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

Limitations of Boltzmann's Principle

B. H. Lavenda[1]

Received June 13, 1994

The usual form of Boltzmann's principle assures that maximum entropy, or entropy reduction, occurs with maximum probability, implying a unimodal distribution. Boltzmann's principle cannot be applied to nonunimodal distributions, like the arcsine law, because the entropy may be concave only over a limited portion of the interval. The method of subordination shows that the arcsine distribution corresponds to a process with a single degree of freedom, thereby confirming the invalidation of Boltzmann's principle. The fractalization of time leads to a new distribution in which arcsine and Cauchy distributions can coexist simultaneously for nonintegral degrees of freedom between $\sqrt{2}$ and 2.

1. THE MANY FACETS OF BOLTZMANN'S PRINCIPLE

Boltzmann's principle,

$$S = k \ln \Omega + \text{const} \tag{1}$$

relating entropy S logarithmically to the "thermodynamic" probability Ω of a given state means many things to different people. We shall assume that temperature is measured in energy units where Boltzmann's constant is unity.

A system at thermodynamic equilibrium will pass through all states, Γ_1, $\Gamma_2, \ldots, \Gamma_n$. In the course of a long time t_0, the system will spend a total of t_i time units in the state Γ_i. The longer the system spends in any of these states, the more probable it will be, and consequently we may expect that the thermodynamic probability $\Omega(\Gamma_i)$ for the realization of the state Γ_i will be proportional to t_i. More precisely, we shall set it equal to the fraction t_i/t_0 and speak about the reduction in entropy

$$\Delta S(t_i) = \ln\left(\frac{t_i}{t_0}\right) \tag{2}$$

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

0020-7748/95/0400-0605$07.50/0

where the entropy of the reference state $S_0 = \ln t_0 + \text{const}$ is the entropy it would have if it did not leave that state in the course of the long time interval t_0.

Let $F(t_i | t_0)$ represent the probability for the realization of the time interval t_i. Then it seems reasonable that we can replace the proper fraction t_i / t_0 in Boltzmann's principle (2) by $F(t_i | t_0)$ to obtain a further representation of Boltzmann's principle

$$\Delta S(t_i) = \ln F(t_i | t_0) \tag{3}$$

which is necessarily negative on account of the fact that $F(t_i | t_0)$ is a proper fraction.

To simplify matters, let us consider only two states: the positive and the negative axes in one spatial dimension. We shall refer to t_0 as the length of a path. Let us first apply Boltzmann's principle in the form (1). There are 2^{t_0} paths of length t_0 which contribute an entropy of $S_0 = t_0 \ln 2$. The length of the path is made up of a number p of positive steps and a number of negative steps q such that $t_0 = p + q$. The excess $k = p - q$ will determine where the system is after t_0 steps. The number of different paths to k is the number of ways that p positive, or q negative, steps can be chosen from the total $t_0 = p + q$ number of steps (Feller, 1968)

$$N_{2t_0, |t_0 - k|} = \binom{t_0}{(t_0 + k)/2} \tag{4}$$

where the binomial coefficient is understood to be zero if k and t_0 are not of the same parity. Using Stirling's approximation in the form

$$x! \approx x^x e^{-x} \tag{5}$$

gives an entropy of

$$\begin{aligned} S(k) &= \ln N_{2t_0, |t_0 - k|} \\ &= t_0 \ln(t_0) - \left(\frac{t_0 + k}{2}\right) \ln\left(\frac{t_0 + k}{2}\right) - \left(\frac{t_0 - k}{2}\right) \ln\left(\frac{t_0 - k}{2}\right) \end{aligned}$$

which is maximized by the value $k = 0$. That is to say, for $k = 0$, (4) becomes the number of paths that return to the origin, for which the entropy has the maximum value

$$S(k = 0) = t_0 \ln 2$$

which is the entropy of 2^{t_0} of length t_0, and consequently the entropy difference vanishes.

Doubling the length of the path, the probability of a return to the origin at epoch $2t_0$ is

$$F(t_0) = \binom{2t_0}{t_0} 2^{-2t_0} = \exp\{\Delta S(k = 0)\} \tag{6}$$

which is certain, with probability 1, since the entropy difference $\Delta S(k = 0) = S(k = 0) - S_0$ vanishes. This is indeed surprising, since it is independent of the value of t_0. The problem has to do with the crudeness of Stirling's approximation (5). A more refined estimate of the factorial would be

$$x! \approx x^x e^{-x}(2\pi x)^{1/2} \tag{7}$$

Using this approximation to evaluate the factorials in the binomial coefficient in (6) leads to

$$F(t_0) = \frac{1}{(\pi t_0)^{1/2}} \tag{8}$$

which holds even for moderate values of t_0 (Feller, 1968, p. 79). This is a rather counterintuitive result, for it states that the longer the time is, the less probable it will be for the system to return to the origin. We now show that it also leads to an erroneous expression for the change in entropy.

The probability that in the time interval $2t_0$, the particle will spend $2t$ units of time on the positive axis, and hence $2(t_0 - t)$ units on the negative axis, is

$$F(t|t_0) = \binom{2t}{t}\binom{2(t_0 - t)}{t_0 - t} 2^{-2t_0}, \qquad t = 0, 1, \ldots, t_0 \tag{9}$$

This is the same as the probability that up to epoch $2t_0$, the last visit to the origin occurred at epoch $2t$ (Feller, 1968, p. 79). In other words, (9) can be viewed as the product of two probabilities of return to the origin,

$$F(t|t_0) = \binom{2t}{t} 2^{-2t} \times \binom{2(t_0 - t)}{t_0 - t} 2^{-2(t_0 - t)} \tag{10}$$

for the return in $2t$ units of time to the origin on the positive axis and in $2(t_0 - t)$ units of time on the negative axis.

We already know that Stirling's approximation (5) will not suffice, and application of (7) to the binomial coefficient in (9) converts what was a *discrete* distribution into a *continuous* probability density,

$$f(t|t_0) = \frac{t_0}{\pi[t(t_0 - t)]^{1/2}} \tag{11}$$

This is the arcsine probability density because its integral is

$$F(t|t_0) = \frac{2}{\pi} \sin^{-1}\left(\frac{t}{t_0}\right)^{1/2} \tag{12}$$

The arcsine probability density (11) has the curious property that its weight is concentrated at the extremes of the interval with minimum probability at the midpoint, as shown in Fig. 1.

Since we possess the probability distribution (9), we might be tempted to apply Boltzmann's principle in the form (3). We then find that although $\ln[2\sin^{-1}(\sqrt{x})/\pi]$ is negative semidefinite on the entire interval [0, 1], it is not a concave function on the same interval.[2] Hence, Boltzmann's principle in the form (3) is to be excluded.

Availing ourselves of the approximation

$$\Pr(t \leq \mathbf{T} \leq t + \Delta t) \approx f(t)\,\Delta t$$

where we have used the mean-value theorem and evaluated $f(\psi)$ at t and $t \leq \psi \leq t + \Delta t$, for a small latitude Δt in the specification of the probability, we can replace the discrete $F(t)$ in Boltzmann's principle by $f(t)\,\Delta t$; we get

$$\Delta S(t) = \ln f(t|t_0) + \ln \Delta t$$

Since variations in the statistical latitude Δt are so small that they can be neglected with respect to variations in the first factor (Onsager, 1931), Boltzmann's principle is essentially

$$\Delta S(t) = \ln f(t) + \text{const} \tag{13}$$

Introducing the arcsine probability density (11) into (13) leads to

$$\Delta S(t) = -\frac{1}{2}\ln t - \frac{1}{2}\ln(t_0 - t) + \text{const} \tag{14}$$

If the entropy is to be identified with a convex function, possessing a minimum at $t = t_0/2$, which would normally be the most probable value, then it would

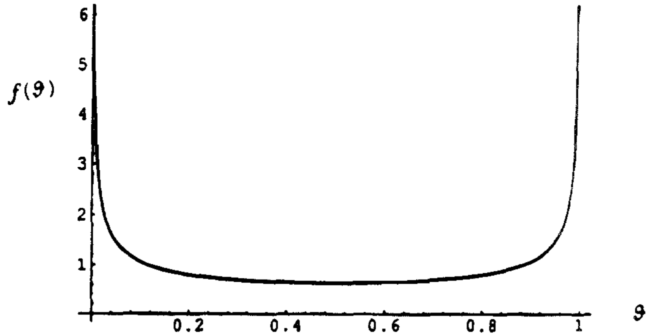

Fig. 1. The arcsine density.

[2]It transforms from a concave to a convex function at $x \approx 0.72106$.

lead to a contradiction with the second law. Hence, expression (14) is to be ruled out on the grounds that it is a convex function and consequently violates the second law.

Since the arcsine probability density (11) has almost all of its weight concentrated at the extremes of the interval $[0, t_0]$, we might be tempted to consider it in the limits $t << t_0$ or $\tau << \tau_0$, where $\tau/\tau_0 = 1 - t/t_0$. On account of the fact that the distribution is symmetric, we need only treat a single limit. For small x, $\sin^{-1} x \approx x$ and, in this limit, Boltzmann's principle (3) becomes

$$\Delta S(t) = \frac{1}{2} \ln\left(\frac{t}{t_0}\right) + \text{const} \tag{15}$$

Although this has the same form as (2), it is only valid in the asymptotic limit for $t << t_0$. This implies that the entropy reduction is very small and consequently it gives the erroneous result of indicating a low probability, as compared to (2).

As thermodynamic equilibrium is reached, the different regions of phase space $\Gamma_1, \ldots, \Gamma_n$ of equal extent tend to accommodate the system for equal periods of time. Consider two such regions for which the system spends t_1 and $t_2 = t_0 - t_1$ amounts of time. The entropy reduction

$$\Delta S(t_1) = \ln\left(\frac{t_1}{t_0}\right) + \ln\left(\frac{t_0 - t_1}{t_0}\right) \tag{16}$$

will be maximum when the system spends equal amounts of time in the two regions of phase space, $t_1^* = t_0/2$. This gives a maximum entropy reduction $\Delta S(t_0/2) = -2 \ln 2$. In contrast, (15) is valid for small t, where the entropy reduction is very small. A similar conclusion can be made for an entropy reduction given by $\Delta S(\tau) = \frac{1}{2}\ln(\tau/\tau_0)$, where again $\tau/\tau_0 = 1 - t/t_0$. Therefore, even though (15) is a concave function, it does not represent a physically acceptable form of the entropy reduction: *Either the entropy reduction is valid over the entire interval or it is valid nowhere in the interval.*

2. FRACTAL ARCSINE LAWS THROUGH SUBORDINATION

Subordination is a very powerful method to derive the probability distribution of one component of a composite system. In so doing, it supersedes the composition law for the structure function (Khinchin, 1949), and gives a correct identification of the entropy reduction. The subordinated probability distributions that we will be dealing with here are power laws, and the subordination method by which they are derived can be considered as the origin of such distributions (Lavenda, 1994).

We now specify that the particle diffusing in one dimension is a Brownian particle whose space and temporal evolutions are connected through

$$r = r_0\left(\frac{t_0}{t}\right)^{1/2} \tag{17}$$

where r_0 and t_0 are two characteristic space and time cutoffs. We shall refer to (17) as the Lévy transform, after Lévy (1965), who used it to derive the strictly stable distribution of characteristic exponent 1/2 from the half-normal or Brownian motion probability density

$$f(r|t_0) = \left(\frac{2}{\pi t_0}\right)^{1/2} e^{-r^2/2t_0} \tag{18}$$

This is to say that when (17) is applied to (18) there results

$$f(t|r_0) = \frac{1}{(2\pi t)^{1/2}} \frac{r_0}{t} e^{-r_0^2/2t} \tag{19}$$

which is the only strictly-stable distribution known in closed form. We may say that the Lévy transform (17) has generated a time distribution whose probability density is (19) by *randomizing* time. Alternatively, due to the symmetry of the process, we can say that the application of Lévy's transformation on (19) *randomizes* space and generates a probability density given by (18).

Suppose that we consider the randomization of the spatial increments so that Lévy's density (19) is the transition probability and the *directing* process (Feller, 1971) has a density given by (18). Then, equating $r = r_0$ in (18) and (19) and integrating over all values of r gives

$$\begin{aligned} f(t|t_0) &= \int_0^\infty f(t|r)f(r|t_0)\,dr \\ &= \frac{1}{\pi}\left(\frac{t_0}{t}\right)^{1/2} \frac{1}{t+t_0} \end{aligned} \tag{20}$$

the inverted beta density, or a beta density of the second kind (Kendall and Stuart, 1969), as the subordinated probability density. Expressed in terms of the fraction $\vartheta = t/(t + t_0)$, (20) becomes

$$f(\vartheta) = \frac{1}{\pi[\vartheta(1-\vartheta)]^{1/2}} \tag{21}$$

Thus, we have derived the arcsine probability density (21) from the method of subordination. There is yet another way at arriving at the arcsine density, and one that will shed some light on the nature of the probability density in relation to Boltzmann's principle. Assume that the frequency ω is distributed according to the gamma density

$$f(\omega|t_0) = t_0 \frac{(\omega t_0)^{m-1}}{\Gamma(\omega)} e^{-\omega t_0} \tag{22}$$

where t_0 is a parameter which we may think of as a characteristic time. Randomizing this time and setting $\omega t_0 = \omega_0 t$ in (22) give the Bayes distribution (Lavenda, 1994)

$$f(t|\omega_0) = \omega_0 \frac{(\omega_0 t)^{m-1}}{\Gamma(m)} e^{-\omega_0 t} \tag{23}$$

where the frequency ω_0 is now the parameter.

Imposing the condition of resonance $\omega = \omega_0$, we find that the subordinated process has the temporal probability density

$$f(t|t_0) = \int_0^\infty f(\omega|t_0) f(t|\omega)\, d\omega$$

$$= \frac{1}{B(m, m)} \frac{t^{m-1} t_0^m}{(t + t_0)^m} \tag{24}$$

where $B(m, m) = \Gamma^2(m)/\Gamma(2m)$ is the beta function. Now, the subordinated probability density (24) will coincide with the inverted beta density (20) for $m = 1/2$. Since m is half the number of degrees of freedom, we conclude that an entropy cannot be defined for a process with a single degree of freedom.

The arcsine probability density (21) thus corresponds to

$$f(\vartheta) = \frac{[\vartheta(1 - \vartheta)]^{m-1}}{B(m, m)} \tag{25}$$

for $m = 1/2$. The beta density (25) transforms from a $\cup$-shaped to $\cap$-shaped form as the number of degrees of freedom is increased from one to three. The intermediary case of two degrees of freedom corresponds to the uniform, or rectangular, distribution. Due to the fractalization of time, this behavior will change, and even for $m = 1$, the probability density will turn out to be unimodal.

It is also important to bear in mind that the fractalization of the time dimension has no effect on the spatial dimension. In other words, the subordinated process to Brownian motion is the Cauchy process, independent of the temporal process. The same is true when we fractalize the spatial dimension; this does not affect the temporal process in the least.

In light of the analogy of Brownian motion and the distribution in length of a polymer chain, the time interval is analogous the number of kinks n.

The length of a polymer chain is usually assumed to be given by the half-normal distribution[3]

$$f(r) = \left(\frac{2}{\pi\langle \mathbf{R}^2\rangle}\right)^{1/2} e^{-r^2/2\langle \mathbf{R}^2\rangle} \tag{26}$$

The mean-square end-to-end distance of the chain containing n kinks is given by

$$\langle \mathbf{R}^2\rangle = b^2 n^\alpha \tag{27}$$

where b is the length of a kink, and the exponent $1 \leq \alpha \leq 2$. The lower limit corresponds to a Gaussian chain, while the upper limit represents a fully extended chain. The value of $\alpha \neq 1$ is related to excluded volume effects.

Now, for Brownian motion the mean-square distance is related to the time elapsed by $\langle \mathbf{R}^2\rangle = Dt$, where D is the coefficient of diffusion. Thus, in analogy with (27) we can generalize Brownian motion to

$$\langle \mathbf{R}^2\rangle = Dt^\alpha \tag{28}$$

which for values $\alpha \neq 1$ is known as *fract(ion)al* Brownian motion (Mandelbrot and Ness, 1968). Thus, the transition probability density for Brownian motion (18) generalizes to

$$f(r|t_0) = \left(\frac{1}{\pi t_0^\alpha}\right)^{1/2} e^{-r^2/2t_0^\alpha} \tag{29}$$

where we have set $D = 1$, for simplicity. Its complementary probability density is obtained by applying the Lévy transform to (29). This results in

$$f(t|r_0) = \frac{\alpha}{(2\pi t^\alpha)^{1/2}} \frac{r_0}{t} e^{-r_0^2/2t^\alpha} \tag{30}$$

which is a generalization of the Lévy probability density, (19).

If $\mathbf{R}(t)$ is a fractal Brownian motion with transition probability density (29) and $\mathbf{T}(t)$ has the generalized Lévy probability density (30), then the subordinated process $\mathbf{T}(\mathbf{R}(t))$ will have a density that is derived from (20) and explicitly given by

$$f(t|t_0) = \frac{\alpha}{\pi}\left(\frac{t_0^\alpha}{t^{2-\alpha}}\right)^{1/2} \frac{1}{t_0^\alpha + t^\alpha}$$

This can be considered as a generalization of the inverted beta density (20), and by the same change of variable it can be converted into

$$f(\vartheta) = \frac{\alpha}{\pi} \frac{[\vartheta(1-\vartheta)]^{\alpha/2-1}}{\vartheta^\alpha + (1-\vartheta)^\alpha} \tag{31}$$

[3] See Orr (1947), or, from a pedagogical viewpoint, de Gennes (1979).

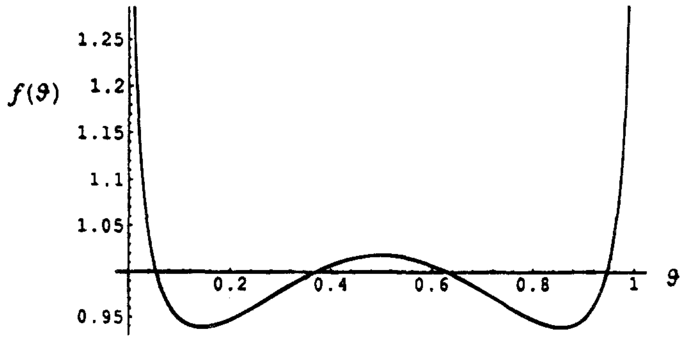

Fig. 2. The appearance of a local maximum in the probability density beyond the bifurcation point.

which is the fractal generalization of the arcsine density (21).

The fractal arcsine probability density (31) is able to cover the entire spectrum from the arcsine density where the weight is concentrated at the extremes, corresponding to $\alpha = 1$, to a unimodel probability density for $\alpha = 2$. The bifurcation between the two occurs at $\alpha = \sqrt{2}$. The shape of the probability density beyond the bifurcation point is shown in Fig. 2, while Fig. 3 shows the transition from a distribution which has most of its weight concentrated at the extremes to a unimodel distribution symmetrical about $\vartheta = 1/2$. Transforming back to the original time variable gives a half Cauchy distribution.

In the range $\sqrt{2} < \alpha < 2$, a local maximum of the midpoint coexists with the relative weights still found at the extremes. In this range, there is a *coexistence* of unimodal and arcsine-type distributions, like the coexistence of separate phases in a first-order phase transition. An entropy reduction cannot be defined for such a system since it would be concave in a restricted region about the midpoint and convex in the remainder.

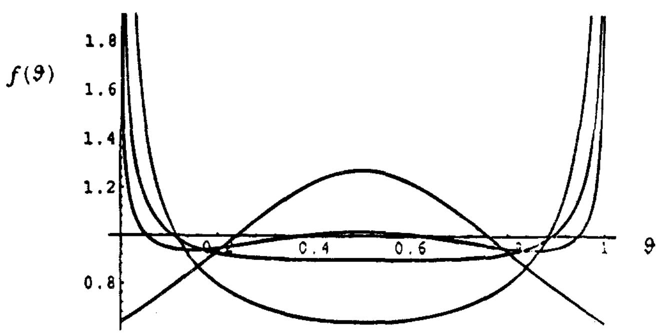

Fig. 3. The transition from the arcsine density to a unimodal probability density.

The fractalization of time permits a continuous transformation of the arcsine law, for $\alpha = 1$, into the half Cauchy distribution for $\alpha = 2$. The latter is an extreme-value distribution for largest value, whose entropy reduction is asymptotically given by (Lavenda, 1993)

$$\Delta S(t) = -e^{-(\pi/2)t/t_0} \tag{32}$$

for $t >> t_0$.[4] The process of subordination transforms the Fréchet distribution for maximum value, (30), with entropy reduction $\Delta S(t) = -r_0^2/2t^2$, into a half Cauchy distribution with entropy reduction (32). There is no relationship between the two, because the random variable and the parameter, which undergoes randomization, do not share a common distribution (Lavenda, 1995).

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

de Gennes, P.-G. (1979). *Scaling Concepts in Polymer Physics,* Cornell University Press, Ithaca, New York, Chapter 1.

Feller, W. (1968). *An Introduction to Probability Theory and Its Applications*, Vol. I, 3rd ed., Wiley, New York, p. 68.

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

Kendall, M. G., and Stuart, A. (1969). *The Advanced Theory of Statistics,* Vol. 1, 3rd ed., C. Griffin, London, §6.6.

Khinchin, A. I. (1949). *Mathematical Foundations of Statistical Mechanics,* Dover, New York, pp. 74–75.

Lavenda, B. H. (1993). *Zeitschrift für Naturforschung,* **48**, 557.

Lavenda, B. H. (1995). Subordination and Bayes' theorem in the thermodynamics of composite systems, *International Journal of Theoretical Physics,* **34**, 615.

Lévy, P. (1965). *Processus Stochastiques et Movement Brownien,* 2nd ed., Gauthier-Villars, Paris, §42.

Mandelbrot, B. B., and Ness, J. W. (1968). *SIAM Review,* **10**, 422.

Onsager, L. (1931). *Physical Review,* **38**, 2265.

Orr, W. J. (1947). *Transactions of the Faraday Society,* **43**, 12.

[4]In this limiting case, the asymptotic Cauchy distribution merges with the double exponential distribution for maximum value.