International Journal of Theoretical Physics, Vol. 34, No. 4, 1995

Subordination and Bayes' Theorem in the Thermodynamics of Composite Systems

B. H. Lavenda[1]

Received June 13, 1994

A completely new derivation of the probability distribution in energy of a small component of a composite system, which was previously derived from the composition law of the structure function and Boltzmann's principle, is given by the method of subordination in which the conjugate, intensive parameter is randomized and is shown to possess a Bayes distribution. The derivation allows an extension of the class of structure functions to include strictly stable laws which possess domains of attraction like the normal law.

1. THE COMPOSITION LAW

The composition law for the structure function determines the distribution in energy of a small component of a composite system (Khinchin, 1949). The derivation rests upon the assumption that the phase space volume element is the direct product of the phase spaces of all components (Khinchin, 1949, pp. 40–41), which can be taken to be synonymous with the assumption of the statistical independence of the components of the composite system (Lavenda, 1991). As such the composition law must be an asymptotic result for systems comprised of a very large number of degrees of freedom.

Boltzmann's principle,

$$S(\varepsilon) = \ln \Omega(\varepsilon) + \text{const} \tag{1}$$

relates the entropy of any component of a composite system $S(\varepsilon)$ whose energy is ε to the surface area $\Omega(\varepsilon)$ enclosing the volume of phase space occupied by the component, where temperature is measured in energy units for which Boltzmann's constant is unity. On the strength of this principle, the probability density for the energy of a small component is seen to be an

[1] Università degli Studi, Camerino 62032 (MC), Italy.

0020-7748/95/0400-0615$07.50/0

exponential function of the entropy difference between the sum of the entropies of the n components and n times the entropy of the composite system [cf. the first line in equation (5)], instead of the entropy of the composite system itself (Khinchin, 1949, p. 145).

It is a common practice in statistical estimation to estimate parameters appearing in a distribution in terms of a statistic, like the sample mean. This means that the parameters are something more than "mere" parameters, which in thermodynamics characterize the reservoirs, and can undergo randomization. Whereas the extensive thermodynamic random variables are distributed in terms of the "frequency" of occurrence, their intensive conjugate random variables are distributed according to "degree-of-belief" that certain values are more probable than others (Lavenda, 1991, pp. 204–214).

Equipped with the distribution in the intensive variable, we will show that the distribution of an extensive, and globally conserved, quantity, like that of the energy of a small component of a composite system, is actually derived from the process of subordination. The probability distribution of the subordinated process will thus always be associated with the microcanonical ensemble, and allow a complete characterization of the mechanical structure of the physical system. The method of subordination will also enable us to enlarge the class of acceptable structure functions (Lavenda and Florio, 1992a) and justify the generalized Boltzmann principle of extreme-value distributions (Lavenda and Florio, 1992b).

2. THERMODYNAMIC MEANING OF SUBORDINATION

For purposes of illustration, we choose the gamma density for the energy ε,

$$f(\varepsilon|\beta_0) = \beta_0 \frac{(\beta_0\varepsilon)^{m-1}}{\Gamma(m)} e^{-\beta_0\varepsilon} \tag{2}$$

which is parametrized by the inverse temperature β_0, where m represents half the number of degrees of freedom. The parameter is usually referred to as the "state of nature," which characterizes the heat reservoir in which the system is in thermal contact. As a problem in statistical estimation, the inverse temperature can be estimated in terms of the sample mean which is constructed from observations made on the internal energy. Therefore, it, too, must fluctuate.

In order to convert the gamma density (2) for the energy into a distribution for the inverse temperature, we introduce the transform $\beta_0\varepsilon = \beta\varepsilon_0$, where ε_0 can be taken as the energy of the reservoir, which if large enough can be considered as a constant. Introducing this transform into (2) results in a new gamma density

$$f(\beta|\varepsilon_0) = \varepsilon_0 \frac{(\varepsilon_0\beta)^{m-1}}{\Gamma(m)} e^{-\varepsilon_0\beta} \tag{3}$$

for the inverse temperature, which is now parametrized by the energy of the reservoir, ε_0.

If we want the distribution in the energy, we must first establish thermal equilibrium: $\beta = \beta_0$. This is the *directing process* (Feller, 1971) whose density is (3). Multiplying it by (2) and integrating over all values of β gives

$$\begin{aligned} f(\varepsilon|\varepsilon_0) &= \int_0^\infty f(\varepsilon|\beta) f(\beta|\varepsilon_0)\, d\beta \\ &= \frac{1}{B(m, m)} \frac{\varepsilon_0(\varepsilon\varepsilon_0)^{m-1}}{(\varepsilon + \varepsilon_0)^{2m}} \end{aligned} \tag{4}$$

where $B(m, m) = \Gamma^2(m)/\Gamma(2m)$ is the beta function. With the change of variable $p = \varepsilon/(\varepsilon + \varepsilon_0)$, (4) becomes the beta distribution (Lavenda, 1991)

$$f(p) = \frac{p^{m-1}(1-p)^{m-1}}{B(m, m)} = \begin{cases} \exp\{S(\varepsilon) + S(\varepsilon_0) - 2S(\varepsilon + \varepsilon_0) - \ln B(m, m)\} \\ \exp\{\Delta S(\varepsilon) + \Delta S(\varepsilon_0) - \ln B(m, m)\} \end{cases} \tag{5}$$

where the entropy is

$$S(x) = (m - 1)\ln x \tag{6}$$

while the reduction in entropy is

$$\Delta S(x) = (m - 1) \ln\left(\frac{x}{x + x_0}\right) \tag{7}$$

For an entropy of the form (6), its concavity property ensures that it will be a monotonically increasing function,

$$2S(\varepsilon + \varepsilon_0) \geq 2S\left(\frac{\varepsilon + \varepsilon_0}{2}\right) \geq S(\varepsilon) + S(\varepsilon_0)$$

which, in turn, ensures that (5) will be a proper probability density.

Moreover, the second equality in (5) holds only for $m > 1$ (Lavenda, 1994), which is a manifestation of *asymptotic independence*. The distribution can be expressed entirely in terms of the entropies of the composite system, which are statistically independent.

Like the gamma density (2), the beta density (4) is parametrized by $m > 0$. For a single degree of freedom, we obtain a particular form of the beta density (of the first kind),

$$f(\vartheta) = \frac{1}{\pi} \frac{1}{[\vartheta(1 - \vartheta)]^{1/2}} \tag{8}$$

where $\vartheta = \varepsilon/(\varepsilon + \varepsilon_0)$, called the arcsine probability density, so named because its integral is the arcsine function. The arcsine probability density is ubiquitous in the theory of random walks. No thermodynamic characterization for this distribution is possible (Lavenda, 1995). As the number of degrees of freedom increases from two to three, the probability density transforms from a uniform to a $\cap$-shaped distribution. For $m > 2$, the beta density (4) has a mode at 1/2, implying that the most probable case is where the two subsystems share the same amount of energy. A thermodynamic characterization for such distributions is possible, where the subordinated distribution becomes asymptotically independent.

It is important to recognize the fact that the subordinated probability density (4) has the same form as the prior probability density of the microcanonical ensemble. In fact, the canonical probability density (2) has been constructed from the Laplace transform of prior probability density,

$$\Omega(\varepsilon) = \varepsilon^{m-1} \tag{9}$$

namely,

$$Z(\beta_0) = \int_0^\infty e^{-\beta_0\varepsilon}\Omega(\varepsilon)\, d\varepsilon \tag{10}$$

in such a way that the associated probability density

$$f(\varepsilon \mid \beta_0) = \frac{e^{-\beta_0\varepsilon}}{Z(\beta_0)}\, \Omega(\varepsilon)$$

is properly normalized.[2]

A rather remarkable feature is that the subordinated probability densities, like (4), will always turn out to be proper probability densities, at least in the asymptotic limit of a large number of degrees of freedom or for large values of the variate [cf. equation (26) below]. The most important case where subordination comes into play is when the same probability density determines both the probability distribution of the variable and the distribution

[2] A similar situation often appears in the Bayes formulation, where the prior probability density, representing a vague knowledge about the parameter (Jeffreys, 1961), can also be improper. The associated probability density will, however, always turn out to be a proper probability density [cf. discussion following equation (12) below].

of the randomized parameter. This choice of the probability density for the randomized parameter coincides with the one derived from Bayes' theorem (Lavenda, 1991, pp. 192–195).

In order to derive the Bayes distribution, we cast the gamma density (2) as the law of error

$$f(\varepsilon\,|\,\beta_0) = A(\beta_0)e^{S(\varepsilon)-S(\varepsilon_0)-S'(\varepsilon_0)(\varepsilon-\varepsilon_0)}$$

where the prime denotes differentiation,

$$\varepsilon_0 = (m - 1)/\beta_0 \tag{11}$$

the norming constant $A(\beta_0) = \beta_0$, and the gamma function is substituted for $(m - 1)^{m-1}e^{-(m-1)}$. Introducing the dual to the entropy through the Legendre transform

$$\ln Z(\beta_0) = S(\varepsilon_0) - S'(\varepsilon_0)\varepsilon = -(m - 1)\ln \beta_0 + \ln \Gamma(m)$$

where (11) has been used to obtain the second equality, we obtain the Bayes distribution (Lavenda, 1991, p. 206)

$$f(\beta\,|\,\varepsilon_0) = A(\varepsilon_0)e^{-\ln Z(\beta)+\ln Z(\beta_0)+(\ln Z)'(\beta_0)(\beta-\beta_0)}\Omega(\beta) \tag{12}$$

upon interchanging the endpoints of the interval, where $\Omega(\beta)$ is the prior probability density and $A(\varepsilon_0)$ is a norming constant.

By involving the law of equipartition of energy in the form (11), which was the form found by Gibbs (1902), we are naturally led to the Bayes–Laplace principle of setting the prior probability density equal to the uniform probability density. In the absence of any prior information about β, the proponents of the Bayes–Laplace principle argue that only the uniform prior would correspond to knowing nothing or providing for a uniform assessment over the entire positive half-axis. In this case, we would set the norming constant $A(\varepsilon_0) = \varepsilon_0$.

This choice of the prior probability density was criticized by Jeffreys (1961) precisely because we *do know* that β can take on values only on the positive half-axis. Jeffreys argued that we should take the logarithm of the parameter as the prior distribution so that its density is

$$\Omega(\beta) = \beta^{-1} \tag{13}$$

This requires that the law of equipartition of energy be given by

$$\varepsilon_0 = m/\beta_0 \tag{14}$$

in which case there is no need of the norming constant. In the latter case, we note a parallelism between the improper density chosen for the structure function (9) and Jeffreys' rule (13). It is interesting to remark that for small values of m, which was definitely not excluded in Gibbs' analysis (Gibbs, 1902, p. 118), we can discriminate between the Bayes–Laplace principle and Jeffreys' rule, (13). In either case, the change of variable $\varepsilon\beta_0 = \varepsilon_0\beta$, used to go from (2) to (3), is essentially a short-cut to the Bayes distribution.

Another observation to be made is the intimate relation between the Laplace and Legendre transforms. The Legendre transform selects out that value of the parameter of the distribution which maximizes overwhelmingly the probability distribution.

Furthermore, provided the distribution belongs to the exponential family, the subordination procedure leads to a beta distribution of the second kind. Hence, *subordination can be considered as the probabilistic origin of power laws in physics.* In thermodynamics, the process of subordination transforms the probability distribution of the canonical ensemble to that of the microcanonical ensemble. Thus, subordination is to be associated with the introduction of a constraint, related to the appearance of ε_0, that leads to a reduction in the ensemble description.

3. SUBORDINATION OF EXTREME-VALUE DISTRIBUTIONS

On the strength that the probability density of the subordinate process has the same form as the prior probability density, just as (4) gives the same expression for the entropy as (9), subordination may be used to justify the expressions for the entropy reduction in the distribution of extreme values. Hitherto, we have relied essentially on the property of concavity, and the Legendre transform of the generating function, whose form is known (Lavenda and Florio, 1992b). The Legendre transformation thus converts a convex function into a concave one.

For illustrative purposes, we choose the Lévy probability density

$$f(x) = \frac{1}{2(\pi x^3)^{1/2}}\, e^{-1/4x} = \frac{S'(x)}{[\pi(-\Delta S(x))]^{1/2}}\, e^{\Delta S(x)}, \qquad x > 0 \tag{15}$$

because it is the only strictly stable law known in closed form.[3] The entropy reduction,

$$\Delta S(x) = -\frac{1}{4x} \tag{16}$$

has been identified as the Legendre transform of the logarithm of the Laplace transform

[3]This is a consequence of the fact that the generating function is given in terms of a Basset function of order 1/2, which can be expressed in terms of elementary functions [cf. equation (17) below.]

$$\frac{1}{2\sqrt{\pi}}\int_0^{\infty}\frac{1}{x^{3/2}}\,e^{-\lambda_0 x-1/4x}\,dx=\left(\frac{2}{\pi}\right)^{1/2}\lambda_0^{1/4}K_{1/2}(\sqrt{\lambda_0})=e^{(-\lambda_0)^{1/2}} \tag{17}$$

where $K_{1/2}$ is a Bessel function of imaginary argument, or a Basset function. The entropy reduction (16) is a concave function which, because of its superadditivity,

$$\Delta S(x_1+x_2)\geq\Delta S(x_1)+\Delta S(x_2)$$

is also a monotonically increasing function. Superadditivity is a sufficient, but not a necessary, condition for a function to be monotonically increasing. Even subadditive entropies, like that of black radiation, are also monotonically increasing functions.

The "canonicalized" probability density can be written as the error law

$$\begin{aligned} f(x|\lambda_0) &= \frac{S'(x)}{[\pi(-\Delta S(x))]^{1/2}}\,e^{\Delta S(x)-\Delta S(x_0)-S'(x_0)(x-x_0)} \\ &= \frac{1}{2(\pi x^3)^{1/2}}\,e^{-[(\lambda_0 x)^{1/2}-1/2x^{1/2}]^2} \end{aligned} \tag{18}$$

which brings out the profound relation between least squares and maximum probability. The conjugate parameter λ_0 has been introduced as the derivative of the entropy reduction (16),

$$S'(x_0)=\frac{1}{4x_0^2}=\lambda_0 \tag{19}$$

which is the common practice in Legendre transforms, since the dual of (16) is the logarithm of the generating function,

$$\ln Z(\lambda_0)(\lambda_0)=\Delta S(x_0)-S'(x_0)x_0=-\sqrt{\lambda_0} \tag{20}$$

In order to obtain the second equality in (20), we have introduced the definition of the conjugate variable, (19).

The exponent of the Bayes probability density is derived from the error law (18) by introducing the Legendre transform (20), and a similar expression for $\ln(\lambda)$, and interchanging endpoints. Multiplying it by the prior probability density $f(\lambda)$ and a suitable norming constant, we get

$$\begin{aligned} f(\lambda|x_0) &= A(x_0)e^{-\ln Z(\lambda)+\ln Z(\lambda_0)+(\ln Z)'(\lambda_0)(\lambda-\lambda_0)}\Omega(\lambda) \\ &= A(x_0)e^{-[(x_0\lambda)^{1/2}-1/2x_0^{1/2}]^2}\Omega(\lambda) \end{aligned} \tag{21}$$

where λ_0 and x_0 are related by

$$(\ln Z)'(\lambda_0)=-x_0 \tag{22}$$

This relation is obtained by minimizing the exponent in (21), which is known as the likelihood function (Lavenda, 1991, p. 192). The extremum condition (22) is in complete accord with (19). That this is, indeed, the maximum likelihood value of λ follows from the convexity property of $\ln Z(\lambda)$.

Since λ_0 is the only minimum of $[(x_0\lambda)^{1/2} - 1/2x_0^{1/2}]^2$ in the exponent of (21) on $[0, \infty]$, the dominant contribution to the integral of (21) will come from the neighborhood of λ_0. This observation permits us to use Laplace's method in which the limits of integration are extended to $\pm\infty$, and the quadratic form is developed in a Taylor series about λ_0. To lowest order we get

$$\int_{-\infty}^{\infty} e^{-x_0^3(\lambda-\lambda_0)^2}\, d\lambda = \left(\frac{\pi}{x_0^3}\right)^{1/2} \tag{23}$$

Thus, the constant of integration in (21) is $A(x_0) = (x_0^3/\pi)^{1/2}$. As for the initial probability density $\Omega(\lambda)$, it suffices to note that since the limits of integration in (23) have been extended to $\pm\infty$, we must choose the Bayes–Laplace principle and consider a uniform distribution.

Invoking the condition of equilibrium, $\lambda = \lambda_0$, between (18) and (21), we complete the square in the exponent of their product to obtain the quadratic form $\{[(x + x_0)\lambda]^{1/2} - 1/(x + x_0)^{1/2}\}^2$. This has a unique minimum at $\lambda_0 = 1/(x + x_0)^2$, which will be close to the minimum in the Bayes distribution for $x \approx x_0$. Again using Laplace's method, we find the probability density of the subordinate process to be given by

$$\begin{aligned} f(x|x_0) &= \int_{-\infty}^{\infty} f(x|\lambda) f(\lambda|x_0)\, d\lambda \\ &= \frac{1}{2\pi}\left(\frac{x_0}{x}\right)^{3/2} \int_{-\infty}^{\infty} e^{-1/4(x+x_0)^3(\lambda-\lambda_0)^2}\, d\lambda\; e^{1/(x+x_0)-1/4x-1/4x_0} \\ &= \frac{1}{\sqrt{\pi}}\left[\frac{x_0}{x(x+x_0)}\right]^{3/2} e^{1/(x+x_0)-1/4x-1/4x_0} \end{aligned} \tag{24}$$

In terms of the entropy reduction (16), we can express (24) as

$$f(x|x_0) = \frac{2}{\sqrt{\pi}} \frac{-S''(x)}{[-\Delta S(x) - \Delta S(x_0)]^{3/2}}\, e^{\Delta S(x)+\Delta S(x_0)-4\Delta S(x+x_0)} \tag{25}$$

which can be taken as Boltzmann's principle for the probability distribution of one component in a composite system when the prior probability distribution is a strictly stable law. In other words, it is the stable law counterpart of (5), which is a consequence of the central limit theorem. This offers further justification for identifying (16) as the entropy reduction of the Lévy distribution. Moreover, the fact that the entropy reduction (16) is a monotonically increasing function ensures that the exponent in (25) is negative definite.

The probability distribution (15), which is subordinated to the canonicalized Lévy distribution (18), is exactly normalized in the limit of large x_0. This is easily seen by introducing the transformation $u = 1/4(1/x + 1/x_0)$ into (24), and integrating to obtain

$$\frac{e^{1/x_0}}{\sqrt{\pi}} \int_{1/4x_0}^{\infty} \left(2 - \frac{1}{2x_0 u}\right) \frac{1}{\sqrt{u}} e^{-u - 1/4x_0^2 u}\, du \tag{26}$$

For large values of x_0, the lower limit is approximately zero and the integral is unity. We thus conclude that the asymptotic limit of large values of the variate of a strictly stable distribution is analogous to the large sample limit in the central limit theorem. In fact, we will now show that the two domains of attraction are, in a certain sense, complementary to one another.

Interestingly enough, the strictly stable distribution (15) can be derived from a particular form of the gamma density

$$f(u) = \frac{u^{-1/2}}{\Gamma(1/2)} e^{-u} \tag{27}$$

by the inverse transform $u = 1/4x$, known as the chi-square distribution. The chi-square distribution (27) has only a single degree of freedom, for which the central limit theorem is certainly not applicable. In other words, the process corresponding to (27) would have vanishing entropy. Yet, by a simple inverse transform, it can be converted into a strictly stable law of characteristic exponent 1/2, where asymptotic independence is achieved for large values of the variate. Hence, the asymptotic independence of strictly stable laws is complementary to that of the central limit theorem.

ACKNOWLEDGMENTS

This work was supported, in part, by the EC Human Capital and Mobility Network Programme (Contract ERBCHRXCT920007), the Italian Ministry of University and Research (MURST), and the Italian National Science Council (CNR).

REFERENCES

Feller, W. (1971). *An Introduction to Probability Theory and Its Applications*, Vol. II, 2nd ed., Wiley, New York, §X.7.

Gibbs, J. W. (1902). *Elementary Principles in Statistical Mechanics,* Yale University Press, New Haven, p. 120, formula (384).

Jeffreys, H. (1961). *Theory of Probability*, 3rd ed., Clarendon Press, Oxford, p. 119.
Khinchin, A. I. (1949). *Mathematical Foundations of Statistical Mechanics*, Dover, New York, pp. 71–75.
Lavenda, B. H. (1991). *Statistical Physics: A Probabilistic Approach*, Wiley-Interscience, New York, pp. 182–185.
Lavenda, B. H. (1995). Limitations of Boltzmann's principle, *International Journal of Theoretical Physics*, **34**, 605.
Lavenda, B. H., and Florio, A. (1992a). *Foundations of Physics Letters*, **5**, 457.
Lavenda, B. H., and Florio, A. (1992b). *International Journal of Theoretical Physics*, **31**, 1455.