International Journal of Theoretical Physics, Vol. 30, No. 6, 1991

Probability Distributions of Thermodynamic Intensive Variables

B. H. Lavenda[1] **and J. Dunning-Davies**[2]

Received August 22, 1990

Probability distributions of thermodynamic intensive variables are derived from the Gaussian law of error for extensive variables. Bayesian distributions, in the sense of degree-of-belief, are derived by interchanging the endpoints of the interval and transforming the probability distribution into one dependent upon a parameter which is to be estimated from the sample data. Legendre transforms of the probability distribution for extensive variables lead to dual distributions, in the frequency sense, for the conjugate thermodynamic intensive variables.

1. DUALITY IN THERMODYNAMICS AND STATISTICS

Both thermodynamics and mathematical statistics classify variables into two categories according to whether they are *extensive* or *intensive*, in thermodynamics, and *observable* or *estimable*, in statistics. The Bayesian method makes the further distinction of whether the variables are to be interpreted in the *limit-of-frequency* sense or in the sense of *degree-of-belief.* Moreover, there exists an inherent asymmetry in the two categories where prominence is given to the primary set of variables which are extensive and observable. The primary set defines the state of the system in thermodynamics, while in statistics it comprises the sample data. The secondary set links the system to the outside world; it consists of the intensive variables in thermodynamics and those parameters which define the *state of nature* in statistics. These quantities are *derived*; in thermodynamics, they are obtained by partially differentiating the fundamental relation that describes the thermodynamic properties of the system, while in statistics they are estimated in terms of the observations. The estimators of the intensive

---

[1] Università degli Studi, Camerino 62032 (MC), Italy.
[2] University of Hull, Hull HU6 7RX, Great Britain.

0020-7748/91/0600-0907$06.50/0

variables, which are random quantities since they are functions of the sample data, can be interpreted both in the sense of limit-of-frequency and degree-of-belief. Therefore, the secondary set must also possess well-defined probability distributions and, in this paper, we will derive these distributions from the Gaussian law of error for their conjugate extensive variables.

A thermodynamic state of the system is specified by a set of extensive thermodynamic variables, which, according to statistical mechanics, coincides with the average energy $\bar{U}$, volume $\bar{V}$, and mole number $\bar{N}$. The thermodynamic properties of the system are deduced from the fundamental relation which expresses the entropy S in terms of the independent extensive variables (Callen, 1985). The secondary set of thermodynamic variables consists of the temperature $\hat{T}$, the pressure, and the chemical potential. These intensive quantities are defined by the partial derivatives of the fundamental relation with respect to their conjugate extensive variables. By a rotation of the Gibbs space, the fundamental relation can be inverted so that the energy U is made a function of the independent extensive variables, which now consist of $\bar{S}$, $\bar{V}$, and $\bar{N}$. This is known as the energy representation and is limited to the analysis of reversible processes.

Observable variables are measurable and since a measurement can never be performed with unlimited precision, a distribution in the possible values that the extensive variables can assume will be set up about their most probable values. Since these variables are *additive*, the most probable value will be assumed to coincide with the sample mean, which will be assumed to coincide with the mean of the distribution. This, in general, will be rigorously so when the sample size is allowed to increase without bound. However, due to the particular properties of the distribution, it will be true for any value of the sample size.

Estimable variables are parameters upon which the probability distributions of the extensive variables depend; they define the state of nature of the system by relating the probability distributions to the properties of the physical system under investigation. If the state of nature was known with certainty, an observation would not be informative and there would be no need to deal with probability distributions. The estimators of the parameters defining the state of nature are random variables since they are functions of the sample data.

Statistical thermodynamics deals only with estimators that are *sufficient* statistics (Mandelbrot, 1962). In a physical sense, sufficient statistics summarizes all the "information" there is about the parameter that is contained in the sample. In a mathematical sense, the conditional probability of any other statistic, given the sufficient statistics, is independent of the parameter to be estimated (Lindley, 1970). Sufficient statistics, whatever the sample size, exists for the exponential family of distributions. These distributions introduce a certain duality, which will be explored in this paper, in the sense of a Legendre transform between observable and estimable variables.

In the entropy representation, the most fundamental pair of conjugate variables is the energy $\bar{U}$ and the inverse temperature $\hat{\beta}$, which will be measured in energy units where Boltzmann's constant is unity. The distribution of the energy depends upon a parameter β whose "true" value is the inverse temperature $\hat{\beta}$. The problem of point estimation consists in estimating β from observations U_i of the energy. These values can be taken from a single system or different subsystems which are subsequently allowed to interact thermally so as to form a composite system. Although sampling from a large population is the usual procedure in statistics, the structure of statistical thermodynamics favors the latter interpretation (Mandlebrot, 1962; Lavenda and Scherer, 1988). If the sample size was to increase without bound, certainty would be obtained in the limit. However, at some stage, the additivity of the extensive variables would break down together with the property of sufficiency (Mandelbrot, 1956).

The natural choice for the sufficient statistic is the sample mean,

$$\bar{U} = \frac{1}{m} \sum_{i=1}^{m} U_i \tag{1}$$

owing to the additivity property of the energy. The assertion that (1) is a sufficient statistic for the estimation of the inverse temperature simply means that the conditional probability of observing the value U_i of any subsystem, given the total energy $m\bar{U}$ of the composite system, is independent of the common value of the temperature when the subsystems are brought into thermal contact and left alone for a sufficiently long period of time so as to secure thermal equilibrium (Mandelbrot, 1956).

Any estimate of the temperature is itself a random variable, since it is a function of the sample energies. A sufficient statistic requires that the converse is not true. This is in complete harmony with the second law: The concavity of the entropy with respect to the energy ensures that there is a unique solution, $\hat{\beta} = \hat{\beta}(\bar{U})$, to

$$S'(\bar{U}) = \hat{\beta} \tag{2}$$

where the prime stands for differentiation, while the convexity of the entropy with respect to the inverse temperature is independent of the energy.

The statistical correlations between any estimator of the inverse temperature and the sample values of the energy, of which it is a function, is summarized by the thermodynamic uncertainty relations (Lavenda, 1987; Lavenda and Scherer, 1988; Mandelbrot, 1956; Tikochinsky and Levine, 1984). The thermodynamic uncertainty relations in the case of the energy and inverse temperature can be discussed in terms of the size of the thermostat with which the system is in contact (Mandelbrot, 1956). As the thermostat shrinks in size, measurements of the energy become increasingly more precise at the expense of those of the temperature. In the limit where the thermostat vanishes, it is no longer possible to define a temperature. As the thermostat increases in size, measurements of the energy become increasingly less precise, while those of the temperature become more precise.

There are two different types of probability distributions for the inverse temperature; both can be derived from the fundamental error law for the energy. In one case, we want a rational degree-of-belief of the inverse temperature given the sample mean energy. The randomness of β is in the sense of degree-of-belief that certain values are more probable than others. In Section 3, the likelihood function is obtained simply by interchanging the endpoints of the interval in the error law for the energy. This function is a measure of the "likelihood" of β and it is only natural to choose that value whose "likelihood" is a maximum. Since we are dealing with the Legendre transform of the entropy, the stationary condition is entirely equivalent to the second law which defines the inverse temperature. In Section 5, we inquire about the probability of making an error by "guessing" a certain value of β whose true value is $\hat{\beta}$. This probability distribution is defined in the limit-of-frequency and is derived from the Legendre transform of the entropy in the fundamental error law for the energy. In exactly the same way that the error law for an extensive variable is expressed in terms of the concavity criterion of the entropy, the error law governing its conjugate intensive variable is expressed in terms of the concavity condition of the Legendre transform of the entropy.

2. LAW OF ERROR FOR THE ENERGY

For simplicity, we will consider deviations in the energy, holding constant the volume and mole number. The thermodynamic stability properties of the system are couched in the *strict* concavity properties of the entropy (Galgani and Scotti, 1970; Lavenda and Dunning-Davies, 1990)

$$S''(\bar{U}) < 0 \tag{3}$$

An alternative definition of concavity employs the truncated Taylor series expansion (Hardy *et al.*, 1952)

$$S(U_i) = S(\bar{U}) + S'(\bar{U})(U_i - \bar{U}) + \tfrac{1}{2} S''(U)(U_i - \bar{U})^2$$

where U lies between U_i and $\bar{U}$. On the strength of inequality (3) we have

$$S(U_i) - S(\bar{U}) - S'(\bar{U})(U_i - \bar{U}) < 0 \tag{4}$$

as an alternative criterion of concavity. This criterion of concavity is related to the existence of a probability distribution for the energy (Lavenda and Dunning-Davies, 1990).

Denote by $f(U_i|\bar{U})$ the probability that an observation of the energy of any of the m subsystems will result in an energy U_i whose most probable value is the sample mean (1), which, by assumption, is equal to the mean of the distribution. Since the energies of the m subsystems are independent, the joint or "sample" probability is

$$f(U_1, \ldots, U_m|\bar{U}) = \prod_{i=1}^{m} f(U_i|\bar{U}) \tag{5}$$

If the sample mean (1) is the most probable value of the energy, then the joint probability distribution (5) will satisfy the stationary condition

$$\sum_{i=1}^{m} f_i'/f_i = 0 \tag{6}$$

where we have abbreviated the notation, $f_i = f(U_i|\bar{U})$, and the prime stands for differentiation with respect to $\bar{U}$.

If we fix the sample mean, which is another way of saying that the composite system is isolated, having a total energy $m\bar{U}$, it will act as a constraint on the extremum principle. The constrained variational principle can be converted into a free one by the usual method of Lagrange multipliers. Instead of (6), the stationary condition of the free variational principle is

$$\sum_{i=1}^{m} f_i'/f_i + S''(\bar{U})(U_i - \bar{U}) = 0 \tag{7}$$

where the second derivative of the entropy plays the role of a Lagrange multiplier. Since the energies of the subsystems that form the composite system are now all independent, condition (7) must be satisfied for each of the m subsystems. Integrating by parts then gives

$$\log f(U_i|\bar{U}) = \psi(U_i) - S(\bar{U}) - (U_i - \bar{U})S'(\bar{U}) + \log A$$

where A is a normalization constant. In order that $f(U_i|\bar{U})$ lie between 0 and 1, the right side must be negative for all values of U_i and $\bar{U}$ that are physically possible. This is guaranteed by condition (4) that the entropy is a concave function. Alternatively, we may argue that since $U_i = \bar{U}$ maximizes the probability distribution, the integration function $\psi(U_i)$ must be the same function of U_i that the entropy is of $\bar{U}$. Greene and Callen (1951) have raised this to a principle, for if it were otherwise, they argue that there would be a separate thermodynamics for the microcanonical and canonical ensembles. We have shown that this principle rests on the validity of Stirling's approximation in the case where the independent extensive variable is the average mole number (Lavenda and Dunning-Davies, in press).

Consequently, the relation between the concavity of the entropy and the law of error leading to the average value as most probable value of the energy is

$$f(U_i|\bar{U}) = A\exp\{S(U_i) - S(\bar{U}) - S'(\bar{U})(U_i - \bar{U})\} \tag{8}$$

where we have identified $\psi(U_i)$ with the "stochastic" entropy $S(U_i)$ since it is a function of the observed energy. The link between the error law for the energy (8) and the state of nature of the system is given by the second law, (2), which defines the inverse temperature $\hat{\beta}$. Introducing (2) into the law of error (8) gives

$$f(U_i|\hat{\beta}) = A\exp\{-U_i\hat{\beta} + S(U_i) + L(\hat{\beta})\} \tag{9}$$

where

$$L(\hat{\beta}) = \hat{\beta}\bar{U} - S(\bar{U}) \tag{10}$$

which is the Legendre transform with respect to the energy.

The error law, written in the form (9), is clearly seen to satisfy Neyman's factorization criterion in order for distribution to admit a sufficient statistic. As a problem of point estimation, we would replace the true value of β by another estimate and use the sample mean (1), which is now seen to be a sufficient statistic, to estimate the true value of β. The usual way is to use Fisher's method of maximum likelihood, which constructs the likelihood function from the joint distribution (5). We shall now give an alternative method which simply interchanges the endpoints of the interval in the law of error for the extensive variable to produce a likelihood function which is independent of the sample size.

3. ERROR LAW AS A LIKELIHOOD FUNCTION

The second law (2) defines the inverse temperature, which provides knowledge of the "state of nature" of the system, since it relates the probability distribution to the physical state of the system. However, in the same way that it is essential to determine the probability of deviations in the extensive variables from their most probable value, so it is necessary to consider the statistical character of the knowledge of the state of nature. For if the state of nature were known exactly, then no observation of the energy would provide a better estimate of it than what is already known. In other words, no experiment would be informative.

The second law (2) converts the error law for the energy (8) into (9), for which the best estimator of β, namely $\hat{\beta}$, has the highest probability. However, if we want to find a measure of rational belief in a value of β, we would start with a probability distribution

$$f(U_i|\beta) = A\exp\{S(U_i) + L(\beta) - \beta U_i\} \tag{11}$$

where $L(\beta)$ is related to the moment generating function by

$$\exp[-L(\beta)] = \sum_{i=1}^{m} A \exp\{-\beta U_i + S(U_i)\}$$

The distribution (11) depends upon a parameter β which is to be estimated from a sample of m observations on the energy. To this end, we form the joint or "sample" distribution (5),

$$f(U_1, \ldots, U_m|\beta) = A^m \exp\{m[-\beta\bar{U} + S(\bar{U}) + L(\beta)]\} = f(\bar{U}|\beta) \quad (12)$$

Owing to the extensitivity of the entropy and its Legendre transform (12) depends on the observations of the energy only through the sample mean (1). In other words, the entropy obtained by combining m subsystems, all having the same value of the parameter β, is the sum of the entropies of the individual subsystems. An increase in entropy results from the thermal interaction of subsystems that do not have a common value of the parameter β.

The "likelihood" function (12), for a fixed sample mean, provides a subjective probability measure for different values of β. The maximum likelihood method inverts the functional dependence in (12), where β is now considered as the variable and $\bar{U}$ the parameter. Only in the case (i.e., the Bayes case) where β is equipped with a prior probability density can a probabilistic interpretation be given to (12); otherwise, we must be content with comparing the likelihoods of different values of β.

The maximum likelihood estimate is obtained as the solution of

$$\frac{\partial}{\partial\beta}\log f(\bar{U}|\beta) = m[L'(\beta) - \bar{U}] = 0 \quad (13)$$

since

$$\frac{\partial^2}{\partial\beta^2}\log f(\bar{U}|\hat{\beta}) = mL''(\hat{\beta}) < 0$$

This is a consequence of the fact that the strict concavity property of the entropy is transferred to its Legendre dual,

$$L''(\hat{\beta}) = 1/S''(\bar{U}) < 0 \quad (14)$$

By the implicit function theorem, the likelihood equation (13) may be solved for the maximum likelihood estimate $\hat{\beta} = \hat{\beta}(\bar{U})$. Since S and L are Legendre transforms of one another, the maximum likelihood estimate $\hat{\beta}$ will coincide with the thermodynamic definition of the inverse temperature (2).

The joint distribution (12) is an m-fold product of distributions

$$f^{\#}(\bar{U}|\beta) = A \exp\{-\beta\bar{U} + S(\bar{U}) + L(\beta)\} \quad (15)$$

It is quite remarkable that this distribution results from interchanging the endpoints of the interval in the error law (8),

$$f^{\#}(\bar{U}|U_i) = A \exp\{S(\bar{U}) - S(U_i) + S'(U_i)(U_i - \bar{U})\} \tag{16}$$

Owing to the concavity condition,

$$S(\bar{U}) - S(U_i) + S'(U_i)(U_i - \bar{U}) < 0 \tag{17}$$

(16) is, again, a proper probability distribution. In fact, the sum of the two conditions of concavity, (4) and (17), is

$$(U_i - \bar{U})[S'(U_i) - S'(\bar{U})] < 0$$

showing that the slope of the entropy is monotonically *decreasing.* In thermodynamics, this inequality guarantees that heat will not flow spontaneously from a hot to a cold body or, equivalently, that the heat capacity is always positive definite.

In the Bayes sense, the error laws (8) and (16) can be considered to be inverses of each other for which "cause" and "effect" have been interchanged. We are reasoning from a sample of m observations made on the energies of m subsystems, each being characterized by a parameter β_i. However, nothing is known about these parameters, so that, in accordance with the principle of indifference, we set

$$\beta = S'(U_i) \tag{18}$$

for each of the subsystems. Certainly, when the subsystems are allowed to interact thermally and left alone for a sufficiently long period of time we should expect that all the β's will be the same. Inserting (18) into the error law (16) results in (15), where the Legendre transform

$$L(\beta) = \beta U_i - S(U_i) \tag{19}$$

is the same for each subsystem.

The maximum likelihood estimate maximizes the posterior probability, which, according to Bayes' theorem, is

$$f^{\#}(\beta|\bar{U}) = f^{\#}(\bar{U}|\beta) f^{\#}(\beta) / f^{\#}(\bar{U}) \tag{20}$$

where the prior densities are $f^{\#}(\beta)$ and

$$f^{\#}(\bar{U}) = \int_0^{\infty} f^{\#}(\bar{U}|\beta) f^{\#}(\beta)\, d\beta \tag{21}$$

If the prior probability density $f^{\#}(\beta)$ is uniform, the posterior probability (20) is proportional to the conditional probability, or likelihood function, (15) and $\hat{\beta} = \hat{\beta}(\bar{U})$ will be a solution of both the second law (2) and the likelihood equation (13). For $\beta = \hat{\beta}$, the exponent in (15) attains its supremum value of zero on account of the Legendre transform (10). In the case where the prior density is not uniform, there is a distinction between the probability densities (12) and (15). Since the logarithm of the former increases linearly with the sample size, it will ultimately dwarf the logarithm of the prior density as the sample size increases without limit. The same is not true of the conditional probability (15) and another form of asymptotic equivalence must be sought.

In statistical thermodynamics, other asymptotic limits are of interest, such as when the number of degrees of freedom of a system is very great (Gibbs, 1902). The logarithm of the conditional probability (15) will then dwarf the prior density, since the latter is independent of the number of degrees of freedom, as the following example will show.

Consider an ideal gas whose fundamental relation is

$$S(\bar{U}, \bar{V}, \bar{N}) = \bar{N} \log(\bar{U}^n \bar{V}^r \bar{N}^s)$$

Similar relations apply to each of the m subsystems. The exponents must add to zero in order that the entropy be extensive. It will prove more convenient, however, to work with molar quantities where the fundamental relation for the composite and ith subsystem are, respectively,

$$s(\bar{u}, \bar{v}) = \log(\bar{u}^n \bar{v}^r) \tag{22}$$

and

$$s_i = \log(u_i^n v_i^r) \tag{23}$$

Due to our lack of knowledge, each subsystem will be characterized by the same value of β, where

$$\left(\frac{\partial s_i}{\partial u_i}\right)_{v_i} = \frac{n}{u_i} = \beta \tag{24}$$

For, if each system did not have the same value of β, the entropy would not be additive. The Legendre transform of the molar entropy s_i, with respect to the molar energy u_i, is

$$l(\beta) = n \log \beta - r \log v_i + n - n \log n \tag{25}$$

In terms of molar quantities, the conditional probability density (15) is

$$f^{\#}(\bar{u}|\beta) = A(\beta\bar{u})^n \exp\{-\beta\bar{u} + n - n \log n\} \tag{26}$$

Choosing the prior density of the scale parameter as

$$f^{\#}(\beta) = 1/\beta \tag{27}$$

we have that the prior density of (21) will be unity if we set $A = n!/\Gamma(n)$. Then, according to Bayes' theorem (20), the posterior probability density for the scale parameter is

$$f^{\#}(\beta|\bar{u}) = \bar{u}\frac{(\beta\bar{u})^{n-1}}{\Gamma(n)}e^{-\beta\bar{u}} \tag{28}$$

which is again a gamma density. As a function of $\beta, f^{\#}(\bar{u}|\beta)$ is the likelihood function $\mathscr{L}$ whose logarithm is

$$\ln \mathscr{L}(\beta|\bar{u}) = -\beta\bar{u} + n \ln(\beta\bar{u})$$

The likelihood function has a unique maximum at $\beta = \hat{\beta}$, where $\hat{\beta}$ is the solution to the likelihood equation

$$\frac{\partial}{\partial\beta}\ln\mathscr{L} = -\bar{u} + \frac{n}{\beta} = 0$$

Apart from supplying the correct normalization so that the posterior density (28) will be a proper probability density (27) is Jeffreys' (1961) second rule that applies to scale parameters which can assume values from 0 to ∞. The integral of (27) over all values of β is infinite and Jeffreys uses it as his definition of certainty. Hence, the improper probability density is used to represent ignorance about β. The maximum likelihood estimate will coincide with the one that maximizes the *a posteriori* density (20) when $n \gg 1$, where n is half the number of degrees of freedom. In this limit, the logarithm of (26) dwarfs the logarithm of (27).

4. THERMODYNAMIC UNCERTAINTY RELATIONS

We have seen that fluctuations in the energy cause fluctuations in its conjugate intensive variable. In order to emphasize the symmetry between the two, we consider the energy to be a continuous random variable and compare the variances of the probability densities (9) and (15).

The logarithm of the moment generating function of the probability density (9),

$$L(\hat{\beta}) = -\log\int_0^\infty \exp\{-\hat{\beta}u + S(u)\}\,du - \log A \tag{29}$$

is a completely monotone function, since its derivatives alternate in sign. The first moment is just the stationarity condition of the likelihood function (13), while the second central moment is

$$L''(\hat{\beta}) = -\overline{(u-\bar{u})^2} \doteq -\overline{(\Delta u)^2} < 0 \tag{30}$$

where the bar denotes averaging. Expression (30) clearly brings out the relation between the concavity of the Legendre transform of the entropy and the stability properties of thermodynamic systems; since $\overline{(\Delta u)^2} = -\bar{u}'(\hat{\beta})$, the heat capacity per particle $C = -\hat{\beta}^2 \bar{u}'(\hat{\beta})$ is necessarily positive.

The logarithm of the moment generating function of the conditional probability density (15),

$$S(\bar{u}) = -\log \int_0^\infty \exp\{-\beta \bar{u} + L(\beta)\}\, d\beta \tag{31}$$

is also a completely monotone function. A comparison of (29) and (31) shows that the roles of the Legendre dual functions have been interchanged, although their physical meanings must be kept quite distinct (Lavenda, 1988). The "structure" function $\Omega(u) = \exp[S(u)]$ represents the density of states in the energy interval between u and $u + du$. Although we may formally consider $\Omega(\beta) = \exp[L(\beta)]$ as a "density of states" for the parameter to lie in the interval between β and $\beta + d\beta$, it cannot be given a physical interpretation analogous to the structure function, which contains all the mechanical information about the state of the system (Khinchin, 1949). This is certainly a manifestation of the fact that the random variable whose probability density is (15) cannot be interpreted in the limit-of-frequency, but, rather, in the sense of degree-of-belief.

The first moment calculated from (31) is the second law (2), provided the average and most probable values coincide. The second central moment is

$$S''(\bar{U}) = -\overline{(\beta - \hat{\beta})^2} \doteq -\overline{\Delta\beta^2} < 0 \tag{32}$$

which again shows how concavity is related to thermodynamic stability. The average and most probable values of β will coincide if $\beta(u)$ is an *unbiased* estimator of $\hat{\beta}(\bar{u})$; that is (Lavenda, 1987),

$$\int_0^\infty [\beta(u) - \hat{\beta}(\bar{u})] f(u|\hat{\beta})\, du = 0$$

Differentiating with respect to $\hat{\beta}$ yields

$$-1 + \int_0^\infty [\beta(u) - \hat{\beta}(\bar{u})][L'(\hat{\beta}) - u] f(u|\hat{\beta})\, du = 0$$

or, equivalently,

$$\overline{\Delta\beta\Delta u} \doteq \overline{(\beta - \hat{\beta})(u - \bar{u})} = -1 \tag{33}$$

which is the covariance of the random variables. The correlation coefficient,

$$\rho \doteq \overline{\Delta\beta\Delta u} / [\overline{(\Delta\beta)^2}\, \overline{(\Delta u)^2}]^{1/2} \tag{34}$$

is a measure of the "degree-of-association" between the random variables. Since $\rho \in [-1, 1]$, squaring both sides of (34) and using the value of the covariance given in (33) gives

$$\overline{(\Delta\beta)^2(\Delta u)^2} \geq 1 \tag{35}$$

which is a prototype of a thermodynamic uncertainty relation (Lavenda, 1987).

In the statistical literature, inequalities of the form (35) are known as Cramér-Rao inequalities. These inequalities place lower bounds on the variance. From the relation between the second derivatives of the Legendre transforms (14) and their relations to the variances (30) and (32), it is seen that the equality applies in the uncertainty relation (35). Fisher refers to the negative of (30) as the *amount of* information in the sample. The uncertainty relation gives the notion an intuitive appeal, since the more information we have, the better will be our estimation of the inverse temperature. The inverse of the information represents the lower bound for the variance of any unbiased estimator of the inverse temperature. This interpretation lends further support to the asymmetry of the two densities (9) and (15) in that the likelihood function (15) determines the best estimate of the parameter in terms of the sample mean energy and not contrariwise. The inverse of the variance of the probability density (9) is the lower limit of the variance among *all* unbiased estimators.

5. DUAL PROBABILITY DISTRIBUTIONS

In this section, we derive the probability distributions for intensive variables that can be interpreted in the limit-of-frequency. The error laws for extensive variables will all have a structure similar to (8), which identifies the sample mean as the most probable value of the quantity measured. This is due to their property of *additivity*. There is no reason to suppose that the error laws for intensive variables will identify the mean value with the most probable one, and consequently, their laws of error may have a structure different from (8). But just as the intensive variables are derived quantities in thermodynamics, so, too, will be their error laws. It is precisely the Legendre transform which allows the error laws for intensive variables to be derived from the error laws for their conjugate variables.

Introducing the Legendre transform (10) and

$$L(\beta_i) = \beta_i u_i - S(u_i) \tag{36}$$

into the error law for the energy (8) gives

$$f(\beta_i|\hat{\beta}) = A \exp\{L(\hat{\beta}) - L(\beta_i) - L'(\beta_i)(\hat{\beta} - \beta_i)\} \tag{37}$$

which is a proper probability distribution on account of the concavity criterion of the Legendre transfrom of the entropy. A comparison of the Legendre transforms (19) and (36) shows that we are not claiming ignorance about the initial values of the parameters β_i that determine the state of nature of the m subsystems. We will now show that (37) is a law of error which identifies mean values, other than the sample mean, as the most probable values of the quantity measured.

Again consider the fundamental relations for composite, (22), and individual subsystems, (23), consisting of an ideal gas. According to the second law (2), the composite system will have an inverse temperature $\hat{\beta} = n/\bar{u}$, while those of the individual subsystems will be given by $\beta_i = n/u_i$. The Legendre transform (10), in molar form, is

$$l(\hat{\beta}) = n \log \hat{\beta} + n - n \log n - r \log \bar{v} \tag{38}$$

while

$$l(\beta_i) = n \log \beta_i + n - n \log n - r \log v_i \tag{39}$$

which replace (25) since the temperatures of the subsystems are the sample data. Assuming the gas to be uniform, $v_i = \bar{v}$, the dual (37) of the error law for the energy (8) is

$$f(\beta_i|\hat{\beta}) = A\left(\frac{\hat{\beta}}{\beta_i}\right)^n \exp\left\{-n\hat{\beta}\left(\frac{1}{\beta_i} - \frac{1}{\hat{\beta}}\right)\right\} \tag{40}$$

This is precisely the inverted gamma density. Consider, for the moment, that both u and β are continuous random variables. The inverted gamma density is related the canonical distribution, evaluated at the maximum likelihood value of the scale parameter,

$$f(u|\hat{\beta})\, du = \hat{\beta}\frac{(\hat{\beta}u)^{n-1}}{\Gamma(n)}\, e^{-\hat{\beta}u}\, du \tag{41}$$

by a change of variable. Introducing $u = n/\beta$ into (41) results in

$$f(\beta|\hat{\beta}) = \left(\frac{n}{\beta}\right)\left(\frac{\hat{\beta}}{\beta}\right)^n e^{n(1-\hat{\beta}/\beta)} \tag{42}$$

which will be the same as (40) if we set $A = n/\beta$. The derivation of the inverted gamma density (42) from the proper normalized probability density (41) by a simple one-to-one differentiable change of variable guarantees that it, too, is a proper normalized probability density. We shall now show that the inverted gamma density (40) is the error law leading to the *harmonic* mean as the most probable value of the inverse temperature.

In order that (6), where the prime now stands for differentiation with respect to $\hat{\beta}$, be equivalent to the definition of the harmonic mean,

$$\sum_{i=1}^{m}\left(\frac{1}{\beta_i}-\frac{1}{\hat{\beta}}\right)=0 \tag{43}$$

we require (Keynes, 1921)

$$\frac{f'(\beta_i|\hat{\beta})}{f(\beta_i|\hat{\beta})}+\vartheta''(\hat{\beta})\left(\frac{1}{\beta_i}-\frac{1}{\hat{\beta}}\right)=0 \tag{44}$$

where $\vartheta''(\hat{\beta})$ may be interpreted as a Lagrange multiplier for the constraint (43). Integrating by parts then gives

$$f(\beta_i|\hat{\beta})=A\exp\left\{-\vartheta'(\hat{\beta})\left(\frac{1}{\beta_i}-\frac{1}{\hat{\beta}}\right)+\psi(\hat{\beta})-\psi(\beta_i)\right\} \tag{45}$$

where

$$\psi(x)=\int\frac{\vartheta'(x)}{x^2}\,dx$$

The error law (45) identifies the harmonic mean as the most probable value of the inverse temperature. This law of error coincides with (40) for a Lagrange multiplier given by $\vartheta''=n$.

To grasp the physical significance of this result, consider two identical bodies that are initially at the temperatures T_1 and T_2, where $T_1 > T_2$. A quantity of heat $-\delta Q_1$ is withdrawn from the hotter body and an amount δQ_2 is introduced into the colder one. In a completely *irreversible* process, no work is done, and the final temperature of the composite system is the arithmetic mean of the two initial temperatures, $\hat{T}=\frac{1}{2}(T_1+T_2)$. Owing to the fact that no work is performed, this is the highest attainable temperature that the composite system can achieve. Since $\hat{T}$ is the arithmetic mean, $\hat{\beta}$ is the harmonic mean.

According to the deterministic viewpoint of thermodynamics, the composite system will certainly reach the final temperature $\hat{T}$ after thermal equilibrium has been established. However, the outcome of bringing two bodies, initially at different temperatures, into thermal contact is less than certain, since the quantities of heat that are transferred from one body to the other are the uncontrollable part of the energy. The controllable part of the energy is the work and given that this is zero, all we can surmise is that $\hat{T}$ is the most probable value of the temperature of the composite system. In other words, the act of thermal interaction makes the predicted outcome one of high probability rather than a certainty.

At the other extreme, the process is completely *reversible* and the maximum amount of work is performed. The composite system then attains the lowest temperature possible. We now require (6) to be equivalent to the definition of the geometric mean

$$\hat{T}=\left(\prod_{i=1}^{m} T_i\right)^{1/m} \tag{46}$$

which is the minimum attainable final temperature. We may write (46) equivalently as

$$\sum_{i=1}^{m} \log\left(\frac{T_i}{\hat{T}}\right)=0$$

Following the same procedure as before leads to a law of error

$$f(T_i|\hat{T})=A\exp\left\{\varphi'(\hat{T})\log\left(\frac{T_i}{\hat{T}}\right)+\chi(\hat{T})-\chi(T_i)\right\} \tag{47}$$

where

$$\chi(x)=\int\frac{\varphi'(x)}{x}\,dx$$

which identifies the geometric mean as the most probable value of the temperature measured. We now take note of the fact that if the geometric mean of the observations yields the most probable value of the quantity, the arithmetic mean of the logarithm of the observations will yield the most probable value of the logarithm of the quantity (Keynes, 1921). Since the entropy of an ideal gas is proportional to the logarithm of the temperature, we put $\varphi'=n\log\hat{T}=\bar{S}$, where $\bar{S}$ is the average entropy, and $2\chi(\hat{T})=n(\log\hat{T})^2=\bar{S}^2/n$. This has the effect of transforming the error law for the temperature (47) into the normal law of error

$$f(S_i|\bar{S})=A\exp\left\{-\frac{(S_i-\bar{S})^2}{2n}\right\} \tag{48}$$

for the entropy.

Therefore, in a reversible process the entropy becomes the independent additive variable. The normal law (48) usually arises in the limit of small fluctuations when the entropy is taken as the independent variable (Landau and Lifshitz, 1969). The mean-square fluctuation of the entropy $\overline{(\Delta S)^2}=n$ corresponds to that of an ideal gas and the relative fluctuation is proportional to $n^{-1/2}$, which tends to zero as the number of degrees-of-freedom increases. This small-fluctuation limit is incorporated into the error law

$$f(T_i|\hat{T})=A\exp\left\{-\frac{n}{2}\left(\log\frac{T_i}{\hat{T}}\right)^2\right\} \tag{49}$$

by the fact that the average entropy is proportional to the logarithm of the geometric mean value of the temperature which is the lowest attainable temperature. We therefore expect fluctuations in reversible processes to be smaller than in irreversible processes. Furthermore, (49) predicts that positive and negative deviations of the same absolute magnitude are not equally likely, while the corresponding normal law for the entropies (48) satisfies the condition that negative and positive errors of the same absolute magnitude are equally likely.

ACKNOWLEDGMENTS

The work of B.H.L. was supported in part by the Ministry of Universities and Scientific Research and the National Science Council.

REFERENCES

Callen, H. B. (1985). *Thermodynamics*, 2nd ed., Wiley, New York, Chapter 2.

Galgani, L., and Scotti, A. (1970). *Pure and Applied Chemistry*, **22**, 229.

Gibbs, J. W. (1902). *Elementary Principles in Statistical Mechanics*, Yale University Press, New Haven, Connecticut, p. 104.

Greene, R. F., and Callen, H. B. (1951). *Physical Review*, **83**, 1231.

Hardy, G. H., Littlewood, J. E., and Pólya, G. (1952). *Inequalities*, 2nd ed., Cambridge University Press, Cambridge, p. 77.

Jeffreys, H. (1961). *Theory of Probability*, 3rd ed., Oxford University Press, Oxford, p. 122.

Keynes, J. M. (1921). *A Treatise on Probability*, St. Martin's Press, New York, pp. 218–228.

Khinchin, A. I. (1949). *Mathematical Foundations of Statistical Mechanics*, Dover, New York, p. 37.

Landau, L. D., and Lifshitz, E. M. (1969). *Statistical Physics*, 2nd ed., Pergamon Press, Oxford, Section 114.

Lavenda, B. H. (1987). *International Journal of Theoretical Physics*, **26**, 1069.

Lavenda, B. H. (1988). *International Journal of Theoretical Physics*, **27**, 451.

Lavenda, B. H., and Dunning-Davies, J. (1990). *Foundations of Physics Letters*, **3**, 435.

Lavenda, B. H., and Dunning-Davies, J. (in press). *Annalen der Physik.*

Lavenda, B. H., and Scherer, C. (1988). *Nuovo Cimento*, **11**(6).

Lindley, D. (1970). *Introduction to Probability and Statistics*, Part 2, Cambridge University Press, Cambridge, pp. 46–58.

Mandelbrot, B. (1956). *Institute of Radio Engineers Transactions Information Theory*, **IT-2**, 190.

Mandelbrot, B. (1962). *Annals of Mathematics and Statistics*, **33**, 1021.

Tikochinsky, Y., and Levine, R. D. (1984). *Journal of Mathematical Physics*, **25**, 2160.