International Journal of Theoretical Physics, Vol. 31, No. 8, 1992

Thermodynamics of Rare Events

B. H. Lavenda[1] **and A. Florio**[1]

Received November 12, 1991

Scaling laws of physics are derived from extreme value distributions. Small jump processes that comprise a compound Poisson distribution generate the asymptotic distributions of stable laws. These extreme value distributions, or their tails, can be expressed in terms of the entropy decrease. As an example, the scaling law for the radius of gyration of a polymer is derived which is comparable to Flory's formula. The entropy is identified by its property of concavity, which is shown to coincide with Boltzmann's probabilistic definition for first passage in a random walk. A more general definition is required for nonintegral dimensions. The relation to mean-field theory of the kinetic Weiss–Ising model is shown and this distribution of the order parameter is governed by an asymptotic distribution for the smallest value rather than a normal distribution. Finally, the logarithm of the sample size is shown to be the yardstick for the decrease in entropy.

1. INTRODUCTION

Thermodynamics, which is dedicated to the study of systems comprised of an extremely large number of independent and identically distributed (i.i.d.) random variables, is founded on the limit laws of probability theory. The form of the prior distribution, or structure function, containing all the relevant information regarding the mechanical nature of the isolated system (Khinchin, 1949) dictates that the posterior distribution arising from thermal contact with a heat reservoir will be attracted asymptotically to the normal law as the number of degrees of freedom increases without limit.

We will be interested in systems comprised of i.i.d. random variables in which a characteristic parameter tends to infinity. In the asymptotic limit, the process will be brought into the domain of attraction of a stable law. The question of whether the system will be governed by a normal or stable law depends upon whether the variance is finite or not, respectively. Such systems are characterized by an initial distribution which is scale-invariant,

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

0020-7748/92/0800-1455$06.50/0 © 1992 Plenum Publishing Corporation

or a distribution which is said to preserve scaling. These distributions are characterized by their tail

$$1-F(u)=\Pr\{U>u\}\sim\frac{1}{u^{\alpha}} \tag{1}$$

for sufficiently large values of u, where α is the characteristic exponent. An example of such a process is a random walk in which the derivative of (1) is the transition probability density for a single jump of u steps (Barber and Ninham, 1970). Such types of inverse power laws have been considered to describe everything from word frequencies (Zipf, 1949) to the distribution of incomes and have been much publicized by Mandelbrot (1956, 1959, 1960, 1963).

In fact it would appear that there is a close relationship between our formulation and the fractal formulation in which (1) is interpreted as the number of objects greater than u, $\mathrm{Nr}(U>u)$. Then, on the basis of the analogy with (1) Mandelbrot (1983) has proposed that

$$\mathrm{Nr}(U>u)=\frac{1}{u^{\alpha}}$$

and taking the logarithm of both sides gives the fractal dimension as

$$\alpha=\frac{\ln n}{\ln(1/u)} \tag{2}$$

where n is the total number of objects or parts into which a single object has been divided. The fractal formulation is completely *deterministic* and hinges on the analogy between $\mathrm{Nr}(U>u)$, the number of events or objects greater than u, and $\Pr(U>u)$, the probability of an event being greater than u. In Section 6 we will see that the probabilistic analog of (2) is the expression for the expected largest value u_n in a population of size n,

$$n\Pr(U>u_n)=1 \tag{3}$$

Unlike (2), which determines the fractal dimension in terms of the n parts into which a self-similar object has been divided and the scaling ratio $u=n^{-1/\alpha}$, the expected largest value u_n is determined by the sample size n and the characteristic exponent α.

In economics (1) is known as the Pareto (1897) distribution, which summarizes the fact that there are more poor people than rich people. It has no mode and moments $\geq\alpha$ do not exist. All stable distributions have characteristic exponents $0<\alpha<2$ and have infinite variance. These distributions have some surprising and unexpected behavior. A simple scaling argument shows that $U\sim n^{1/\alpha}$. Considering *strictly* stable distributions for which $\alpha<1$, it is easy to see that the sample mean $(U_1+\cdots+U_n)/n$ will have the same distribution as $U_1 n^{-1+1/\alpha}$, which tends to infinity as n increases without limit. This means that the sum will be much larger than any one of its components, which implies that one term will grow exceedingly large and dominate the sum (Feller, 1971). This type of unequal sharing is best contrasted with equal sharing, leading to the law of equipartition of energy, in thermodynamics.

2. THERMODYNAMICS VS. EXTREME VALUES

In thermodynamics we are given a prior distribution, proportional to the volume of phase space occupied by the system. It increases as some fixed power m of the energy ε and has a density, or the surface area, given by

$$\Omega(\varepsilon) = \frac{\varepsilon^{m-1}}{\Gamma(m)} \tag{4}$$

$\Omega(\varepsilon)$ is referred to as the structure function (Khinchin, 1949), which contains all the relevant information concerning the mechanical structure of the isolated system prior to placing it in contact with a heat bath and other types of reservoirs. Its Laplace transform

$$\mathscr{Z}(\beta) = \int_0^\infty e^{-\beta\varepsilon}\Omega(\varepsilon)\, d\varepsilon = \beta^{-m} \tag{5}$$

is known as the partition function and β is the characteristic parameter of the heat reservoir, representing the "state of nature" of the system. The moments of the posterior density

$$f(\varepsilon; \beta) = \frac{e^{-\beta\varepsilon}}{\mathscr{Z}(\beta)}\Omega(\varepsilon) \tag{6}$$

or canonical density, all diverge in the limit as $\beta \to 0$. Hence, β is not a dummy variable as in the usual definition of the generating function, but must be given a physical meaning. With the aid of the second law, β is identified as the inverse temperature (measured in energy units), so that the expression for the average energy

$$\frac{\partial \ln \mathscr{Z}}{\partial \beta} = -\frac{m}{\beta} = -\bar{\varepsilon}$$

is none other than the law of equipartition of energy.

The fact that (4) implies (5) and vice versa is a well-known Tauberian theorem (Feller, 1971). The initial distributions of extreme values are given in terms of their tails for large values of the variate. If $F(u)$ is a probability distribution with Laplace transform $\mathscr{L}(\lambda)$, then the Laplace transform of its tail is

$$\int_0^\infty e^{-\lambda u}[1-F(u)]\,du=\frac{1}{\lambda}\int_0^\infty (1-e^{-\lambda u})\,dF(u)=\frac{1-\mathscr{L}(\lambda)}{\lambda} \tag{7}$$

where an integration by parts has been made. With the tail distribution given by (1), we obtain the analogous Tauberian result that each of the relations

$$1-\mathscr{L}(\lambda)\sim\lambda^\alpha \tag{8}$$

and

$$1-F(u)\sim\frac{1}{\Gamma(1-\alpha)}u^{-\alpha} \tag{9}$$

implies the other provided $0<\alpha<1$. Recalling that $\ln(1-x)\sim -x$ as $x\to 0$, we have $\ln\mathscr{L}\sim-(1-\mathscr{L})$ as $\lambda\to 0$, so that (8) gives

$$\mathscr{L}(\lambda)\sim e^{-\lambda^\alpha} \tag{10}$$

which implies that $\mathscr{L}(\lambda)$ is the Laplace transform of an infinitely divisible probability distribution.[2]

We have therefore related the generating function of a strictly stable distribution, having a characteristic exponent in the open interval (0, 1), to the generating function of a compound Poisson process (7). It is possible to interpret the negative of the tail $\mathscr{F}(u)=-[1-F(u)]$ as the intensity of small, positive jump processes greater than u (de Finetti, 1970). The fact that $\mathscr{F}(0)$ is infinite simply means that there are an infinite number of very tiny jump processes in each interval. If $\mathscr{F}(u)$ were not known, it could be deduced from the facts that it must give a finite contribution to the variance $\int u^2\,d\mathscr{F}(u)<\infty$ with finite limits that contain the origin, and that n times distribution of the intensity of the jumps should have the same form as $\mathscr{F}(u)$ itself (de Finetti, 1970). The resulting behavior is in direct contrast to thermodynamics, where the partition of a thermodynamic system into two large components of the same size leads to the equal sharing of the system's energy. The most probable value of the energy of either component is the mean value of the energy. However, for strictly stable distributions (i.e., $0<\alpha<1$), it appears that there is an overwhelming probability for one of the components to be much larger than the other. We will now show that this conjecture is indeed true by deriving an explicit expression for the asymptotic distribution.

[2]We recall that the family of stable laws is a subset of the set of infinitely divisible distribution laws.

The exponential Chebyshev inequality

$$\int_0^u dF(x) \leq \int_0^u e^{\lambda(u-x)}\, dF(x) \leq e^{\lambda u} \int_0^\infty e^{-\lambda x}\, dF(x) \tag{11}$$

implies that

$$\ln \Pr(U \leq u) \leq \lambda u + \ln \mathscr{L}(\lambda) \equiv S(\lambda, u)$$

The function $S(\lambda, u)$ is a saddle function which is strictly convex in λ and weakly concave in u, as shown in Figure 1. Although this is a characteristic upper bound for large deviations of normal processes (Deuschel and Stroock, 1989), we shall derive a new result showing that the equality sign holds for $S(\lambda, u) = \min$ with respect to the parameter λ, or in other words

$$\ln \Pr(U \leq u) = \{\lambda u + \ln \mathscr{L}(\lambda)\}_{\min} \tag{12}$$

Using the relation $\ln \mathscr{L} \simeq -(1 - \mathscr{L}) = -(\sigma\lambda)^\alpha/\alpha$, in the limit as $\lambda \to 0$, where σ represents the intensity of the jumps, we find

$$\hat{\lambda} = \left(\frac{\sigma^\alpha}{u}\right)^{1/(1-\alpha)} \tag{13}$$

as the stationary condition to (12).

If $\hat{\lambda}$ represents the inverse temperature and u the energy, then (13) predicts that the energy will increase as some root of the temperature rather

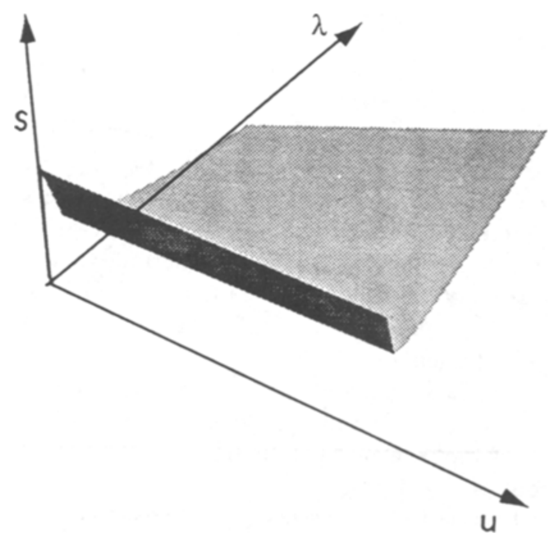

Fig. 1. The saddle function $S(\lambda, u)$ for a strictly stable distribution.

than linearly or as a power of the temperature, which is the usual thermodynamic situation. For a system with a variable particle number, such as a photon gas, the temperature, which is proportional to the average kinetic energy, increases as the fourth root of the energy. The extra energy is accounted for by particle creation (Hagedorn, 1965). On the other hand, (13) predicts that the temperature will increase as some power of the energy which could be accounted for by particle destruction.

Introducing (13) into (12), we come out with

$$\ln \Pr(U \leq u) = S(u) - S_0 \tag{14}$$

where

$$S(u) - S_0 \equiv -\frac{1-\alpha}{\alpha}\left(\frac{\sigma}{u}\right)^{\alpha/(1-\alpha)} \tag{15}$$

defines the thermodynamic entropy $S(u)$ in terms of the concave function $-u^{-\alpha/(1-\alpha)}$ and S_0 is a constant, characterizing the entropy of the unperturbed reference state, as shown in Figure 2. Expression (14) is analogous to Boltzmann's principle for extreme value distributions. Since the exponential of (14) coincides with the cumulative distribution, we obtain the probability density as

$$f(u) = \left(\frac{\sigma^{\alpha}}{u}\right)^{1/(1-\alpha)} \exp\left[-\frac{1-\alpha}{\alpha}\left(\frac{\sigma}{u}\right)^{\alpha/(1-\alpha)}\right] \tag{16}$$

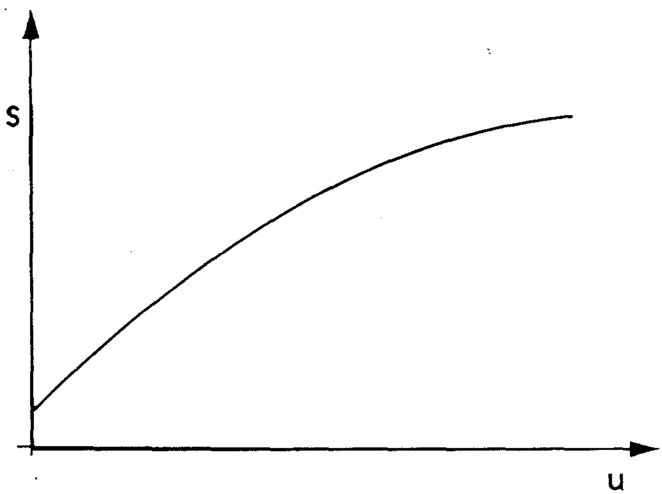

Fig. 2. Introducing the relation $\hat{\lambda} = f(u)$ obtained from the stationary criterion back into the saddle function $S(\lambda, u)$ results in a strictly concave function $S(u)$ whch is identified as the entropy.

The conventional derivation of (16) uses the statistical independence of the random variables. If $F(u)$ is the probability that a single observation will result in a value $\leq u$, then the probability that all of the n independent observations on a continuous variable are $\leq u$ is $F^n(u)$. We may interpret this as giving the probability that u is the largest value among n independent observations (Gumbel, 1954, 1958). Assuming that the tail of the initial probability distribution is of the Pareto type, (1),

$$1 - F(u) = \frac{a}{u^k}$$

with *no restrictions on the value of the characteristic exponent* k, we have

$$F^n(u) = \left\{1 - \frac{n[1 - F(u)]}{n}\right\}^n = \left(1 - \frac{an}{nu^k}\right)^n$$

for n i.i.d. random variables. Then, using the well-known limit relation $(1 - x/n)^n \to e^{-x}$ gives

$$\Pr(M_n \leq u) = F^n(u) = \exp\left[-\left(\frac{s}{u}\right)^k\right] \tag{17}$$

as $n \to \infty$, where the scaling constant is $s = (an)^{1/k}$ and $M_n = \max[U_1, \ldots, U_n]$. Comparing (14), where the entropy decrease is given by (15), to the logarithm of (17) determines $k = \alpha/(1-\alpha)$. Hence, k varies from 0 to ∞ as α goes from 0 to 1. The limit of large n is implicit in the derivation of the asymptotic distribution (14). It is to be found in our assumption that the sum of a large number of random variables has the compound Poisson distribution (de Finetti, 1929).

Just as in thermodynamics, the stable distributions are infinitely divisible, meaning that a random variable can be considered to be the sum of any number of i.i.d. components. The preservation of "additivity" means that we can partition a system into components and ask how a random variable will be partitioned among the components. Partitioning our original system into two components labeled 1 and 2 with the total "energy" $u = u_1 + u_2$, we obtain for the distribution of u_1

$$\frac{f_1(u_1) f_2(u - u_1)}{\int_0^u f_1(u_1) f_2(u - u_1)\, du_1}$$

Unlike the thermodynamic situation illustrated in Figure 3, in which the energy is divided equally for subsystems having the same number of degrees of freedom, the strictly stable distribution shows a splitting in Figure 4 above a certain threshold value of the sum of the two addenda. The

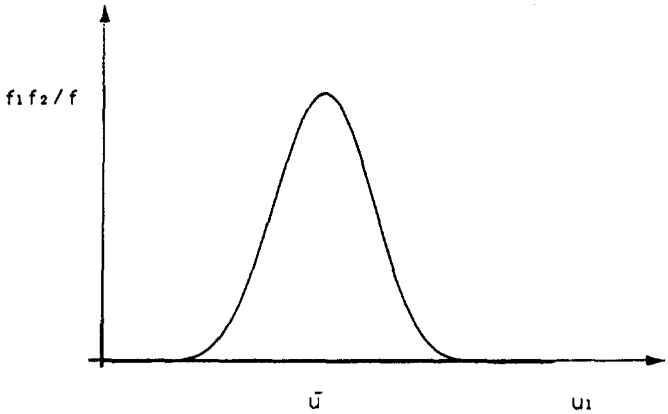

Fig. 3. The probability density of the first of two components in a thermodynamic situation where each has the same number of degrees of freedom.

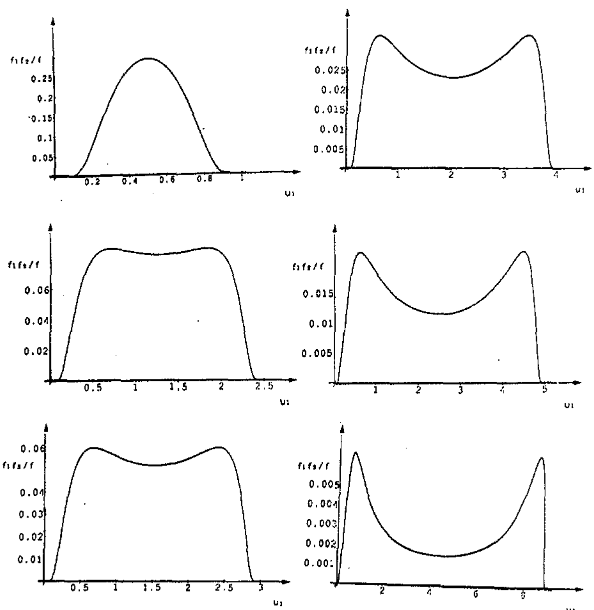

Fig. 4. The probability density of the first of two components for a strictly stable distribution. Increasing the fixed sum of the two variables leads to a splitting at approximately $u = 2.5$. The splitting accentuates as u is further increased.

characteristic exponent is $\alpha = 1/2$ and the threshold value is approximately 5/2.[3]

3. TOWARD THE NORMAL DISTRIBUTION

In the interval $1 < \alpha < 2$, the integral in the expression for the generating function of the compound Poisson process (7) fails to converge. A corrective or centering term $u\lambda$ is required. If such a term is subtracted in the integral and an equivalent term $b\lambda$ is added, the generating function will be left invariant. In analogy with Brownian motion, the b term would represent mean value. However, this interpretation will be seen to be valid only for $\alpha = 2$, while for $1 < \alpha < 2$, the b term must be interpreted as an upper limit on the variate u. Hence, in contrast to strictly stable distributions, where the reduced variate is unbound on the entire positive half-interval, we are now dealing with limited distributions whose variates can take on both positive and negative values. With the introduction of the correction term $u\lambda$, the generating function becomes

$$\begin{aligned} 1 - \mathscr{Z}(\lambda) &= b\lambda + \frac{(\alpha-1)\sigma^\alpha}{\Gamma(2-\alpha)} \int_0^\infty (1 - u\lambda - e^{-\lambda u}) \frac{1}{u^{\alpha+1}}\, du \\ &= b\lambda - (\sigma\lambda)^\alpha/\alpha = -\ln \mathscr{Z}(\lambda) \qquad \text{as} \quad \lambda \to 0 \end{aligned} \tag{18}$$

which can be easily verified by differentiating twice with respect to λ and then integrating twice. Expression (18) shows that the logarithm of the generating function is a convex function of λ, as it must be in order to ensure a positive-definite variance. For $\alpha = 2$, (18) becomes the expression for the logarithm of the generating function of the normal distribution with mean b and variance σ^2 (Cox and Miller, 1972).

From the exponential Chebyshev inequality (11), we now obtain the upper limit on the probability as

$$\ln \Pr(U \le u) \le (u - b)\lambda + (\sigma\lambda)^\alpha/\alpha = S(\lambda, u) \tag{19}$$

The entropy will have a minimum with respect to λ provided that $u \le b$, which indicates that we are dealing with limited distributions in contrast to the exponential and Cauchy distributions. The stationary condition is

$$\hat{\lambda} = \left(\frac{b-u}{\sigma^\alpha}\right)^{1/(\alpha-1)} \tag{20}$$

In the case $\alpha = 2$ there is no restriction on u; it can be either larger or smaller than b, which is precisely its mean value. We are now dealing with

[3]This splitting is not obtained for the initial density proportional to $1/u^{\alpha+1}$, contrary to what is claimed in Mandelbrot (1956).

a "compensated" sum of jumps. For $1 < \alpha < 2$ we must discriminate between positive and negative jumps. Positive jumps result in negative deviations from the limiting value b and lead to the asymptotic distribution of the largest value. Then, by symmetry, negative jumps should lead to positive deviations, where b should be the lower bound on the variate u. However, no exponential Chebyshev inequality can be derived for this process, so that we must resort to the symmetry principle which exists between the distributions of largest and smallest values (Gumbel, 1954, 1958).

Introducing (20) into (19), we obtain the thermodynamic entropy as

$$S(u) = S_0 - \left(\frac{\alpha - 1}{\alpha}\right)\left(\frac{b-u}{\sigma}\right)^{\alpha/(\alpha-1)} \tag{21}$$

which is concave in u, as it should be. According to the generalized form of Boltzmann's principle (14), the distribution function for the largest value is

$$F(u) = \exp\left[-\frac{\alpha - 1}{\alpha}\left(\frac{b-u}{\sigma}\right)^{\alpha/(\alpha-1)}\right]$$

This extreme probability distribution belongs to the third type of asymptotic distributions.[4] Therefore, by our compound Poisson process, we have now converted an initial probability distribution belonging to the Pareto type into an asymptotic one belonging to the third type.

In the case $b = 0$, the variate u is negative. The probability distribution of the smallest value can be obtained from the symmetry principle that the smallest and largest values taken from a symmetrical distribution are mutually symmetrical (Gumbel, 1954, 1958). The probability density of the smallest value is obtained from the largest one by changing the sign of the variate. Since the variate is negative for the largest value, we arrive at a known paradoxical result that largest value is negative while the smallest one is positive (Gumbel, 1954, 1958). This confusion can be avoided by focusing our attention on the probability of the smallest value to exceed u:

$$\Pr(U > u) = \exp\left[-\frac{\alpha - 1}{\alpha}\left(\frac{u}{\sigma}\right)^{\alpha/(\alpha-1)}\right] \tag{22}$$

Its density,

$$f(u) = \left(\frac{u}{\sigma^{\alpha}}\right)^{1/(\alpha-1)} \exp\left[-\frac{\alpha - 1}{\alpha}\left(\frac{u}{\sigma}\right)^{\alpha/(\alpha-1)}\right] \tag{23}$$

bears the name of Weibull, after the Swedish engineer who first used the distribution in the analysis of the breaking strength of materials.

[4]Due to deep-rooted tradition, we will continue to refer to the classification scheme according to which the extreme value distribution of the first type is the exponential distribution.

The so-called age-specific failure rate is defined as the ratio of (23) to (22) (Cox, 1962). In particular, the age-specific failure rate is a constant for the exponential distribution, implying that the failure of a certain component is independent of its age. For $1<\alpha<2$, the age-specific failure rate is a positive, increasing function of u, which indicates positive ageing.

Due to the symmetry between the largest and smallest values (i.e., the negative of the former is equal to the latter), the probability density (23) will have an entropy

$$S(u)=S_0-\frac{\alpha-1}{\alpha}\left(\frac{u}{\sigma}\right)^{\alpha/(\alpha-1)} \tag{24}$$

This can also be deduced from the fact that the third asymptotic distribution of the smallest value can be derived from the second asymptotic distribution of the largest value by a reciprocal transformation (Gumbel, 1954, 1958) (see Section 4). In other words, the invariance of the entropy is a direct consequence of the symmetry principle relating largest and smallest value distributions.

For $0<\alpha<1$ the integral of the generating function of the compound Poisson process converges without the centering term $u\lambda$, while in the range $1<\alpha<2$ it is absolutely necessary. This is a reflection of the fact that in the first case, the compound Poisson process consists of only positive jumps, which results in an extreme value distribution of the largest value for a nonnegative variate, whereas in the second case both positive and negative jumps occur. The positive jumps generate an asymptotic probability distribution of the largest value for a variate which is negative in the case $b=0$. Consequently, we are led to attribute the negative jump processes to the existence of an asymptotic probability distribution for the smallest value for a positive variate. Undoubtedly, it is this phenomenon which is at the root of the symmetry principle between the extreme distributions of largest and smallest values.

4. SCALING LAWS FROM EXTREME-VALUE DISTRIBUTIONS

For $\alpha=3/2$, the Weibull density (23) is also known as the law of the distribution of the nearest neighbor in a random distribution of particles (Chandrasekhar, 1943). Introducing the modulus of the force of gravitation $|\mathbf{F}|\propto 1/r^2$, where r represents distance, leads immediately to the well-known Holtsmark distribution. This distribution, discovered by the Danish astronomer, appeared several years before Lévy's (1925) monograph introducing the notion of a stable law.

Feller (1971) conjectured that the four-dimensional analogue to the Holtsmark distribution would be a symmetrically stable distribution with characteristic exponent $\alpha = 4/3$, since in four dimensions the gravitational force varies inversely as the third power of the distance. We may generalize this to D dimensions where the characteristic exponent $\alpha = D/(D-1)$ and the force varies inversely as the $(D-1)$ power of the distance.

The Weibull density can be derived from a generalized normal process,

$$dF(x_1, x_2, \ldots, x_D) \propto \exp\left\{ -\frac{1}{D}\left(\frac{\sum_{i=1}^{[D]} x_i^2}{\sigma^2}\right)^{D/2}\right\} dx_1\, dx_2 \cdots dx_D \quad (25)$$

where $D = \alpha/(\alpha-1)$ and $[D]$ stands for the largest integer $\leq D$. Introducing the square of the distance in Euclidean space,

$$r^2 = \sum_{i=1}^{D} x_i^2$$

into (25) and noting that the Jacobian of the transformation $dx_1\, dx_2 \cdots dx_D \propto r^{D-1}\, dr$, we see that it transforms into (23).

Consider a polymer chain in D dimensions. The probability that the radius r lies within the interval dr is

$$f(r) = \frac{r^{D-1}}{r_0^D} \exp\left[-\frac{1}{D}\left(\frac{r}{r_0}\right)^D \right] \quad (26)$$

where r_0 is the unperturbed radius of the chain. If there are n monomers, each separated by a distance a, then $r_0 = n^{1/2} a$. The total repulsive energy in D dimensions is (de Gennes, 1979)

$$E = \frac{T\sigma n^2}{r^D} \quad (27)$$

where σ is the excluded volume. Introducing (27) into the radial distribution (26) converts a Weibull density for the smallest value into an asymptotic distribution of the second type for the largest value of E, viz.

$$f(E) = \frac{1}{D} \frac{E_{\max}}{E^2} \exp\left(-\frac{1}{D} \frac{E_{\max}}{E} \right) \quad (28)$$

with characteristic exponent $\alpha = 1/2$, where the maximum repulsive energy is

$$E_{\max} = \frac{T\sigma}{a^D} n^{2-D/2}$$

According to (15), the entropy is

$$S(E) = S_0 - \frac{1}{D} \frac{E_{\max}}{E} \quad (29)$$

The second law then defines the temperature as

$$S'(E) = \frac{E_{\max}}{DE^2} = T^{-1} \tag{30}$$

which upon rearrangement gives

$$E = \left(\frac{\sigma}{Da^D}\, n^{2-D/2}\right)^{1/2} T$$

Since this must be the same as the expression for the repulsive energy (27), we obtain the expression

$$r_G^D = (D\sigma a^D)^{1/2} n^{1+D/4} \tag{31}$$

for the radius of gyration r_G. Therefore, the radius scales as $r_G \sim n^\nu$ with the characteristic exponent $\nu = (4+D)/4D$.

In a dilute solution of separate coils in a good solvent, Flory has shown, by a mean-field argument, that the radius of gyration is related to the degree of polymerization according to $r_G \sim n^{\nu_F}$, where $\nu_F = 3/(D+2)$. Flory derived this formula by minimizing the total energy with respect to the radius, where the total energy is the sum of the total repulsive energy (27) and an attractive term representing the elastic energy which is proportional to the square of the radius (de Gennes, 1979). This variational condition is equivalent to the second law, (30), which implicitly takes into account the dynamical balance through the definition of the temperature.

A comparison between the two characteristic exponents is given in Table I. For dimensions $D = 2$ and $D = 4$ the two characteristic exponents are identical, while for $D = 3$, the difference is well within experimental error. For dimensions $D > 4$, ν decreases more slowly than ν_F.

Alternatively, from the probability density (26), we obtain the entropy decrease due to elongation as

$$S(r) = S_0 - \frac{1}{D}\left(\frac{r}{r_0}\right)^D \tag{32}$$

Table I. Comparison of Characteristic Exponents

Dimension D	$\nu_F = 3/(D+2)$	$\nu = (4+D)/D$	$\lvert(\nu_F - \nu)/\nu_F\rvert$
2	3/4	3/4	0
3	3/5	7/12	2.78%
4	1/2	1/2	0
5	3/7	9/20	5%
6	3/8	5/12	11.11%

Applying the second law gives

$$-S'(r) = \frac{r^{D-1}}{r_0^D} = \frac{\chi}{T}$$

where χ is the force conjugate to r. But since $\chi = -E'$, where E is the repulsive energy given by (27), we immediately obtain (31) and consequently $\nu = (4+D)/4D$.

5. COMBINATORIAL DEFINITION OF ENTROPY

So far our only justification for calling $S(r)$ an entropy lies in its property of concavity. Although this is the defining property of the entropy (Lavenda and Dunning-Davies, 1990), it is of interest to relate it to probability in a way in which Boltzmann did; that is, to the logarithm of the number of ways of selecting a given number of objects from a parent population. Since this cannot be done in general, we must generalize the concept of entropy in a way which liberates it from being defined as the logarithm of a combinatorial factor while retaining its relation to probability.

Consider a random walk along integer points on the real axis. The system is initially placed at the origin and it is equally probable for there to be a transition to the right or to the left. The probability of being at r after n steps is

$$p(r) = \binom{n}{(n+r)/2} \frac{1}{2^n}$$

provided r and n have the same parity; otherwise $p(r) = 0$. In the limit of large n this tends to the normal distribution

$$p(r) = \frac{1}{(n\pi/2)^{1/2}} \exp\left(-\frac{r^2}{2n}\right)^{1/2} \tag{33}$$

provided r is of the order of $\sqrt{n}$.

If, however, we ask for the probability that a point $r > 0$ is reached for the *first* time after n steps, we have (Feller, 1968)

$$f(r) = \frac{r}{n} p(r) = \frac{r}{n} \binom{n}{(n+r)/2} \frac{1}{2^n} \tag{34}$$

This distribution does not tend to the normal density (33) in the limit of large n, but, rather, to the Rayleigh density, which is an asymptotic distribution for the smallest value of r.

According to Boltzmann's principle, the entropy is defined as (Lavenda, 1991)

$$S(r) = \ln\binom{n}{(n+r)/2}$$

For n sufficiently large, we may use Stirling's approximation to obtain

$$S(r) = n \ln n - \frac{n+r}{2}\ln\left(\frac{n+r}{2}\right) - \frac{n-r}{2}\ln\left(\frac{n-r}{2}\right) \tag{35}$$

showing that the ratio $r/n \equiv \xi$ plays the role of an order parameter. Expression (35) is the entropy of mixing and in terms of it we may write the distribution (34) as

$$\ln f(r) = \ln(r/n) + S(r) - S_0 \tag{36}$$

where

$$S_0 = n \ln 2$$

is the maximum value of the entropy. However, on account of the first term in (36), it will only be a proper probability density in the limit $r/n \ll 1$ where the entropy is approximated by the quadratic expression

$$S(r) = S_0 - \frac{r^2}{2n}$$

The resulting probability density bears the name of Rayleigh (Lord Rayleigh, 1880, 1919), who first derived it in connection with the problem of random vibrations in two dimensions. The now continuous density (36) has a tail distribution given by

$$1 - F(r) = \exp[S(r) - S_0] \tag{37}$$

as can easily be seen by differentiation.

The Rayleigh density is the stationary solution to the Fokker-Planck equation (Stratonovich, 1963)

$$\frac{1}{\mathscr{D}}\frac{\partial f}{\partial t} = -\frac{\partial}{\partial r}\left[\left(\frac{1}{r} - \frac{r}{n}\right)f - \frac{\partial f}{\partial r}\right]$$

in the case of a vanishing probability current, where $\mathscr{D}$ is the diffusion coefficient. The corresponding Langevin equation is

$$\dot{r} = -\mathscr{D}\left(\frac{r}{n} - \frac{1}{r}\right) + (2\mathscr{D})^{1/2}\dot{w} \tag{38}$$

where w is a Wiener process. This term describes the influence of random fluctuations. In its absence the stochastic differential equation (38) reduces to a deterministic rate equation having a nontrivial stationary solution given by $r_0 = \sqrt{n}$, or in terms of the order parameter,

$$\xi_0 = \frac{1}{\sqrt{n}} \tag{39}$$

If ξ is to retain its interpretation as an order parameter, we must explain the reason for there being the nonvanishing stationary value (39).

Returning to the expression for the entropy of mixing (35), we have

$$-S'(\xi) = \frac{1}{2}\ln\left(\frac{1+\xi}{1-\xi}\right) = \frac{\chi}{T} \tag{40}$$

on the strength of the second law. Solving for ξ, we get

$$\xi = \tanh\left(\frac{\chi}{T}\right) \tag{41}$$

This equation has the same form as in the mean-field approximation to the Weiss-Ising model of a ferromagnetic (Griffiths *et al.*, 1966). Suppose that the force χ is given by the linear phenomenological relation

$$\chi = T_c \xi \tag{42}$$

which would correspond to the zero magnetic field situation in the mean-field approximation, where T_c is the critical temperature. It is now evident that (41) possesses nonzero solutions only for temperatures $T < T_c$. For small values of the order parameter, the hyperbolic tangent may be approximated by the first two terms in its Taylor series expansion. We thus obtain

$$\xi_0 = T\left[\frac{3(T_c - T)}{T_c^3}\right]^{1/2} \tag{43}$$

which is only possible provided $T < T_c$; otherwise, we have the trivial solution $\xi_0 = 0$. Now, (43) must be consistent with (39) and this fixes the number of steps as

$$n = \frac{T_c^3}{3T^2(T_c - T)} \tag{44}$$

which becomes infinite as $T \uparrow T_c$. This shows that the order parameter tends to zero as the square root of the difference in the temperature, while the number of random steps, which is proportional to the time elapsed, tends to infinity as the inverse of the temperature difference.

Although the interpretation of the mean-field approximation to a lattice spin system given by (41) is well known, it is generally assumed that the probability density centered about the metastable state has the normal form (33). On the contrary, we have shown that the pertinent probability density is the extreme value density:

$$f(r)=\frac{r}{n}\exp\left(-\frac{r^2}{2n}\right) \tag{45}$$

for the *smallest* value of the order parameter.

6. THERMODYNAMICS OF EXTREME VALUES

One of the two basic parameters which characterize extreme value distributions is the return period τ (Gumbel, 1954, 1958). It is defined as

$$\tau(r)=\frac{1}{1-F(r)} \tag{46}$$

In flood control $\tau(r)$ would represent the mean time interval between two discharges of a river.

According to (14), the second asymptotic distribution for the largest value is

$$F(r)=\exp[S(r)-S_0]=\exp\left[-\left(\frac{\vartheta}{r}\right)^k\right] \qquad \mathrm{II} \tag{47}$$

while from (22), the tail of the third asymptotic distribution for the smallest value is

$$1-F(r)=\exp[S(r)-S_0]=\exp\left[-\left(\frac{r}{\vartheta}\right)^k\right] \qquad \mathrm{III} \tag{48}$$

In (47), $\vartheta=\sigma((1-\alpha)/\alpha)^{1/k}$ and $k=\alpha/(1-\alpha)$, while in (48), $\vartheta=\sigma((\alpha-1)/\alpha)^{1/k}$ and $k=\alpha/(\alpha-1)$. Note that whereas the characteristic exponent k can be greater or less than one in the first case, k must be necessarily greater than one in the second case.

In Section 1 we have defined the expected largest value r_n according to (3), or equivalently as

$$n[1-F(r_n)]=1$$

Alternatively, the expected smallest value r_1 is defined by (Gumbel, 1954, 1958)

$$nF(r_1)=1$$

The definition of the expected largest value is just another way of writing the return period (46). The expected largest value of the third distribution is

$$S_0 - S(r_n) = \ln n \qquad \text{III}$$

or

$$r_n = \vartheta(\ln n)^{1/k} \qquad \text{III}$$

Since $k > 1$, the expected largest value increases more slowly than $\ln n$. Likewise, the entropy change due to the smallest value of the second asymptotic distribution increases as the sample size increases:

$$S_0 - S(r_1) = \ln n \qquad \text{II}$$

or

$$r_1 = \frac{\vartheta}{(\ln n)^{1/k}} \qquad \text{II}$$

In both cases, the extreme values increase with increasing sample size and the entropy change increases as the logarithm of the sample size, which can also be interpreted as the amount of time elapsed. Therefore, the logarithm of the sample size is the yardstick for determining the entropy change. This behavior is analogous to our previous interpretation of the expected time of exit from the domain of attraction of a stable stationary state, where the logarithm of the expected exit time is equal to the entropy difference between the stationary state and the state with maximum entropy on the boundary (Lavenda, 1985). In extreme value theory, the corresponding relation between the logarithm of the return period and the entropy change is

$$\ln \tau(r^*) = S_0 - S(r^*) \tag{49}$$

where $r^* = r_n$, the largest expected value of the third asymptotic distribution, while $r^* = r_1$, the smallest expected value of the second asymptotic distribution.

The second basic parameter which characterizes extreme value distributions is the intensity $\mu(r)$ (Gumbel, 1954, 1958). In mortality statistics,

$$\mu(r)\, dr = \frac{f(r)}{1 - F(r)}\, dr \tag{50}$$

is the probability that a person aged r will die in the interval between r and $r + dr$. The distribution $F(r)$ has density $f(r)$ and $\mu(r)$ is referred to as the "force of mortality," or simply as the intensity function. The intensity μ elicits a thermodynamic interpretation for the third asymptotic distribution for the smallest value. Differentiating the tail (48) of the third asymptotic distribution, we obtain the probability density as

$$f(r)=\frac{\chi(r)}{T}\exp[S(r)-S_0]$$

Then, according to the definition of the intensity (50), we find

$$\mu(r)=\frac{\chi(r)}{T}$$

The intensity is the thermodynamic force of the third asymptotic distribution for the smallest value, which is necessarily a positive quantity. This condition was used in the last section to ensure the existence of a nontrivial solution to (41). In the case of the second asymptotic distribution, the thermodynamic force is negative and hence cannot be related to the intensity.

7. PERSPECTIVES

We have shown that the entropy expressions that are related to asymptotic distributions for extreme values are generalizations of Boltzmann's combinatorial definition of the entropy, which for small fluctuations corresponds to the case $D=2$. In general, we can expect nonintegral values of D that cannot be derived from any combinatorial formula by using Stirling's approximation and developing the logarithmic expressions in a Taylor series. We have therefore relied on the characterizing property of concavity to identify the entropy (Lavenda and Dunning-Davies, 1990).

Although there are numerous examples of extreme value distributions in the social sciences, relatively few have been identified in the physical sciences thus far. A classic exception is the Cauchy distribution for the energy distribution of unstable states in the reaction of decay products, commonly known as the Lorentz distribution (Zolotarev, 1986). We have shown that the order parameter in an order–disorder transition is governed by an asymptotic distibution of the third type for the smallest value. Extreme value distributions should also find a role in critical phenomena in which a transition point can be modified by stochastic processes such as the buildup of nucleation centers, structural flaws, etc.

Finally, it is worth noting that since the extreme value distributions are expressed in terms of the entropy decrease alone, they can only account for the statistics of the process. External interactions that modify the energy of the system can be taken into account by the method of conjugate distributions (Lavenda, 1991), which transforms, for example, an isolated system into a closed system by placing it in contact with a heat reservoir.

ACKNOWLEDGMENTS

The work of B.H.L. was supported in part by the Consiglio Nazionale delle Ricerche and the Ministero dell'Università e della Ricerca Scientifica e Tecnologica.

REFERENCES

Barber, M. N., and Ninham, B. W. (1970). *Random and Restricted Random Walks*, Gordon & Breach, New York.

Chandrasekhar, S. (1943). *Reviews of Modern Physics*, **15**, 1.

Cox, D. R. (1962). *Renewal Theory*, Methuen, London.

Cox, D. R., and Miller, H. D. (1972). *Theory of Stochastic Processes*, Chapman & Hall, London.

De Finetti, B. (1929). *Atti Accademia Nazionale dei Lincei Rendiconti Classe di Scienze Fisiche Matematiche e Naturali*, **10**, 163.

De Finetti, B. (1970). *Teoria delle Probabilità*, Einaudi, Turin.

De Gennes, P.-G. (1979). *Scaling Concepts in Polymer Physics*, Cornell University Press, Ithaca, New York, p. 44.

Deuschel, J.-D., and Stroock, D. W. (1989). *Large Deviations*, Academic Press, Boston.

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

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

Griffiths, R. B., Weng, C.-Y., and Langer, J. S. (1966). *Physical Review*, **149**, 301.

Gumbel, E. J. (1954). *Statistical Theory of Extreme Values and Some Practical Applications*, National Bureau of Standards, Washington, D.C.

Gumbel, E. J. (1958). *Statistics of Extremes*, Columbia University Press, New York.

Hagedorn, R. (1965). *Nuovo Cimento Supplemento*, **3**, 147-186.

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

Lavenda, B. H. (1985). *Nonequilibrium Statistical Thermodynamics*, Wiley-Interscience, New York.

Lavenda, B. H. (1991). *Statistical Physics: A Probabilistic Approach*, Wiley-Interscience, New York.

Lavenda, B. H., and Dunning-Davies, J. (1990). *Foundations of Physics Letters*, **3**, 435-441.

Lévy, P. (1925). *Calcul des Probabilités* (Gauthier-Villars, Paris).

Lord Rayleigh (1880). *Philosophical Magazine*, **10**, 73.

Lord Rayleigh (1919). *Philosophical Magazine*, **37**, 321.

Mandelbrot, B. (1956). In *Information Theory, the Third London Symposium*, C. Cherry, ed., Academic Press, New York.

Mandelbrot, B. (1956). *Comptes Rendus des Séances Hebdomadaires de l'Académie des Sciences de Paris*, **242**, 2223-2226.

Mandelbrot, B. (1959). *Comptes Rendus des Séances Hebdomadaires de l'Académie des Sciences de Paris*, **249**, 613-615, 2153-2155.

Mandelbrot, B. (1963*a*). *International Economic Review*, **1**, 79-106.

Mandelbrot, B. (1963*b*). *International Economic Review*, **4**, 517-543.

Mandelbrot, B. (1983). *The Fractal Geometry of Nature*, Freeman, New York.
Pareto, V. (1897). *Cours d'économie politique*, Rouge & Co., Lausanne.
Stratonovich, R. (1963). *Topics in the Theory of Random Noise*, Vol. I, Gordon & Breach, New York.
Zipf, G. K. (1949). *Human Behavior and the Principle of Least Effort*, Addison-Wesley, Cambridge, Massachusetts.
Zolotarev, V. M. (1986). *One-Dimensional Stable Distributions*, American Mathematical Society, Providence, Rhode Island.