the value E_n of some quantity E." The type of evidence $\hat{d}_\epsilon$ for which Jaynes's method is designed is a statement of the numerical value of the average of E, specifically that this average is ϵ. Hence $\hat{d}_\epsilon$ must be such that

$$\sum_{i=1}^{n} p(h_i|\hat{d}_\epsilon)E_i = \epsilon, \tag{4.5}$$

where h_i is the hypothesis that a randomly chosen individual from the population has value E_i of the quantity E. Jaynes does not specify the character of the evidence $\hat{d}_\epsilon$ which he would consider to be suitable for the applicability of his method, and it is very likely that most types of evidence which he would admit are not expressible in extremely simple languages like Carnap's $\mathfrak{L}$. Fortunately for our purpose, however, there is one candidate for $\underline{\hat{d}}_\epsilon$ which appears suitable from Jaynes' point of view and which is also expressible in $\mathfrak{L}$: specifically, the disjunction of all state descriptions $\underline{D}_N$ having n-tuples $\{N_1,\ldots,N_n\}$ such that

$$\sum_{i=1}^{n} \frac{N_i E_i}{N} = \epsilon. \tag{4.6}$$

For a given exact ϵ it is sometimes possible to find more than one n-tuple satisfying Eq. (4.6); and if the average ϵ is only given within an interval (ϵ_1,ϵ_2), it is always possible to find more than one n-tuple satisfying Eq. (4.6) if N is sufficiently large. In order to illustrate the first half of the preceding sentence, we consider $n = 3$, $E_i = i$, N an even integer, and $\epsilon = 2$. Then the following n-tuples all satisfy Eq. (4.6):

$$\begin{aligned}\{N_1,N_2,N_3\} &= \{0,N,0\}\\ &= \{1,N-2,1\}\\ &= \{2,N-4,2\}\\ &\ \ \vdots\\ &= \{N/2,0,N/2\}.\end{aligned} \tag{4.7}$$

The evidence $\underline{\hat{d}}_2$, which is the disjunction of all the state descriptions having the n-tuples listed in (4.7), clearly does not specify the frequency of occurrence of each $\underline{P}_i$ in the population. We shall make use of this simple illustration throughout the remainder of Section IV, even though some of the results could be generalized.

It is straightforward to show from the axioms of probability, which all of the functions $\underline{c}_\lambda$ satisfy, that

$$\underline{c}_\lambda\left(\underline{h}_j|\underline{\hat{d}}_\epsilon\right) = \frac{\displaystyle\sum_{\{N_1,\ldots,N_n\}}' \frac{N_j}{N}\underline{m}_\lambda(\underline{S}_N)}{\displaystyle\sum_{\{N_1,\ldots,N_n\}}' \underline{m}_\lambda(\underline{S}_N)}$$

$$= \frac{\displaystyle\sum_{\{N_1,\ldots,N_n\}}' \frac{N_j}{N}\frac{N!}{\prod_{i=1}^{n} N_i!}\underline{m}_\lambda(\underline{D}_N)}{\displaystyle\sum_{\{N_1,\ldots,N_n\}}' \frac{N!}{\prod_{i=1}^{n} N_i!}\underline{m}_\lambda(\underline{D}_N)}$$

$$= \frac{\displaystyle\sum_{\{N_1,\ldots,N_n\}}' \frac{N_j}{N} \frac{N!}{\prod_{i=1}^{n} N_i!} \prod_{j=1}^{n} \frac{\left\{\frac{\lambda}{n}\left(\frac{\lambda}{n}+1\right)\cdots\left(\frac{\lambda}{n}+N_j-1\right)\right\}}{\lambda(\lambda+1)\cdots(\lambda+N-1)}}{\displaystyle\sum_{\{N_1,\ldots,N_n\}}' \frac{N!}{\prod_{i=1}^{n} N_i!} \prod_{j=1}^{n} \frac{\left\{\frac{\lambda}{n}\left(\frac{\lambda}{n}+1\right)\cdots\left(\frac{\lambda}{n}+N_j-1\right)\right\}}{\lambda(\lambda+1)\cdots(\lambda+N-1)}}, \tag{4.8}$$

where Σ' designates summation over the n-tuples which satisfy Eq. (4.6). Actually, Eq. (4.8) concerns an idealized situation, since for arbitrary ϵ (for example, any irrational number) there may be no n-tuples satisfying Eq. (4.6). For greater applicability, we could replace Eq. (4.8) by a similar equation for $\underline{c}_\lambda\left(\underline{h}_j | \underline{d}_{(\epsilon_1,\epsilon_2)}\right)$, where $\underline{\hat{d}}_{(\epsilon_1,\epsilon_2)}$ is the evidence that the average ϵ of E falls in the interval (ϵ_1, ϵ_2), and the summation Σ' would be over n-tuples which make the ϵ of Eq. (4.6) fall in this interval. By restricting our attention to the case of $\epsilon = 2$, for which there is an abundance of n-tuples (actually triples) satisfying Eq. (4.6), we can use Eq. (4.8) as it stands, thereby simplifying the calculations somewhat. In the case of $\epsilon = 2$, Jaynes' method yields

$$p\left(\underline{h}_i | \underline{\hat{d}}_2\right) = \frac{e^{-\beta(2)i}}{e^{-\beta(2)} + e^{-2\beta(2)} + e^{-3\beta(2)}} = \frac{1}{3}, \tag{4.9}$$

since the value of the parameter β which makes $\epsilon = 2$ is obviously $\beta(2) = 0$. We now assert the following

THEOREM. (i) *For every nonnegative real number λ there exist a $\delta_\lambda > 0$ and an integer N_λ such that $\left|\underline{c}_\lambda\left(\underline{h}_j | \underline{\hat{d}}_2\right) - \frac{1}{3}\right| > \delta_\lambda$ if the number of individuals N of the language $\mathfrak{L}$ is greater than N_λ, where $j = 1,2,3$.* (ii) *For $\lambda = \infty$ we have*

$$\lim_{N\to\infty} \underline{c}_\infty\left(\underline{h}_j | \underline{\hat{d}}_2\right) = \tfrac{1}{3}.$$

Furthermore, for all ϵ in [1, 3],

$$\lim_{N\to\infty} \underline{c}_\infty\left(\underline{h}_j | \underline{\hat{d}}_{(\epsilon - N^{-1/2}, \epsilon + N^{-1/2})}\right) = \frac{e^{-j\beta}}{e^{-\beta} + e^{-2\beta} + e^{-3\beta}}, \tag{4.10}$$

where $\beta = \beta(\epsilon)$ is the parameter which makes the expectation value of E equal to ϵ in Eq. (4.5).

The proof of this theorem is rather lengthy, and therefore we have relegated it to Appendix B. Most of the remainder of the present section will be concerned with the significance of the theorem.

First of all, we wish to point out that the theorem does not simply identify Jaynes' method with one of Carnap's continuum of inductive methods (specifically, with the one in which $\lambda = \infty$). One cannot reasonably expect a simple identification, because of the great differences between their approaches. Carnap restricts his attention to a very simple language $\mathfrak{L}$, but within it he admits any noncontradictory sentence as an evidential sentence. Jaynes does not specify the class of languages within which his evidence is expressed, but his method is most directly applicable to evidence which can be cast in the form of assigning a numerical value to the average of some variable quantity. Our strategy has been to investigate a case to which both Carnap's and Jaynes' methods apply. It is clear that the sentences $\underline{\hat{d}}_\epsilon$ discussed above can be expressed in Carnap's language $\mathfrak{L}$, and we also believe that $\underline{\hat{d}}_\epsilon$ signifies an evidential proposition which Jaynes would accept as grounds for assigning the value ϵ to the average of E. Once $\underline{\hat{d}}_\epsilon$ is accepted as signifying evidence that both Carnap and Jaynes can use, then we can say on the basis of the above theorem that Jaynes' method disagrees with every one of the methods in the λ-continuum with finite λ, at least when $\epsilon = 2$. The discrepancy between $\underline{c}_\lambda(\underline{h}_2|\underline{\hat{d}}_2)$ and Jaynes' value of p_2 when $\epsilon = 2$, namely $\frac{1}{3}$, does not go to 0 with increasing N. On the other hand, in the particular problem that has been set up, the probability evaluations made by Jaynes' method and those made by Carnap's $\underline{c}_\infty$ are in closer and closer agreement as N increases. Clearly, then, $\underline{c}_\infty$ comes nearest to Jaynes' method among all those in Carnap's continuum of inductive methods.

There are, of course, other inductive methods than those of Carnap, and indeed we are not committed to remaining within the family of methods which he studied. (The formality of Carnap's approach was criticized in Shimony [26] and a more informal and flexible method was recommended there.) Nevertheless, we feel that the exhibition of convergent agreement between Jaynes' probability evaluations and those of c_∞ is very significant. It is particularly illuminating concerning the anomaly exhibited in the theorem of FS and in the theorems of Sections II and III, to the effect that a necessary condition for the consistency of Jaynes' prior and posterior probabilities with each other and with the general rules of probability theory is that there is a prior probability 1 of finding the value of the parameter β to be 0. This is precisely the probability that we find in the limit of $N \to \infty$ if we use the confirmation function $\underline{c}_\infty$. We shall now give a proof of this claim.

We need to evaluate $\underline{c}_\infty(\underline{\hat{d}}_{(\epsilon_1,\epsilon_2)}|\underline{t})$, where $\underline{t}$ is a tautology, for arbitrarily small intervals (ϵ_1, ϵ_2) about the value $\epsilon = 2$. The easiest way to do this is to treat the value of E assigned to the ith individual in the population as a random variable X^i with possible values 1, 2, 3, and we can suppose that the evaluation of $X^1, \ldots, X^N$ takes place sequentially. Carnap shows that

$$\underline{c}_\infty(\underline{P}_j\underline{a}_i|\underline{D}_{i-1}) = \underline{c}_\infty(\underline{P}_j\underline{a}_i|\underline{t}) = \tfrac{1}{3}, \tag{4.11}$$

if $\underline{D}_{i-1}$ is a state description restricted to the first $i-1$ individuals [3, p. 37]. Even if $\underline{D}_{i-1}$ says that all of the first $i-1$ individuals have the property designated by $\underline{P}_1$, $\underline{c}_\infty$ assigns exactly the same probability $\frac{1}{3}$ to the sentence $\underline{P}_1\underline{a}_i$ as to $\underline{P}_2\underline{a}_i$ and $\underline{P}_3\underline{a}_i$. This statement will seem less surprising if one recalls that λ is a weight placed upon logical considerations (here, the number of mutually exclusive and exhaustive predicates) as opposed to empirical considerations (the evidence about the first $i-1$ individuals). Since $\lambda = \infty$ places all weight upon logical considerations, $\underline{D}_{i-1}$ is irrelevant to the probability of sentences involving only the ith individual. If we now express Eq. (4.11) in terms of the random variables $X^1, \ldots, X^N$, we may say that these are independent random variables according to $\underline{c}_\infty$, each with mean $\mu = 2$ and variance $\sigma^2 = (1-2)^2/3 + (2-2)^2/3 + (2-3)^2/3 = 2/3$. We are now in a position to apply the Central Limit Theorem [6]

$$\lim_{N\to\infty} P\left\{\frac{|\mathcal{S}_N - N\mu|}{\sigma N^{1/2}} < \eta\right\} = \Phi(\eta), \tag{4.12}$$

where $\mathcal{S}_N = X^1 + \cdots + X^N$, η is any positive real number, and

$$\Phi(\eta) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\eta} e^{-y^2/2} dy.$$

We may rewrite Eq. (4.12) in terms of the population average $\mathcal{S}_N/N$ as

$$\lim_{N\to\infty} P\left\{|\mathcal{S}_n/N - 2| < \sqrt{\frac{2}{3N}}\,\eta\right\} = \Phi(\eta). \tag{4.13}$$

But $P\{|\mathcal{S}_N - 2| < \sqrt{2/3N}\,\eta\}$ is just an alternative notation which we have introduced for $\underline{c}_\infty(\underline{\hat{d}}_{(2-\sqrt{2/3N}\,\eta,\,2+\sqrt{2/3N}\,\eta)}|\underline{t})$. Using the fact that $\epsilon = 2$ corresponds to $\beta = 0$ and that

$$\left.\frac{d\epsilon}{d\beta}\right|_{\beta=0} = -\frac{8}{3},$$

(from Eq. (2.6)), we have

$$\lim_{N\to\infty} \underline{c}_\infty(\underline{d}_{(-\sqrt{3/32N}\,\eta,\,\sqrt{3/32N}\,\eta)}|\underline{t}) = \Phi(\eta), \tag{4.14}$$

where the sentence $\underline{d}_{(\,,\,)}$ concerns the parameter β, as explained in Section I. Since $\Phi(\eta)$ converges to 1 as $\eta \to \infty$, we can decide antecedently how close to 1 we wish $\underline{c}_\infty(\underline{d}_{(-\sqrt{3/32N}\,\eta,\sqrt{3/32N}\,\eta)} | \underline{t})$ to be, and then we make the interval about $\beta = 0$ as small as we wish by choosing N large enough. Roughly speaking, then, in the limit of $N \to \infty$, there is a prior probability of 1 of finding β to be 0, if the confirmation function $\underline{c}_\infty$ is used.

Some one may be tempted to present the foregoing calculation as a vindication of Jaynes' program, since it shows that the assignment of prior probability 1 to the value $\beta = 0$ is not just a peculiar by-product of comparing prior and posterior probabilities computed on the basis of the maximum entropy principle, but rather the natural result of an independently developed inductive method. The price of this vindication, however, is that one accept the confirmation function $\underline{c}_\infty$ as a reasonable function to employ in calculating probabilities upon which decisions and actions presumably would be based. But we saw, in the preceding discussion, that $\underline{c}_\infty$ has the idiosyncratic property of regarding all evidence concerning other individuals as irrelevant to a particular individual of interest, and it therefore is entirely unsuitable as an intellectual instrument for obtaining reliable guidance from experience. Furthermore, the question of whether it is reasonable to accept as virtually certain that the parameter β is 0, given only the background information, can be considered in detachment from the properties of $\underline{c}_\infty$. Even if we make an effort to divest ourselves of all evidence not contained in this background information, such as the evidence that a variety of heat baths with differing temperatures and hence differing reservoir parameters can be found, we still regard the assignment of probability 1 to $\beta = 0$ and of probability 0 to the entire remaining range of values of β to be completely counterintuitive. This assignment seems like too much knowledge to extract out of a situation of presumed ignorance, and indeed it goes directly counter to Jaynes' sensible maxim that we should be honest about the extent of our ignorance.

There is one further characteristic of $\underline{c}_\infty$ which is relevant to Jaynes' method: that is, that $\underline{c}_\infty$ assigns equal prior probability to all state descriptions:

$$\underline{c}_\infty(\underline{D}_N | \underline{t}) = (\text{number of state descriptions})^{-1} = n^{-N}. \tag{4.15}$$

A corollary is that if evidence $\underline{e}$ is considered which is consistent with a subclass $\mathfrak{D}$ of all the state descriptions and inconsistent with all the others, then if $\underline{D}_N$ belongs to $\mathfrak{D}$,

$$\underline{c}_\infty(\underline{D}_N | \underline{e}) = (\text{Number of members of } \mathfrak{D})^{-1}. \tag{4.16}$$

Equation (4.16) is reminiscent of one of the common procedures of statistical mechanics, the use of the microcanonical ensemble, which assigns equal weights to all states in an energy hypersurface. It is well known that once one has the microcanonical ensemble to represent an isolated system, one can straightforwardly derive the canonical distribution to represent a system in contact with a heat bath: one lets the system and the heat bath together constitute an isolated system, characterized by the microcanonical distributions, and then combinatorial considerations lead to the desired conclusion. In this way one comes by the standard route to the canonical distribution, which Jaynes derives by means of his maximum entropy principle. Does this consideration provide a vindication of Jaynes' method —or even somewhat more than a vindication, for his method arrives at the conclusion more efficiently than do the usual statistical mechanical arguments?

The answer, we believe, is negative. It is only in rather special situations that one can assert the equiprobability of all states of a given class in statistical physics, and when that can be done there is a physical reason, such as ergodicity or an appropriate symmetry. If one is employing the logical concept of probability, such a reason should be included as part of the evidential proposition. In the absence of such a reason, one may try to justify equiprobability on purely logical or epistemological grounds, perhaps by some version of the Principle of Indifference; but there are notorious ambiguities and paradoxes which plague such a program [15, 19]. In any case, the assignment of equal prior probabilities to state descriptions, as in Eq. (4.15), cannot be justified on epistemological grounds, for if it could, then the confirmation function $\underline{c}_{\infty}$ would be the appropriate one to use in inductive logic, and we saw above that this cannot be correct.

We conclude that Jaynes is mistaken in his program of founding statistical mechanics upon epistemology. His central principle, the maximum entropy principle, cannot be universally true. It may be true under special circumstances, and we have given reasons to believe that these are the circumstances in which it is appropriate to employ something like the c-function $\underline{c}_{\infty}$. The locution "something like" is needed partly because Carnap only defined $\underline{c}_{\infty}$ for a rather special class of languages, and for the purpose of formulating statistical physics one would need to use a language richer than any of these; and partly because inductive logicians may have to abandon as hopelessly idealistic the program of choosing a single confirmation function to be applied to all the admissible sentences in a given language, and use instead confirmation functions which are relativized to special bodies of evidence. In a relativized application of confirmation function one might set aside Eq. (4.15), but still use Eq. (4.16) if the evidence $\underline{e}$ is suitable. What might serve as "suitable" evidence $\underline{e}$? One candidate might be a large body of data concerning a gambling device, indicating that the outcomes of individual plays can reasonably be taken to be independent random variables. Another candidate is a physical characterization of a many-body system which is sufficient to permit ergodicity to be proved, as Sinai [28] succeeded in doing for a box of hard spheres governed by Newtonian dynamics. Another candidate is the mechanism of memory loss postulated by Tisza and Quay [30, p. 54]. With such $\underset{\sim}{e}$ Eq. (4.16) might be justified, and then one could proceed to derive the standard formulae of statistical mechanics along one of several well-known lines. And, as a bonus, one would have a justification for a *relativized* use of the maximum entropy principle, which would permit the derivation of the standard formulae in Jaynes' manner. It would be wrong, however, to say that Jaynes' method has greater generality or proceeds from weaker assumptions than other, more overtly physical methods.

V. Replies to Some Criticisms

Jaynes [17] stated a number of criticisms of FS, which presumably he would apply to the present paper.

He says first that the evidential propositions $\hat{d}_\epsilon$ considered by FS are "ill-defined." That is true, but it is only to achieve generality. FS let "$\hat{d}_\epsilon$ be the evidence that the expected value of E is ϵ," and their reasoning goes through no matter what this evidence is taken to be.

Jaynes' central objection is the following, which will be quoted at length. ("PME" is an abbreviation for "Principle of Maximum Entropy.")

> If a statement d referring to a probability distribution in space S is testable (for example, if it specifies a mean value $\langle f \rangle$ for some function $f(i)$ defined on S), then it can be used as a constraint in PME; but it cannot be used as a conditioning statement in Bayes' theorem because it is not a statement about any event in S or any other space.
>
> Conversely, a statement D about an event in the space S'' (for example, an observed frequency) can be used as a conditioning statement in applying Bayes' theorem, whereupon it yields a posterior distribution on S'' which may be contracted to a marginal distribution on S; but D cannot be used as a constraint in applying PME in space S, because it is not a statement about any event in S, or about any probability distribution over S; i.e., it is not testable information in S." [17, p. 54]

This objection raises some issues about the language which is appropriate in probability theory. We agree with Jaynes that a probability statement (e.g., "The probability of hypothesis h on evidence e is $\frac{1}{4}$") is not a statement about an event, since the concept of probability which he is using, and which we accept, is that of reasonable degree of belief. The evidence e, however, does concern an event, which can be taken to be a point in an evidence space. Since the $\hat{d}_\epsilon$ which FS discuss is a body of evidence, it does concern an event. Because of the generality of the discussion of FS, they do not specify the evidence explicitly, but rather refer to it by the way in which it will be used in the maximum entropy prescription: i.e., it is that evidence which will warrant taking ϵ to be the expected value of E. The objection may be raised at this point that this mode of reference implicitly refers to a probability distribution, since the expected value of E is $\Sigma E_i p_i$; since probability statements are not about events, it would then follow that $\hat{d}_\epsilon$ also is not about an event. Our answer is that the mode of reference must be carefully distinguished from the thing referred to. The phrases "the Evening Star" and "the Morning Star" refer to the same entity, but in different ways; in Frege's terminology, they have the same "Bedeutung" but different "Sinn" [8, pp. ix, 56–78). An event can be referred to in different ways: by a name, by a description of its intrinsic properties, and by indirect descriptions, such as a phrase which contains a probabilistic expression like "expected value."

In Section IV of the present paper we make the general considerations of FS more concrete by proposing specific candidates for the evidence $\hat{d}_\epsilon$ and for the sentence $\underline{\hat{d}}_\epsilon$ which expresses that evidence. In the passage quoted above [17, p. 54] there seems to be a claim that one and the same sentence cannot be used as a constraint in PME and as a conditioning statement in Bayes' theorem, because sentences performing these two functions must belong to different kinds of language. However, the sentence $\underline{\hat{d}}_\epsilon$ of Section IV does indeed perform both functions. Since it is a noncontradictory sentence in the language $\mathfrak{L}$, it can be used as an evidential statement in any of Carnap's confirmation functions $\underline{c}_\lambda$, all of which conform to general Bayesian principles. But $\underline{\hat{d}}_\epsilon$ is also a testable statement which seems to determine the expected value of E, so that it can be used in an application of the maximum entropy principle; it certainly resembles closely some of the sentences which Jaynes himself deploys in his illustrations of the principle (e.g., [13, pp. 183–187]).

The exhibition of the sentence $\underline{\hat{d}}_\epsilon$ in Section IV also provides a positive answer to a question of Cyranski [5, p. 298], "whether evidence is a proposition in the same 'language' as is the hypothesis, at least when 'evidence' is of the form required by the MEP." Cyranski gives a negative answer to this question and thus essentially offers the same objection to FS as does Jaynes, though in greater detail.

Jaynes says that "informed students of statistical mechanics will be astonished at the suggestion that there is any inconsistency between application of PME in space S and of Bayes' theorem in S^n, since the former yields a canonical distribution, while the latter is just the Darwin–Fowler method, originally introduced as a rigorous way of justifying the canonical distribution!" [17, p. 54]. This is a reasonable point. But it is just because of this kind of consideration that we have tried in Section IV to determine precisely under what circumstances Jaynes's methods are legitimate and agree with more conventional Bayesian methods. Our conclusion was that the requisite circumstances are quite special and are physical rather than epistemological in nature.

Finally, Jaynes remarks that FS wrote, between them, three critical articles, but never indicated what their positive preferences are. That is true, and we shall briefly indicate some preferences here. In foundations of statistical mechanics we are impressed by the success of Tisza and Quay [30] in deriving statistical thermodynamics from a very moderate phenomenological assumption of memory loss and by the related method of Mandelbrot [21]. In foundations of inductive logic one of us has developed an informal Bayesian theory [26], but modifications of it are envisaged.

APPENDIX A: Proof of FS Theorem

Let

$$g(E)=\int_{-\infty}^{\infty}\frac{e^{-\beta E}}{\sum_{j=1}^{n}e^{-\beta E_j}}d\mu, \qquad (A.1)$$

Then

$$\frac{dg}{dE}=\int_{-\infty}^{\infty}\frac{-\beta e^{-\beta E}}{\sum_{j=1}^{n}e^{-\beta E_j}}d\mu \qquad (A.2)$$

is a monotonically increasing function of E unless μ is concentrated entirely at 0, because the integrand

$$\frac{-\beta e^{-\beta E}}{\sum e^{-\beta E_j}}$$

is monotonically increasing for $\beta \neq 0$. But a function on R which everywhere has a monotonically increasing derivative cannot be equal to a given number for more than two distinct values of its argument. Hence, if there are three or more distinct values among $E_1,\ldots,E_n$ then

$$\frac{1}{n}=\int_{-\infty}^{\infty}\frac{e^{-\beta E_i}}{\sum_{j=1}^{n}e^{-\beta E_j}}d\mu, \qquad i=1,\ldots,n \qquad (A.3)$$

(which is equivalent to Eq. (1.10), implies

$$\mu(\{0\})=1, \qquad \mu(R-\{0\})=0. \qquad \text{Q.E.D.} \quad (A.4)$$

APPENDIX B: PROOF OF THE THEOREM OF SECTION IV

Taking the limit $\lambda \to \infty$ in Eq. (4.9) yields

$$c_\infty\left(\underline{h}_j \mid \hat{\underline{d}}_2\right) = \frac{\sum' \frac{N_j}{N} \frac{N!}{N_1!N_2!N_3!}}{\sum' \frac{N!}{N_1!N_2!N_3!}}, \tag{B.1}$$

where the sums are over the triples of (4.7). Eq. (B.1) can also be obtained from Eq. (4.16). The evaluation of the rhs of Eq. (B.1) is a standard problem in statistical mechanics, which can be solved, for example, by the method of Darwin and Fowler (see, for example, [23]). The result is

$$\underline{c}_\infty\left(\underline{h}_j \mid \hat{\underline{d}}_2\right) \approx \frac{z_0^{E_j}}{\sum_{i=1}^{3} z_0^{E_i}}, \tag{B.2}$$

where $z_0 = e^{-\beta}$. Since $\epsilon = 2$ implies that $\beta = 0$, we find

$$\underline{c}_\infty\left(\underline{h}_j \mid \hat{\underline{d}}_2\right) \approx \tfrac{1}{3}, \tag{B.3}$$

in good agreement with Jaynes's result in Eq. (4.9). (Throughout Appendix B we shall use expressions of the form "$F \approx G$" to mean that there is a function $\omega(N)$ which converges to 0 as N goes to ∞, and $F = G(1 + \omega(N))$. Similarly "$F \gtrsim G$" will mean "$F > G(1 + \omega(N))$".) It is clear that we can also obtain a generalization of Eq. (B.2),

$$\underline{c}_\infty\left(\underline{h}_j \mid \underline{d}_{(\epsilon - o(1/N),\, \epsilon + o(1/N))}\right) \approx \frac{z_0^{E_j}}{\sum_{i=1}^{n} z_0^{E_i}} \tag{B.2$'$}$$

for any number n of hypotheses $\underline{h}_1, \ldots, \underline{h}_n$ and any values of the $E_1, \ldots, E_n$, provided that the summations occurring in the rhs of Eq. (B.1) are replaced by sums over n-tuples which yield averages of E in the interval $(\epsilon - o(1/N), \epsilon + o(1/N))$.

We shall not present of derivation of Eq. (B.2), since the Darwin–Fowler method is both standard and lengthy, and since the result Eq. (B.3) is also the immediate result of Eq. (B.32), which will be derived below.

We are not able to treat all the finite values of λ by a uniform method, and therefore we consider the following subcases: $\lambda = 0, 0 < \lambda < \frac{3}{2}, \lambda = \frac{3}{2}, \frac{3}{2} < \lambda < 3, \lambda = 3, 3 < \lambda < \infty$.

For $\lambda = 0$ we have

$$\underline{c}_0(\underline{D}_N|\underline{t}) = \lim_{\lambda\to 0} \underline{c}_\lambda(\underline{D}_N|\underline{t}) = \lim_{\lambda\to 0} \frac{\prod_{j=1}^{n}\left\{\frac{\lambda}{n}\left(\frac{\lambda}{n}+1\right)\cdots\left(\frac{\lambda}{n}+N_j-1\right)\right\}}{\lambda(\lambda+1)\cdots(\lambda+N-1)}$$

$$= \lim_{\lambda\to 0}\left(\frac{\lambda}{n}\right)^p \frac{1}{\lambda},$$

where p is the number of predicates $\underline{P}_j$ which characterize at least one individual in $\underline{D}_N$. Clearly,

$$\underline{c}_0(\underline{D}_N|\underline{t}) = 0 \qquad \text{if } p \neq 1,$$
$$\underline{c}_0(\underline{D}_N|\underline{t}) = 1/n \qquad \text{if } p = 1.$$

When $\epsilon = 2$, the only $\underline{D}_N$ with $p = 1$ is the one in which $N_2 = N, N_1 = N_3 = 0$. Hence Eq. (4.8) implies

$$\underline{c}_0(\underline{h}_1|\underline{\hat{d}}_2) = \underline{c}_0(\underline{h}_3|\underline{\hat{d}}_2) = 0, \tag{B.4a}$$

$$\underline{c}_0(\underline{h}_2|\underline{\hat{d}}_2) = 1. \tag{B.4b}$$

In the case of $\lambda = 3$ we have

$$\prod_{j=1}^{3}\left\{\frac{\lambda}{3}\left(\frac{\lambda}{3}+1\right)\cdots\left(\frac{\lambda}{3}+N_j-1\right)\right\} = \prod_{j=1}^{3} N_j!.$$

Hence Eq. (4.8) yields

$$\underline{c}_3(\underline{h}_1|\underline{\hat{d}}_2) = \underline{c}_3(\underline{h}_3|\underline{\hat{d}}_2) = \frac{\sum_{N_1=0}^{N/2} N_1/N}{\sum_{N_1=0}^{N/2} 1} = \frac{\frac{1}{2}\cdot\frac{1}{2}(N/2+1)}{N/2+1} = \tfrac{1}{4}. \tag{B.5a}$$

Likewise

$$\underline{C}_3(\underline{h}_2|\underline{\hat{d}}_2) = \tfrac{1}{2}. \tag{B.5b}$$

Some preliminary remarks and calculations will be useful for the remaining cases, which are more difficult.

(i) The conditions $N_1 + N_2 + N_3 = N$ and $\epsilon = 2$ imply

$$N_1 = N_3 = (N - N_2)/2. \tag{B.6}$$

Hence, the summation Σ' over triples in Eq. (4.8) becomes a summation over N_2 from 0 to N in steps of 2, and Eq. (4.8) becomes (for the case of $j = 2$)

$$\underset{\sim}{c}_\lambda(\underset{\sim}{h}_2|\underset{\sim}{d}_2) = \frac{\displaystyle\sum_{N_2=0}^{N}{}' \frac{N_2}{N}\left[\frac{\frac{\lambda}{3}\left(\frac{\lambda}{3}+1\right)\cdots\left(\frac{\lambda}{3}+N_2-1\right)}{N_2!}\right] \times \left[\frac{\frac{\lambda}{3}\left(\frac{\lambda}{3}+1\right)\cdots\left(\frac{\lambda}{3}+\frac{N-N_2}{2}\right)}{[(N-N_2)/2]!}\right]^2}{\displaystyle\sum_{N_2=0}^{N}{}' \left[\frac{\frac{\lambda}{3}\left(\frac{\lambda}{3}+1\right)\cdots\left(\frac{\lambda}{3}+N_2-1\right)}{N_2!}\right] \times \left[\frac{\frac{\lambda}{3}\left(\frac{\lambda}{3}+1\right)\cdots\left(\frac{\lambda}{3}+\frac{N-N_2}{2}\right)}{[(N-N_2)/2]!}\right]^2}$$

$$\equiv \frac{\displaystyle\sum_{N_2=0}^{N}{}' \frac{N_2}{N}\Omega(N_2)}{\displaystyle\sum_{N_2=0}^{N}{}' \Omega(N_2)}, \tag{B.7}$$

where the notation $\Omega(N_2)$ is defined by its context. To simplify the notation in the following calculations we assume that N is even, and the summation Σ' is only over even values of N_2.

(ii) One would like to use Stirling's formula in evaluating the rhs of Eq. (B.7), but there is a difficulty in the fact that values of N_2 close to 0 and close to N might contribute strongly to the sums, and for these values the error due to using Stirling's formula for $N_2!$ and $(N - N_2)!$ can be large. Hence, we use the strategy of dividing the integers between 0 and N into three classes:

(a) $0 \leq N_2 \leq M$, so that $(N - M)/2 \leq (N - N_2)/2 \leq N/2$,
(b) $M \leq N_2 \leq N - 2M$, so that $M \leq (N - N_2)/2 \leq (N - M)/2$,
(c) $N - 2M \leq N_2 \leq N$, so that $0 \leq (N - N_2)/2 \leq M$,

where we take

$$M = N^{(1-\alpha)}, \qquad 0 < \alpha < 1. \tag{B.8}$$

Our strategy will be to choose α close enough to 1 and yet N large enough that Stirling's formula will give an excellent approximation to $N_2!$ and $(N - N_2)!$ for all N_2 in class (b), and yet classes (a) and (c) have many fewer members than class (b). We shall then be able to show that if $\frac{3}{2} < \lambda < \infty$ the overwhelming contribution to the rhs of Eq. (B.7) comes from class (b); for $0 < \lambda < \frac{3}{2}$ the overwhelming contribution comes from class (c); and for $\lambda = \frac{3}{2}$ the overwhelming contribution comes from the union of (b) and (c), and the relative contributions from these two classes need not be determined.

(iii) Stirling's formula asserts that

$$\Gamma(x) \approx x^x(2\pi x)^{1/2}e^{-x}. \tag{B.9}$$

If $N \gg \alpha$ then another useful approximation is

$$\left(1 + \frac{\alpha}{N}\right)^{N+1/2} \approx e^{\alpha}, \tag{B.10}$$

and therefore for $N_j \gg 1, N_j \gg \lambda$ we have

$$\frac{\frac{\lambda}{3}\cdots\left(\frac{\lambda}{3} + N_j - 1\right)}{N_j!} = \frac{\Gamma\left(\frac{\lambda}{3} + N_j\right)}{\Gamma\left(\frac{\lambda}{3}\right)\Gamma(1 + N_j)} \approx \frac{1}{\Gamma\left(\frac{\lambda}{3}\right)} N_j^{\lambda/3-1}. \tag{B.11}$$

Hence the function $\Omega(N_2)$ introduced in Eq. (B.7) satisfies

$$\Omega(N_2) \approx \left[\Gamma\left(\frac{\lambda}{3}\right)\right]^{-1} N_2^{\lambda/3-1}\left(\frac{N - N_2}{2}\right)^{2(\lambda/3-1)}. \tag{B.12}$$

We shall freely use this expression whenever N_2 is in class (b), i.e., $M \leq N_2 \leq N - 2M$. For large N we can replace the summation over terms of class (b) by an integral:

$$\sum_{N_2=M}^{N-2M} \Omega(N_2) \approx \frac{1}{2}\int_M^{N-2M} \Omega(N_2)\,dN_2, \tag{B.13}$$

$$\sum_{N_2=M}^{N-2M} \frac{N_2}{N}\Omega(N_2) \approx \frac{1}{2}\int_M^{N-2M} \frac{N_2}{N}\Omega(N_2)\,dN_2, \tag{B.14}$$

where the factor $\frac{1}{2}$ is due to the fact that the summations proceed in steps of 2. If we introduce the notation

$$y \equiv N_2/N,$$

$$D \equiv \int_{N^{-\alpha}}^{1-2N^{-\alpha}} y^{\lambda/3-1}(1-y)^{2(\lambda/3-1)}dy, \tag{B.15}$$

$$I_2 \equiv \int_{N^{-\alpha}}^{1-2N^{-\alpha}} y^{\lambda/3}(1-y)^{2(\lambda/3-1)}dy, \tag{B.16}$$

$$\phi \equiv \left[\Gamma\left(\frac{\lambda}{3}\right)\right]^{-1} \frac{N^{\lambda-2}}{2^{2\lambda/3-1}}, \tag{B.17}$$

then

$$\sum_{N_2=M}^{N-2M} \Omega(N_2) \approx \phi D, \tag{B.18}$$

and

$$\sum_{N_2=M}^{N-2M} \frac{N_2}{N}\Omega(N_2) \approx \phi I_2. \tag{B.19}$$

Observing that

$$D - I_2 = \int_{N^{-\alpha}}^{1-2N^{-\alpha}} y^{\lambda/3-1}(1-y)^{2(\lambda/3-1)+1}dy$$

and integrating by parts, we obtain

$$I_2 \approx \frac{\lambda}{3(\lambda-1)}D - \frac{1}{\lambda-1}\Big[(1-2N^{-\alpha})^{\lambda/3}(2N^{-\alpha})^{2\lambda/3-1}$$
$$- (N^{-\alpha})^{\lambda/3}(1-N^{-\alpha})^{2\lambda/3-1}\Big]. \tag{B.20}$$

In order to handle the values of N_2 in classes (a) and (c) we use the following

LEMMA. *If* $\lambda > 3$, *then* $\Omega(N_2)$ *is monotonically increasing in* $0 \leq N_2 \leq M$ *and monotonically decreasing in* $N - 2M \leq N_2 \leq N$, *for proper choices of* α *and sufficiently large* N. *If* $\lambda < 3$, $\Omega(N_2)$ *is a monotonically decreasing function for* $0 \leq N_2 \leq M$ *and monotonically increasing for* $N - 2M \leq N_2 \leq N$, *for proper choices of* α *and sufficiently large* N.

Proof.

$$\frac{\Omega(N_2)}{\Omega(N_2+2)} = \frac{\left\{\left[\frac{\lambda}{3}+\frac{1}{2}(N-N_2)-1\right]\Big/\frac{1}{2}(N-N_2)\right\}^2}{\dfrac{(\lambda/3+N_2)(\lambda/3+N_2+1)}{(N_2+1)(N_2+2)}}. \tag{B.21}$$

Choose N and α large enough that $0 < M < (N-4)/3$, which ensures $\frac{1}{2}(N-M) > M+2$ and $M < N-2M+1$. Then, for $\lambda > 3$ and $0 \le N_2 \le M$

$$\frac{\lambda/3+\frac{1}{2}(N-N_2)-1}{\frac{1}{2}(N-N_2)} = 1+\frac{\lambda/3-1}{\frac{1}{2}(N-N_2)} < 1+\frac{\lambda/3-1}{N_2+2}$$
$$< 1+\frac{\lambda/3-1}{N_2+1} = \frac{\lambda/3+N_2}{N_2+1}. \tag{B.22}$$

For $\lambda > 3$ and $N-2M \le N_2 \le M$,

$$\frac{\lambda/3+\frac{1}{2}(N-N_2)-1}{\frac{1}{2}(N-N_2)} = 1+\frac{\lambda/3-1}{\frac{1}{2}(N-N_2)} > 1+\frac{\lambda/3-1}{N_2+1}$$
$$> 1+\frac{\lambda/3-1}{N_2+2} = \frac{\lambda/3+N_2+1}{N_2+2}. \tag{B.23}$$

For $\lambda < 3$ and $0 \le N_2 \le M$,

$$\frac{\lambda/3+\frac{1}{2}(N-N_2)-1}{\frac{1}{2}(N-N_2)} = 1-\frac{1-\lambda/3}{\frac{1}{2}(N-N_2)} > 1-\frac{1-\lambda/3}{N_2+2}$$
$$= \frac{\lambda/3+N_2+1}{N_2+2} > 1-\frac{1-\lambda/3}{N_2+1} = \frac{\lambda/3+N_2}{N_2+1}. \tag{B.24}$$

For $\lambda < 3$ and $N-2M \le N_2 \le N$,

$$\frac{\lambda/3+\frac{1}{2}(N-N_2)-1}{\frac{1}{2}(N-N_2)} = 1-\frac{1-\lambda/3}{\frac{1}{2}(N-N_2)} < 1-\frac{1-\lambda/3}{N_2+1}$$
$$= \frac{\lambda/3+N_2+1}{N_2+1} < 1-\frac{1-\lambda/3}{N_2+2} = \frac{\lambda/3+N_2+1}{N_2+2}. \tag{B.25}$$

Inserting Ineqs. (B.22), (B.23), (B.24), and (B.25) into Eq. (B.21) immediately yields all cases of the lemma. Q.E.D.

From the lemma we shall now obtain the valuable Ineqs. (B.26)–(B.30) regarding summations over the values of N_2 in classes (a) and (c), if N and α are chosen as required in the lemma. For $\lambda > 3$

$$\sum_{N_2=0}^{M}{}' \Omega(N_2) \leq \left[\frac{\frac{\lambda}{3} \cdot \cdots \cdot \left(\frac{\lambda}{3} + M - 1 \right)}{M!} \right] \left[\frac{\frac{\lambda}{3} \cdot \cdots \cdot \left(\frac{\lambda}{3} + \frac{N-M}{2} - 1 \right)}{\left[\frac{1}{2}(N-M) \right]!} \right]^2$$

$$\times \sum_{N_2=0}^{M}{}' 1$$

$$\approx [\Gamma(\lambda/3)]^{-1} M^{\lambda/3-1} \left[\frac{1}{2}(N-M) \right]^{2(\lambda/3-1)} \left(\frac{1}{2}M + 1 \right)$$

$$= \phi(1 - M/N)^{2(\lambda/3-1)}(1 + 2/M)(N^{-\alpha})^{\lambda/3}$$

$$\approx \phi o(N^{-\alpha\lambda/3}), \qquad \text{(B.26)}$$

where Eqs. (B.12) and (B.17) have been used. Also for $\lambda > 3$,

$$\sum_{N_2=N-2M}^{N}{}' \Omega(N_2) \leq \frac{\frac{\lambda}{3} \cdot \cdots \cdot \left(\frac{\lambda}{3} + N - 2M - 1 \right)}{(N-2M)!}$$

$$\times \left[\frac{\frac{\lambda}{3} \cdot \cdots \cdot \left(\frac{\lambda}{3} + M - 1 \right)}{M!} \right]^2 \times \sum_{N_2=N-2M}^{N}{}' 1$$

$$\approx \left[\Gamma\left(\frac{\lambda}{3} \right) \right]^{-3} (N - 2M)^{\lambda/3-1} M^{2(\lambda/3-1)}(M+1)$$

$$= \phi(1 - 2M/N)^{\lambda/3-1}(1 + M^{-1})(N^{-\alpha})^{2\lambda/3-1}$$

$$\approx \phi o[N^{-\alpha(2\lambda/3-1)}]. \qquad \text{(B.27)}$$

For $0 < \lambda < 3$ we have (using the remark about $N_j = 0$ after Eq. (4.4a)),

$$\sum_{N_2=0}^{M}{}' \Omega(N_2) \le 1\left[\frac{\frac{\lambda}{3}\cdot\cdots\cdot\left(\frac{\lambda}{3}+\frac{N}{2}-1\right)}{\left(\frac{1}{2}N\right)!}\right]^2 \sum_{N_2=0}^{M}{}' 1$$
$$\approx\left[\Gamma\left(\frac{\lambda}{3}\right)\right]^{-2}(2N^{-1})^{2(1-\lambda/3)}\left(\frac{1}{2}M+1\right)$$
$$= \phi\Gamma\left(\frac{\lambda}{3}\right)(1+2M^{-1})N^{-[\alpha-(1-\lambda/3)]}$$
$$= \phi o\left(N^{-[\alpha-(1-\lambda/3)]}\right). \tag{B.28}$$

Also for $0 < \lambda < 3$,

$$\sum_{N_2=N-2M}^{N}{}' \Omega(N_2) \le \frac{\frac{\lambda}{3}\cdot\cdots\cdot\left(\frac{\lambda}{3}+N-1\right)}{N!}\cdot 1\cdot \sum_{N_2=N-2M}^{N}{}' 1$$
$$\approx[\Gamma(\lambda/3)]^{-1}(N^{-1})^{1-\lambda/3}(M+1)$$
$$= \phi\cdot 2^{2\lambda/3-1}[\Gamma(\lambda/3)]^2(1+M^{-1})N^{-\alpha+2(1-\lambda/3)}$$
$$\approx \phi o\left(N^{-\alpha+2(1-\lambda/3)}\right). \tag{B.29}$$

Finally, for $0 < \lambda < 3$, it is convenient to have the following inequality, which asserts a lower bound rather than an upper bound as in the foregoing inequalities:

$$\sum_{N_2=N-2M}^{N}{}' \Omega(N_2) > \Omega(N) \approx [\Gamma(\lambda/3)]^{-1}N^{\lambda/3-1}$$
$$= \phi 2^{2\lambda/3-1}[\Gamma(\lambda/3)]^2 N^{1-2\lambda/3}. \tag{B.30}$$

Ineq. (B.30) will be of interest only when $\lambda < \frac{3}{2}$, since only in this subcase does $o(N^{1-2\lambda/3})$ go to ∞.

After all these preparations, we can now prove the theorem of Section IV for the remaining cases.

For $\lambda > 3$ we take $\alpha > 0$ and use Eq. (B.20) to write

$$I_2 \approx \frac{1}{3}\frac{\lambda}{\lambda-1}D, \tag{B.31}$$

and it is easily checked that $0 < D < \infty$. Then from Eqs. (B.7), (B.8), (B.19), and (B.31), and Ineqs. (B.26) and (B.27) we find

$$\underline{c}_\lambda(\underline{h}_2|\underline{d}_2) \approx \frac{I_2 + \phi\cdot 0 + \phi\cdot 0}{\phi D + \phi\cdot 0 + \phi\cdot 0} \approx \frac{1}{3}\frac{\lambda}{\lambda - 1}. \tag{B.32}$$

For $\frac{3}{2} < \lambda < 3$, we use Eq. (B.20) to write

$$\begin{aligned} I_2 &= \frac{1}{3}\frac{\lambda}{\lambda-1}D - \frac{1}{\lambda-1}N^{-\alpha(2\lambda/3-1)}\left[2^{2\lambda/3-1}(1-2N^{-\alpha})^{\lambda/3}\right. \\ &\quad \left. - N^{-\alpha(1-\lambda/3)}(1-N^{-\alpha})^{2\lambda/3-1}\right] \\ &= \frac{1}{3}\frac{\lambda}{\lambda-1}D\left[1 - o\left(N^{-\alpha(2\lambda/3-1)}\right)\right], \end{aligned}$$

so that $\alpha > 0$ implies

$$I_2 \approx \frac{1}{3}\frac{\lambda}{\lambda-1}D. \tag{B.33}$$

It is easily checked that $D > 0$, and one finds that $D < \infty$ as follows:

$$\begin{aligned} D &= (D - I_2) + I_2 < \int_{N^{-\alpha}}^{1-2N^{-\alpha}} \frac{dy}{y^{1-\lambda/3}} + \int_{N^{-\alpha}}^{1-2N^{-\alpha}} \frac{dy}{(1-y)^{2(1-\lambda/3)}} \\ &= \frac{3}{\lambda}\left[(1-2N^{-\alpha})^{\lambda/3} - (N^{-\alpha})^{\lambda/3}\right] \\ &\quad + \frac{1}{2\lambda/3 - 1}\left[(1-N^{-\alpha})^{(2\lambda/3-1)} - (2N^{-\alpha})^{1-2\lambda/3}\right] \\ &< \infty. \end{aligned}$$

Then from Eqs. (B.7), (B.13). (B.18), (B.19), and (B.33), and Ineqs. (B.28) and (B.29) we have

$$\underline{c}_\lambda(\underline{h}_2|\underline{d}_2) \approx \frac{I_2 + \phi\cdot 0 + \phi\cdot 0}{\phi D + \phi\cdot 0 + \phi\cdot 0} \tag{B.34},$$

provided that in Ineqs. (B.28) and (B.29) we choose $\alpha > 2(1 - \lambda/3) > (1 - \lambda/3)$. But the bounds $\frac{3}{2} < \lambda < 3$ in the case under consideration guarantee that an α in $(0, 1)$ can be chosen to satisfy this condition.

For the case of $0 < \lambda < \frac{3}{2}$ we write

$$\begin{aligned} D - I_2 &< \int_{N^{-\alpha}}^{1-2N^{-\alpha}} \frac{dy}{y^{1-\lambda/3}} = \frac{3}{\lambda}\left[(1-2N^{-\alpha})^{\lambda/3} - (N^{-\alpha})^{\lambda/3}\right] \\ &\approx \frac{3}{\lambda}. \end{aligned} \tag{B.35}$$

For y in the range of integration, $(1-y)^{2(1-\lambda/3)} < 1-y$ and $y^{\lambda/3} > y$, and therefore

$$I_2 > \int_{N^{-\alpha}}^{1-2N^{-\alpha}} \frac{y}{1-y}\,dy = \left[-\ln(1-y) - y\right]^{1-2N^{-\alpha}}$$
$$= \ln N^{\alpha} + \ln\{\tfrac{1}{2}(1-N^{-\alpha}) - 1 + 3N^{-\alpha}, \tag{B.36}$$

which diverges as $N \to \infty$. Therefore, by Ineq. (B.35) D also diverges, so that

$$D \approx I_2. \tag{B.37}$$

Furthermore,

$$I_2 < \int_{N^{-\alpha}}^{1-2N^{-\alpha}} \frac{dy}{(1-y)^{2(1-\lambda/3)}}$$
$$= \frac{1}{1-2\lambda/3}\left\{\frac{N^{\alpha(1-2\lambda/3)}}{2^{1-2\lambda/3}} - (1-2N^{-\alpha})^{2\lambda/3-1}\right\}. \tag{B.38}$$

Then

$$\sum_{N_2=0}^{N}{}' \Omega(N_2) < \sum_{N_2=N-2M}^{N}{}' \Omega(N_2) + \sum_{N_2=M}^{N-2M} \Omega(N_2) + \phi\cdot o[N^{-\alpha+(1-\lambda/3)}]$$
$$\approx \sum_{N_2=N-2M}^{N}{}' \Omega(N_2) + \phi D + \phi\cdot 0$$
$$= \sum_{N_2=N-2M}^{N}{}' \Omega(N_2)\left\{1 + \frac{\phi D}{\sum_{N_2=N-2M}^{N}{}' \Omega(N_2)}\right\}$$
$$\lesssim \sum_{N_2=N-2M}^{N}{}' \Omega(N_2)\left\{1 + \frac{\phi N^{\alpha(1-2\lambda/3)}/\left[(1-2\lambda/3)2^{1-2\lambda/3}\right]}{\phi 2^{2\lambda/3-1}\left[\Gamma(\lambda/2)\right]^2 N^{1-2\lambda/3}}\right\}, \tag{B.39}$$

where we have used Ineq. (B.30) and have assumed that $\alpha > 1 - \lambda/3$. Also

$$\sum_{N_2=0}^{N}{}' \frac{N_2}{N}\Omega(N_2) > \sum_{N_2=N-2M}^{N}{}' \frac{N_2}{N}\Omega(N_2)$$
$$> (1-2M/N) \sum_{N_2=N-2M}^{N}{}' \Omega(N_2)$$
$$= (1-2/N^2) \sum_{N_2=N-2M}{}' \Omega(N_2). \tag{B.40}$$

Inequalities (B.39) and (B.40) imply

$$\underline{c}_\lambda(\underline{h}_2|\underline{d}_2) = \frac{\sum'^{N}_{N_2=0} (N_2/N)\Omega(N_2)}{\sum'^{N}_{N_2=0} \Omega(N_2)} \gtrsim 1, \tag{B.41}$$

provided that $\alpha < 1$. But since probability theory places an upper bound of 1 on $\underline{c}_\lambda(\underline{h}_2|\underline{d})$, we have

$$\underline{c}_\lambda(\underline{h}_2|\underline{d}_2) \approx 1. \tag{B.42}$$

For $\lambda = \frac{3}{2}$ we use the fact that $\underline{c}_\lambda(\underline{h}_2|\underline{d}_2)$ is continuous in λ by Eq. (4.15), so that from Eqs. (B.34) and (B.42) we have

$$\underline{c}_{3/2}(\underline{h}_2|\underline{d}_2) = \lim_{\lambda\to 3/2^-} \underline{c}(\underline{h}_2|\underline{d}_2) = 1, \tag{B4.43a}$$

and

$$\underline{c}_{3/2}(\underline{h}_2|\underline{d}_2) = \lim_{\lambda\to 3/2^+} \underline{c}_\lambda(\underline{h}_2|\underline{d}_2) \approx \lim_{\lambda\to 3/2^+} \frac{1}{3}\frac{\lambda}{\lambda - 1} = 1. \tag{B4.43b}$$

It is also possible to show that $\underline{c}_{3/2}(\underline{h}_2|\underline{d}_2) = 1$ by analyzing summations in the manner of the other cases.

We note that Eq. (B.4b), for the case of $\lambda = 0$, agrees with the limit of Eq. (B.42) as $\lambda \to 0$; and that Eq. (B.5b), for the case of $\lambda = 3$, agrees with the limits of Eqs. (B.32) and (B.34). Finally, the case of $\lambda = \infty$, which we said could be treated by the Darwin–Fowler method, also can be handled by using Eq. (B.32):

$$\underline{c}_\infty(\underline{h}_2|\underline{d}_2) = \lim_{\lambda\to\infty} \underline{c}_\lambda(\underline{h}_2|\underline{d}_2) \approx \lim_{\lambda\to\infty} \frac{1}{3}\frac{\lambda}{\lambda - 1} = \frac{1}{3}.$$

Thus, the theorem stated in Section IV has been proved in all cases.

References

1. L. Brillouin, "Science and Information Theory," Academic Press, New York, 1956.
2. R. Carnap, "Logical Foundations of Probability," Univ. of Chicago Press, 1950.
3. R. Carnap, "The Continuum of Inductive Methods," Univ. of Chicago Press, 1952.
4. R. Carnap, A basic system of inductive logic, part I, *in* "Studies in Inductive Logic and Probability," Vol. I. (R. Carnap and R. Jeffrey, Eds.), Univ. of California Press, Berkeley/Los Angeles, 1965.

5. J. F. Cyranski, Measurement, theory, and information, *Inform. and Control* **33** (1979), 275–304.
6. W. Feller, "An Introduction to Probability Theory and its Applications" Vol. I, 2nd ed., Wiley, New York, 1957.
7. K. Friedman, Replies to Tribus and Motroni and to Gage and Hestenes, *J. Statist. Phys.* **9** (1973), 265–269.
8. G. Frege, "Translations from the Philosophical Writings of Gottlob Frege" (P. Geach and M. Black, Eds.), Blackwell, Oxford, 1952.
9. K. Friedman and A. Shimony, Jaynes's maximum entropy prescription and probability theory, *J. Statist. Phys.* **3** (1971), 381–384.
10. D. W. Gage and D. Hestenes, Comments on the paper "Jaynes's maximum entropy prescription and probability theory", *J. Statist. Phys.* **7** (1973), 89–90.
11. A. Hobson, The interpretation of inductive probabilities, *J. Statist. Phys.* **6** (1972) 189–193.
12. E. T. Jaynes, Information theory and statistical mechanics, I, *Phys. Rev.* **106** (1957), 620–630; II, *Phys. Rev.* **108** (1957), 171–190.
13. E. T. Jaynes, Information theory and statistical mechanics, *in* "Statistical Physics" (K. W. Ford, Ed.), Brandeis University Summer Institute Lectures in Theoretical Physics, 1962, Vol. 3, Benjamin, New York, 1963.
14. E. T. Jaynes, Foundations of probability theory and statistical mechanics, *in* "Delaware Seminar in the Foundations of Physics" (M. Bunge, Ed.), Springer, New York, 1967.
15. E. T. Jaynes, Prior probabilities, *IEEE Trans. Syst. Sci. Cybernet.* **SSC-4** (1968), 227–241.
16. E. T. Jaynes, Confidence intervals vs Bayesian intervals, *in* "Foundations of Probability Theory, Statistical Inference, and Statistical Theories of Science" (W. Harper and C. Hooker, Eds.), Vol. II, Reidel, Dordrecht, 1976.
17. E. T. Jaynes, Where do we stand on maximum entropy? *in* "The Maximum Entropy Formalism" (R. Levine and M. Tribus, Eds.), MIT Press, Cambridge, Mass., 1979.
18. H. Jeffreys, "Theory of Probability," 3rd ed., Oxford Univ. Press, London/New York, 1961.
19. J. M. Keynes, "A Treatise on Probability," Macmillan, London/New York, 1921.
20. P. Laplace, "A Philosophical Essay on Probabilities," Dover, New York, 1952.
21. B. Mandelbrot, The role of sufficiency and estimation in thermodynamics, *Ann. Math. Statist.* **33** (1962), 1021–1038.
22. O. Penrose, "Foundations of Statistical Mechanics," Pergamon, Oxford, 1969.
23. E. Schrödinger, "Statistical Thermodynamics," Cambridge Univ. Press, Cambridge, 1967.
24. T. Seidenfeld, Why I am not an objective Bayesian; Some reflections prompted by Rosenkrantz, *Theory and Decision* **11** (1979), 413–440.
25. C. Shannon, A mathematical theory of communication, *Bell Syst. Tech. J.* **27** (1948), 379–523.
26. A. Shimony, Scientific inference, *in* "The Nature and Function of Scientific Theories" (R. Colodny, Ed.), Pittsburgh Univ. Press, 1970.
27. A. Shimony, Comment on the interpretation of inductive probabilities, *J. Statist. Phys.* **9** (1973), 187–191.
28. Y. Sinai, "Introduction to Ergodic Theory," Princeton Univ. Press, Princeton, N.J., 1976.
29. L. Szilard, Über die Entropieverminderung in einem thermodynamischen System, *Z. Phys.* **53** (1929), 840–856.
30. L. Tisza and P. Quay, The statistical thermodynamics of equilibrium, *Rev. Mod. Phys.* **25** (1963), 48–90. Reprinted in L. Tisza, "Generalized Thermodynamics," pp. 245–287, MIT Press, Cambridge, Mass., 1966.
31. A. Wehrl, General properties of entropy, *Rev. Mod. Phys.* **50** (1978), 221–260.