International Journal of Theoretical Physics, Vol. 35, No. 9, 1996

Fractals, Clusters, and Order-Statistics: A New Relation Between Probability and Entropy

B. H. Lavenda[1]

Received October 22, 1995

The relationship between fractals and clusters and order-statistics is developed. The initial distributions for order-statistics coincide with homogeneous power laws, used in fractal geometry and clustering to generate self-similar objects. The entropy reduction is defined in terms of the number of particles or events from the top or the bottom of the ordered set. Expressions for the joint and conditional distributions of order-statistics are given in terms of the entropy differences of the interval. Statistical equivalence principles are given in which the probability of the entropy reduction being less (or greater) than the index of the order-statistic is the same as the probability of that order-statistic being greater (or less) than a given value.

1. POWER LAWS

Homogeneous power laws have been used to relate "rank" to frequency, or probability (Zipf, 1949), and the volume of a hypersphere, or probability, to some power of the radius of the volume (Khinchin, 1949). Examples of the former are the Pareto law of incomes and the Zipf law relating frequency to word ranks, while those of the latter are the mass-to-radius relation and the volume of phase space as a function of the energy. The former and latter are inverse-power and power relations, respectively. The exponent may be nonintegral, leading in the former case to the question of how we are to measure distance.

Regarding inverse-power laws, if L were a smooth curve, it could be determined in the limit where the number of steps $N(l)$ tends to infinity while

[1] Centre for Thermodynamics, ENEA CRE Casaccia, 00060 SM Galeria, Rome, Italy, and Università di Camerino, Camerino 62032 (MC), Italy.

0020-7748/96/0900-1993$09.50/0 © 1996 Plenum Publishing Corporation

the size of the individual steps l tends to zero such that their product

$$\lim_{l \to 0} N(l)l = L$$

is finite. However, for real-life curves, which are often continuous but not differentiable, L does not exist. This motivated Mandelbrot (1982) to define the *fractal* dimension.

As concerns power laws, the probability is measured in terms of the volume of phase space that is occupied by the system. Instead of the number, or probability, decreasing as some inverse power, it now increases as a fixed power. The fact the fixed power turns out to be nonintegral, as in cosmological clustering of galaxies (Lavenda, 1995), motivates the definition of the *cluster* dimension (Hastings and Sugihara, 1993).

But, surely, in order to provide a valid description, these power laws cannot be valid everywhere. Inverse-power laws, like the Pareto law, describe events above a given value, such as those of incomes, while power laws themselves must have some upper limit to them, as in the case of the microcanonical ensemble in statistical mechanics, where the total energy of the system is held constant.

The fractal and cluster nature of these cutoffs is evident, as in the case of trees whose smallest branches have a certain diameter. On smaller scales the fractal properties that are seen on larger scales are lost. Alternatively, when trees are viewed at sufficient distances, the individual nature is lost and the entire collection blends into a forest, whose cluster properties do not mimic those of individual trees.

However, on a scale where self-similarity applies it has been well known since the time of Leonardo da Vinci that the combined cross-sectional areas of two main branches must be equal to that of the trunk. If d_1 and d_2 are the diameters of the branches and d is the diameter of the trunk, then Leonardo's claim was $d^2 = d_1^2 + d_2^2$. The same relationship holds for the standard deviation of the normal distribution, $\sigma = (\sigma_1^2 + \sigma_2^2)^{1/2}$. However, for extreme-value distributions, the standard deviation does not exist and we must replace it by a scale factor $a = (a_1^\Delta + a_2^\Delta)^{1/\Delta}$ (De Finetti, 1970). Such a generalized relation has been proposed by Mandelbrot (1982), who called Δ the *diameter exponent*. And if generalized botanical trees have anything to do with extreme-value distributions, the range of permitted values of Δ is $(0, 2]$, where $\Delta = 2$ corresponds to the normal distribution.

These same power laws appear as initial, or sample, distributions in extreme-value theory (Gumbel, 1954, 1958). Since there are only three types of asymptotic distributions, this may provide some restrictions on the otherwise seeming arbitrariness of the power law distributions.

There is the interesting question of what is the relationship between fractal and clustering phenomena and order-statistics and extreme-value distributions? In thermodynamics we know how to define entropy in terms of a pseudoprobability.[2] In the case of the microcanonical ensemble, entropy is defined as the logarithm of the "thermodynamic probability," or the volume of phase space occupied by the system. However, this is an *asymptotic* relation in the sense that the population, or the number of degrees of freedom, has been allowed to increase without limit. It is in this context that Boltzmann's principle is valid [see equations (24) and (25) below]. Can entropy be determined without going to the "thermodynamic" limit? And if so, how is entropy related to the initial probability distributions, which turn out to be fractal and cluster power laws? The aim of this paper is to provide the answers to these questions.

2. FRACTALS AND CLUSTERS: THINNING OUT AND FILLING IN

The definition of dimensions for nonintegral objects has been widely publicized by Mandelbrot (1982). Although it is not often made clear, there are at least two types of objects that possess nonintegral dimensions. Well-known examples are the Korčak and cluster exponents. In 1938 Korčak studied the distribution of sizes of islands in the Aegean sea. Let $\mathrm{Nr}(A > a)$ be the number of islands greater than an area a. Korčak found that this number satisfies the scale-invariant relation

$$\mathrm{Nr}(A > a) = \text{const} \times a^{-B} \tag{1}$$

Mandelbrot (1982) drew attention to the fact that this power law is analogous to the definition of the Hausdorff dimension $\mathfrak{D}$ of a subset X of Euclidean space as

$$\mathfrak{D} = \lim_{r\to 0} \frac{\log N(r)}{\log(1/r)} \tag{2}$$

where $N(r)$ represents the smallest number of open balls of radius r needed to cover X. In the Kračak formula (1), this minimal number is assumed to be analogous to $\mathrm{Nr}(A > a)$, or the number of islands having an area greater than a. The Korčak exponent can thus be defined rigorously as the limit

$$B = \lim_{a\to 0} \frac{\log \mathrm{Nr}(A > a)}{\log(1/a)} \tag{3}$$

[2]This may appear to make arbitrary the definition of the entropy in terms of a statistical quantity. Actually, the definition is fixed by the fact that the probability density can be expressed in terms of Gauss' law of error whose potential is precisely the entropy (Lavenda, 1991).

which also has an intuitive appeal, and to which we will refer as the *fractal* exponent.

If n represents the total number of islands whose areas are governed by the probability distribution $F(a)$, then

$$\operatorname{Nr}(A > a) = n \Pr(A > a) = n\{1 - F(a)\} = n\left(\frac{a_0}{a}\right)^B \tag{4}$$

where a_0 represents some lower cutoff on the area of the islands for which the power law (1) is valid. According to this relation between the number of islands having an area greater than a and the tail of the probability distribution, the power law (1) will be appreciated as the Pareto distribution. The Pareto tail distribution is the initial distribution which is in the domain of attraction of the Fréchet distribution for the largest value (Gumbel, 1954, 1958).

For clustering phenomena we want to know the number of particles, say, within a given distance $\operatorname{Nr}(R \leq r)$ from a centrally located one. If this number of particles scales as

$$\operatorname{Nr}(R \leq r) = \text{const} \times r^D \tag{5}$$

then the cluster is said to have dimension D (Hastings and Sugihara, 1993). Hence, we can define rigorously the *cluster* exponent as the limit

$$D = \lim_{r \to \infty} \frac{\log \operatorname{Nr}(R \leq r)}{\log r} \tag{6}$$

in analogy with the fractal exponent (3). If n represents the total number of objects, then the number of objects within a radius r of the origin is

$$\operatorname{Nr}(R \leq r) = n \Pr(R \leq r) = nF(r) = n\left(\frac{r}{r_0}\right)^D \tag{7}$$

where r_0 is some upper cutoff, which may be the radius of the entire volume under consideration. In this light it will be appreciated that (7) is the initial distribution which is the domain of attraction of the Weibull distribution for the smallest value (Gumbel, 1954, 1958).

Since there are only three stable distributions for the largest values, one of which we can rule out since it corresponds to a negative variate, there remains one other stable distribution that has not been exploited thus far. This is the double-exponential, or Gumbel, distribution (Gumbel, 1954, 1958), whose initial tail distribution is the exponential function. However, the definition of the number of events exceedings v as

$$\mathrm{Nr}(V > v) = n \exp(-\alpha v) \tag{8}$$

where α is known as the *hazard rate*, is not scale-invariant in the sense that

$$f(x) = nf(a_n x) \tag{9}$$

where the parameter a_n depends on n. In other words, a graph which appears the same on all scales must be of the form x^C, where $a_n = n^{-1/C}$ is the scale factor. Surprisingly enough, this is also the condition for *stability* of the so-called second and third asymptotic distributions (Gumbel, 1954, 1958).

The *stability postulate* asserts that if the distribution of an extreme is equal to the probability distribution except for a linear transform of the variate, then the initial distribution is said to be stable with respect to this extreme (Gumbel, 1954, 1958). Since a linear transform does not change the form of the distribution, the probability that the largest value of the variate is smaller than or equal to x must be equal to the probability distribution of a linear function of x, namely

$$F^n(x) = F(a_n x - b_n) \tag{10}$$

where a_n and b_n are referred to as the scaling and centering constants, respectively. The second and third asymptotic distributions of the largest values, where $x < 0$ for the third distribution, are derived from the stability postulate (10) by setting $b_n = 0$. Upon taking the logarithms of both sides of (10), it is not difficult to recognize that this is the same as the criterion of scale invariance (9) (De Finetti, 1970). Invoking the *symmetry principle*, whereby the probability distribution for the smallest value is obtained from that of the largest value by changing the sign of the variate, what pertains to the third asymptotic distribution for largest value is also valid for the distribution of the smallest value, known as the Weibull distribution. In other words, *the definitions of the fractal and cluster exponents (3) and (6) are none other than the conditions for the existence of the Fréchet and Weibull distributions,* which are extreme distributions for the largest and smallest values, respectively.

As we have mentioned, the only remaining extreme-value distribution, the double-exponential or the Gumbel distribution, does not conform to the same type of scale-invariant criterion (9). Rather, the stability postulate requires a scaling parameter $a_n = 1$. This has the effect of shifting the initial probability distribution to the right by an amount

$$b_n = \alpha^{-1} \log n \tag{11}$$

where α^{-1} is the scale parameter in the exponential distribution [see equation (8)] without a change of shape. Since (11) does not concern scale invariance, we will limit our discussion to the asymptotic stable distributions associated with exponents (3) and (6). In other words, these scale-invariant conditions coincide with the stability postulate for the Fréchet distribution for the largest value and the Weibull distribution for the smallest value. They involve the shape parameters, or the exponents, B and D, whereas (11) involves the centering parameter with constant scaling, α^{-1}.

3. ORDER-STATISTICS

We now investigate the relationship between fractals, clusters, and order-statistics. Apart from its own interest, this will also provide a justification for the proposed relation between the probability of falling within a given interval and the entropy difference of the endpoints of the interval.

If we take a sample of independent and identically distributed random values $y_1, y_2, \ldots, y_k$ and order them such that $y_{(1)} \leq y_{(2)} \cdots \leq y_{(k)}$, the property of statistical independence and the fact that they share a common sample distribution will no longer apply. Nevertheless, some simple and remarkable results can be obtained concerning their sampling, joint, and conditional distributions (Kendall and Stuart, 1969).

The probability distribution for the rth order-statistic $Y_{(r)}$ can be obtained by supposing that the population of size n has a continuous distribution function $F(y)$ with a density $F'(y) = f(y)$, the prime being used to denote the derivative. Let y denote the rth value from the *bottom*. The probability of $Y_{(r)}$ is derived by considering each of the n independent measurements as a Bernoulli trial, either success or failure. A "success" means $Y_i \leq y$, while a "failure" occurs when $Y_i > y$. Thus, the probability that $r - 1$ of the Y_i are less than y while $n - r$ are greater than y with the remaining value falling between y and $y + dy$ is

$$g_r(y)\, dy = n! \frac{[F(y)]^{r-1}}{(r-1)!} \frac{[1 - F(y)]^{n-r}}{(n-r)!} f(y)\, dy \tag{12}$$

This is a beta distribution in F and the factorials can be combined into the Beta function,

$$B(r, n - j + 1) = \Gamma(r)\Gamma(n - r + 1)/\Gamma(n + 1)$$

If x denotes the rth value from the *top*, the beta distribution is

$$dG_r(F(x)) = \frac{[F(x)]^{n-r}[1 - F(x)]^{r-1} dF(x)}{B(r, n - r + 1)} \tag{13}$$

The beta distributions are special cases of the multinomial distribution in which n elements are taken at random and we want to determine into which of the k intervals they belong.[3] On the assumptions of independence and replacement, the number of elements in each interval is a multinomial random variable [Castillo (1988), equation (2.11)], and, consequently, the joint probability distribution (12) from the bottom for k intervals is given by

$$g_{r_1,\ldots,r_k}(y_1, y_2, \ldots, y_k)\, dy_1\, dy_2 \cdots dy_k = n! \prod_{i=1}^{k} \frac{[F(y_i) - F(y_{i-1})]^{r_i - r_{i-1} - 1}}{(r_i - r_{i-1} - 1)!} f(y_i)\, dy_i \tag{14}$$

with $y_1 \leq y_2 \cdots \leq y_k$. The fact that the distribution function $F(y)$ is completely monotone ensures that the probability distribution (14) will be positive definite. The property of being completely monotone will soon be transferred to the entropy through a new relation between probability and entropy reduction [equations (16) and (17) below].

As a particular case of (14), consider the joint distribution of any two order-statistics $Y_{(r)}$ and $Y_{(s)}$ from the bottom with $r < s$. This is given by

$$g_{r,s}(y_1, y_2)\, dy_1\, dy_2 = n! \frac{[F(y_1)]^{r-1}}{(r-1)!} \frac{[F(y_2) - F(y_1)]^{s-r-1}}{(s-r-1)!} \frac{[1 - F(y_2)]^{n-s}}{(n-s)!} \times f(y_1) f(y_2)\, dy_1\, dy_2 \tag{15}$$

where $y_2 \geq y_1$, so that (15) is always positive, as it must be. When the joint probability distribution for order-statistics from the top is considered, all that is needed is to interchange the distribution function $F(y)$ with the tail $1 - F(x)$ and the tail $1 - F(y)$ with the distribution function $F(x)$ at the terminal states.

We now focus our attention on the asymptotic behaviour of these probability distributions when the sample size is allowed to increase without limit. To this end, it is convenient to introduce the entropy reduction ΔS (Lavenda and Florio, 1992), whose magnitude represents the number of particles, events, or elements $\leq y$, namely

$$nF(y) = -\Delta S(y) := S_0 - S(y) \tag{16}$$

where S_0 is the value of the entropy in the lower terminal state $F(y_0) = 0$. Alternatively, if x denotes the rth value from the top, we would have to replace (16) by

$$n\{1 - F(x)\} = -\Delta S(x) := S_\infty - S(x) \tag{17}$$

[3] One usually considers the multinomial distribution as a generalization of the binomial distribution. The essential distinction between the binomial and beta distributions is that in the former the number of balls in urns of fixed size is considered, whereas in the latter the number of balls is fixed and the urns are allowed to vary in size. In essence, there is statistical equivalence between the two and this will be discussed in the last section.

where S_∞ is the value of the entropy at the upper terminal state, $F(x_\infty) = 1$. From these expressions we are to conclude that the entropies of the extremes are always the largest, and implicit in both (16) and (17) is the assumption that $|\Delta S(x)| < n$.

Gumbel (1954, 1958) defines the *expected largest value* x_{sup} as the solution to

$$n\{1 - F(x_{\text{sup}})\} = 1 \tag{18}$$

This is the value of the variate that has expectation of being exceeded only once in a sample of size n. To see what this means in terms of the entropy reduction (16), we reinstate Boltzmann's constant k and write condition (18) as $\Delta S(x_{\text{sup}}) = -k$. This is the largest reduction in entropy that is possible. The entropy of the largest characteristic value is the closest to the maximum entropy at the extreme. The entropy reduction of the characteristic mth extreme value, x_m with $m \leq n$, $\Delta S(x_m) = -mk$, will always be smaller than the entropy reduction of the largest characteristic value. In other words, the entropy is a monotonically *increasing* function for order-statistics at the upper extreme.

Analogously Gumbel (1954, 1958) defines the *smallest characteristic value* y_{inf} by the condition

$$nF(y_{\text{inf}}) = 1 \tag{19}$$

Comparing this with the mth smallest characteristic value in terms of their entropies, we find $S(y_{\text{int}}) > S(y_m)$, but now since $y_{\text{inf}} < y_m$, the entropy is a monotonically *decreasing* function of its argument.

Inserting (16) into (12), and considering n large enough so that Stirling's approximation can be applied in the form $n! \approx n^n e^{-n}$, we can express the probability density of the order-statistic $Y_{(r)}$

$$g_r(S(y)) = \frac{[S_0 - S(y)]^{r-1}}{(r-1)!} \frac{[1 + (S_0 - S(y))/n]^{n-r}}{(1 - r/n)^{n-r} e^r}$$

in terms of the entropy reduction from its maximum value at the bottom S_0. Now letting $n \to \infty$ results in

$$g_r(S(y))\, dS(y) = \frac{[S_0 - S(y)]^{r-1}}{(r-1)!} e^{S(y)-S_0}\, dS(y) \tag{20}$$

which is easily recognized as the gamma density (Cramér, 1946). In the case $r = 1$, (20) becomes the asymptotic distribution for the *smallest* value.

The same gamma distribution is obtained for the rth value from the top, namely

$$g_r(S(x))\,dS(x) = \frac{[S_\infty - S(x)]^{r-1}}{(r-1)!}\, e^{S(x)-S_\infty}\, dS(x) \tag{21}$$

When we consider the joint distribution of the rth value of x from the top and the rth value of y from the bottom, then, in the limit as $n \to \infty$, we find that the joint distribution factors into a product of the two gamma densities (20) and (21) (Cramér, 1946) [cf. (29) below]. This bears witness to the fact that the statistics from either extreme is independent of the other provided the sample size is sufficiently large to warrant Stirling's approximation. This also happens to be the same condition for the validity of thermodynamics (Lavenda, 1991).

The relation between the entropy reduction and the probability distribution (16) can be used to cast the joint distribution (14) into the suggestive form

$$\begin{aligned} & g_{r_1,r_2,\ldots,r_k}(y_1, y_2, \ldots, y_k)\, dy_1\, dy_2 \cdots dy_k \\ &= \frac{[-\Delta S(y_1)]^{r_1-1}}{(r_1-1)!} S'(y_1)\, dy_1 \\ &\quad \times \prod_{i=1}^{k-1} \frac{[S(y_i) - S(y_{i+1})]^{r_{i+1}-r_i-1}}{(r_{i+1}-r_i-1)!} S'(y_{i+1})\, dy_{i+1} e^{\Delta S(y_k)} \end{aligned} \tag{22}$$

where we have used the relation $F(y_i) - F(y_{i-1}) = [S(y_{i-1}) - S(y_i)]/n$, $r_{k+1} = n + 1$, and the fact that the population is large enough so that Stirling's approximation applies. Likewise, by introducing (17) into (13), we obtain the joint distribution for a population of size n in k intervals from the top as

$$\begin{aligned} & g_{r_1,r_2,\ldots,r_k}(x_1, x_2, \ldots, x_k)\, dx_1\, dx_2 \cdots dx_k \\ &= e^{\Delta S(x_1)} S'(x_1)\, dx_1 \\ &\quad \times \prod_{i=1}^{k-1} \frac{[S(x_{i+1}) - S(x_i)]^{r_{i+1}-r_i-1}}{(r_{i+1}-r_i-1)!} S'(x_{i+1})\, dx_{i+1} \frac{[-\Delta S(x_k)]^{r_k-1}}{(r_k-1)!} \end{aligned} \tag{23}$$

where the difference in the distributions functions is now $F(x_{i+1}) - F(x_i) = [S(x_{i+1}) - S(x_i)]/n$, $r_1 = n + 1$, and we have considered n large enough to apply Stirling's approximation. For $k = 1$, (22) and (23) reduce to the gamma distributions (20) and (21), respectively. For $r = 1$ they become the asymptotic distributions for the *smallest* and *largest* values, respectively.

Expressed in words, (22) and (23) relate the probability that an element will be found in a given interval to the difference in entropy of the interval. The larger the difference in the entropies of the endpoints of the interval, the greater is the probability that an event will occur in that interval. The presence of the exponential factors is merely a statement of the conservation of probability in the large-population limit. Moreover, (22) shows that the entropy is a monotonically *decreasing* function from the bottom, while (23) shows it is monotonically *increasing* for order-statistics from the top. These properties of the entropy in the two extremes will be of aid in determining the characteristic forms of the entropy reduction for smallest and largest values in the next section.

In both extremes Boltzmann's principle is asymptotically satisfied in a way in which Boltzmann himself would not have appreciated. Instead of a *thermodynamic probability*, which is a large number represented by a binomial or multinomial coefficient, what we can refer to as Boltzmann's principles apply to the probability distribution, or its tail, depending upon whether we are considering the smallest or largest values, respectively. Boltzmann's principle is either given by

$$n[1 - F(a_n x)] = -\Delta S(a_n x) \to -\log \Pr(a_n^{-1} X_{(n)} \leq x) \tag{24}$$

or

$$nF(a_n y) = -\Delta S(a_n y) \to -\log \Pr(a_n^{-1} Y_{(1)} > y) \tag{25}$$

depending upon whether we are considering the largest and smallest values, respectively, in the *asymptotic* limit as $n \to \infty$ with x and y fixed. The asymptotic nature of this result will be readily apparent when we consider the conditional probability distribution (27) below. A comparison of the reference states will show that it is as if the information of the initial state has "worn off" and the system has "forgotten" which state it originated in. Moreover, the entropies of the extremes are always greater than any intermediary value of the entropy, and the extremes are statistically independent in the asymptotic limit as $n \to \infty$ [equation (29) below].

Let us now consider the conditional probability distributions that can be constructed from the ratio of the joint probabilities (22) and (23) and the "single gate" probability distributions (20) and (21), respectively. The probability density of the order-statistic $Y_{(r)}$ given $Y_{(s)}$ from the bottom with $r < s$ is

$$\begin{aligned} g_{r,s-r}(y_1 \mid y_2)\, dy_1 &= \frac{g_{r,s}(y_1, y_2)\, dy_1\, dy_2}{g_s(y_2)\, dy_2} \\ &= \frac{p(y_1)^{r-1}[1 - p(y_1)]^{s-r-1}}{B(r, s - r)}\, dp(y_1) \end{aligned} \tag{26}$$

which is again a beta distribution, except that now the variate is $p(y_1) = \Delta S(y_1)/\Delta S(y_2) < 1$. Again this shows that the entropy of order-statistics from the bottom is a monotonically decreasing function.

The conditional probability distribution of the order-statistic $X_{(r)}$ given $X_{(s)}$ from the top with $r < s$ is

$$g_{s-r}(x_1 \mid x_2)\, dx_1 = \frac{g_{r,s}(x_1, x_2)\, dx_1\, dx_2}{g_r(x_2)\, dx_2}$$
$$= \frac{[S(x_2) - S(x_1)]^{s-r-1}}{(s - r - 1)!}\, e^{S(x_1)-S(x_2)} S'(x_1)\, dx_1 \qquad (27)$$

It is quite remarkable that conditional probability density (27) has the same form as the probability density (21), with the exception that the reference state is *not* the state of maximum entropy, but rather the entropy of that state which conditions the distribution.

Finally, we consider the joint probability distribution of the rth value of x from the top and the sth value of y from the bottom with $x > y$. Following the same line of reasoning as before we obtain the expression (Cramér, 1946)

$$g_{r,s}(x, y)\, dx\, dy$$
$$= n! \frac{[1 - F(x)]^{r-1}}{(r - 1)!} \frac{[F(y)]^{s-1}}{(s - 1)!} \frac{[F(x) - F(y)]^{n-r-s}}{(n - r - s)!} f(x)f(y)\, dx\, dy \quad (28)$$

for the joint probability distribution. In terms of the entropy reductions (16) and (17), the joint probability distribution (28) is

$$g_{r,s}(x, y)\, dx\, dy$$
$$= \frac{[-\Delta S(x)]^{r-1}}{(r - 1)!} \frac{[-\Delta S(y)]^{s-1}}{(s - 1)!}$$
$$\times \frac{[1 + \Delta S(x)/n + \Delta S(y)/n]^{n-r-s}}{[1 - (r + s)/n]^{n-r-s} e^{r+s}} S'(x)S'(y)\, dx\, dy$$

for a sufficiently large population. The additivity of the entropies in the last expression is indicative of a lack of correlation between the order-statistics at the two extremes. This becomes apparent in the asymptotic limit as $n \to \infty$, for in this limit the joint probability distribution reduces to a product of gamma distributions (Cramér, 1946),

$$g_{r,s}(x, y)\, dx\, dy$$
$$= \frac{[S_\infty - S(x)]^{r-1}}{(r - 1)!} e^{S(x)-S_\infty} S'(x)\, dx$$
$$\times \frac{[S_0 - S(y)]^{s-1}}{(s - 1)!} e^{S(y)-S_0} S'(y)\, dy \qquad (29)$$

This implies that order-statistics from the top and bottom are statistically independent in the asymptotic limit of a large population.

The statistical independence of the statistics at the extremes is indeed welcome; for although there exist theoretically continuous frequency distributions uniting upper and lower limits of the order-statistics, very few are known in analytical form. Consequently, we should not expect there to be a single expression for the entropy which is valid at both extremes simultaneously. Rather, we should expect to find *asymptotic* expressions for the entropy in the extremes that are related to the asymptotic extreme-value distributions through the Boltzmann principles (25) and (24), involving the initial distribution and its tail, respectively.

In order to derive the expressions for the entropies in the extremes we use the fact that any putative expression for the entropy must be a *concave* function (Lavenda, 1991), and the conditions imposed on the entropies in the joint probability distributions (22) and (23). The remarkable fact that there are only three classes of asymptotic extreme distributions for the largest value, and the corresponding three classes for the smallest value, together with the property of scale invariance will select two asymptotic expressions for entropy, one for the largest value and the other for the smallest value.

4. GEOMETRY, PROBABILITY, AND ENTROPY

We have seen that order-statistics requires the entropy to be a completely monotone function, increasing for order-statistics from the top and decreasing for order-statistics from the bottom, and because the extremes are independent there will be two asymptotic forms for the entropy for the largest and smallest values. This property of the entropy follows from the fact that the distribution function or its Laplace transform is completely monotone (Feller, 1971) [cf. equations (24) and (25)].

For the distribution of the smallest value, the entropy must be monotonically decreasing and scale invariance implies that it varies as a fixed power of the positive variate y,

$$\Delta S(y) = -n\left(\frac{y}{y_0}\right)^D \tag{30}$$

where y_0 is the range of y, and the positive exponent D is known as the Weibull modulus. The monomial form of the entropy reduction (30) is corroborated by the fact that $p(y_1) = \Delta S(y_1)/\Delta S(y_2) = (y_1/y_2)^D < 1$, where $y_2 \geq y_1$ in the conditional probability distribution (26). We can thus identify the magnitude of the entropy reduction in (30) with the number of particles, say, within a distance y, according to the scaling law (7).

In other words, the number of particles less than or equal to y is given by the reduction in entropy

$$\mathrm{Nr}(Y_i \leq y) = S_0 - S(y) \tag{31}$$

Introducing (31) into (30) and taking logarithms yields

$$\log \mathrm{Nr}(Y_i \leq y) = \log n + D \log\left(\frac{y}{y_0}\right) \tag{32}$$

The parameter n determines the characteristic distance $y_0 n^{-1/D}$ which makes $\mathrm{Nr}(Y_i \leq y) = 1$. It will be readily appreciated that this value coincides with the smallest characteristic value y_{inf} in (19). The scaling parameter is $a_n = n^{-1/D}$ and tends to zero as $n \to \infty$. In terms of the nearest neighbor model, this is the smallest distance at which one particle will be found to a given particle [see equation (52) below and accompanying discussion]. In other words $n y_0^{-D}$ can be thought of as a *cluster density* having dimensions of $(\mathrm{length})^{-D}$.

In the case of the order-statistics from the top, the joint probability distribution (23) requires the entropy to be a monotonically increasing function. This, together with the fact that it must be scale invariant, leads to the identification of the (inverted) gamma density (21) (as a function of x) in the extreme $r = 1$ with the Fréchet density for the largest value. Hence the entropy reduction is

$$\Delta S(x) = -n\left(\frac{x_0}{x}\right)^B \tag{33}$$

where the parameter x_0 represents the lower cutoff on x, like the smallest income for which the Pareto distribution is valid.

According to the scaling relation (4), the number of objects greater than x is given by the following expression for the entropy reduction:

$$\mathrm{Nr}(X_i > x) = S_\infty - S(x) \tag{34}$$

Now introducing (34) into (33) and taking logarithms leads to

$$\log \mathrm{Nr}(X_i > x) = \log n - B \log\left(\frac{x}{x_0}\right) \tag{35}$$

The parameter n now determines the largest characteristic distance $x_0 n^{1/B}$ if x represents a distance. Consequently, the scaling is $a_n = n^{1/B}$ and goes to infinity as n does. And just as $n y_0^{-D}$ represents a cluster density, $n x_0^B$ represents *fractal rarity*, sponginess or hollowness, with units $(\mathrm{length})^B$. Formulas (32) and (35)—and not the expressions for the characteristic exponents (3) and (6)—are what characterize clusters and fractals, respectively.

As an illustration of how the transition from probability to physics can be accomplished, the tail, or the probability, can be set equal to the ratio of the pressure P to the pressure of an ideal gas $P_{\text{mat}} = nT/V$ occupying a volume V. In the former case,

$$1 - F(r) = \frac{P(r)}{P_{\text{mat}}(r)} \tag{36}$$

which cannot be greater than unity without jeopardizing the stability of the system. For instance, if we are discussing stellar stability, P would be identified with the radiation pressure. It is then easy to see that (36) is a proper fraction; the radiation pressure is one-third of the energy density. With $P(r)$ as the radiation pressure, (36) becomes

$$1 - F(r) \sim \left(\frac{T}{\hbar c}\right)^3 \frac{r^3}{n} \tag{37}$$

where r^3 is proportional to the volume of a star containing n particles of mass $m = \mu n$, μ being the mass of a proton.[4] The right-hand side of (37) is the cube of the ratio of the wavelength per particle, $r/n^{1/3}$, to the thermal wavelength, $\lambda_T = \hbar c/T$. In regard to the fact that the star cannot be radiation-dominated, or that (36) must be a proper fraction, the thermal wavelength of radiation λ_T is an upper limit on the linear dimension accessible to a particle.

The ratio (36) can be turned into a stability criterion by considering the condition of thermal equilibrium,

$$T \sim \frac{Gm\mu}{r} \tag{38}$$

in the presence of gravitational attraction, where G is the Newtonian gravitational constant. Expression (38) is the nonrelativistic, nondegenerate virial theorem, which equates the gravitational energy per particle to the thermal energy. Introducing (38) into (37) results in

$$1 - F(r) \sim \left(\frac{n}{n_*}\right)^2 \tag{39}$$

where $n_* = (\hbar c/G\mu^2)^{3/2} = \alpha_G^{-3/2}$ is the maximum number of protons that a star can contain without jeopardizing its stability. This is known as the Chandrasekhar limit, and the corresponding mass is the Chandrasekhar mass.

Therefore, the entropy reduction corresponding to the tail distribution (39) is

[4] In classical thermodynamics of degenerate gases, $n[1 - F(r)]$ would either be identified with the entropy or the particle number.

$$\Delta S(n) \sim -\left(\frac{n}{n_*}\right)^2 n \tag{40}$$

Introducing the scaling relation

$$n(r) \sim n_0 r^D \tag{41}$$

where n_0 is space independent, gives the *thermodynamic force* upon differentiation,

$$\frac{\chi}{T} \sim \left(\frac{n_0}{n_*}\right)^2 n_0 r^{3D-1} \tag{42}$$

Finally, introducing the equilibrium temperature (38) gives the gravitational force as

$$\chi = \frac{d\mathcal{W}}{dr} \sim \left(\frac{n_0}{n_*}\right)^2 \frac{Gm^2}{r^2} \tag{43}$$

which upon integration gives

$$\mathcal{W}(r) \sim \left(\frac{n}{n_*}\right)^2 \frac{Gm^2}{r} \tag{44}$$

The force χ is the derivative of the work (Landau and Lifshitz, 1938) done in bringing n particles from infinity to within a spherical volume of the star with radius r. Expression (46) brings out the degenerate nature of the star, in that it reduces the negative of the potential energy by a factor of $(n/n_*)^2$, or increases the potential energy by the same factor.

In fact these relations apply regardless of whether the scaling exponent D in (41) is positive or negative. In the case of white dwarfs, $n \sim r^{-3}$ (Sexl and Sexl, 1979), while in the case of clustering, D is positive and less than 3.

Consequently, the decrease in the entropy when some external source brings the system from a state of equilibrium to any given state is

$$\Delta S(r) = -\frac{\mathcal{W}(r)}{T(r)} \tag{45}$$

where $\mathcal{W}(r)$ is the work necessary to accomplish the change and $\Delta S(r)$ is the *total* entropy reduction of the n particles. The relation (45) is reminiscent of the principle of minimum work proposed by Landau and Lifshitz (1938) in conjunction with a generalization of the usual Gibbs relation. We emphasize that there is no Gibbs relation which is valid for inhomogeneous systems and that the minimum work is a function of a sole independent coordinate, the radius of the sphere. Moreover, according to (36), we can express the work in (45) as

$$\mathcal{W}(r) = P(r)V \tag{46}$$

where V is the volume of the system. For an ideal gas $PV = \eta U$, where $\eta = 1/3$ for a photon gas or $\eta = 2/3$ for a material gas, and U is the internal energy. The entropy reduction is thus $\Delta S = -\eta n$; the change in entropy is the largest possible and $\eta \leq 1$. As we have appreciated from (40), interactions among the particles can only lead to a decrease in the entropy reduction, i.e., it becomes more negative.

5. STATISTICAL EQUIVALENCE PRINCIPLE

The formulas relating the number of particles less than or greater than a given value to the reduction in entropy, (31) and (34), respectively, can be used to formulate a statistical equivalence principle. Let us observe that the rth order-statistic $Y_{(r)} \leq y$ iff there are *r or more* of the Y_i that are less than or equal to y. Call this number $N(y)$, where we explicitly indicate the dependence on the value y. Consequently,

$$\Pr(N(y) \geq r) = \Pr(Y_{(r)} \leq y) \tag{47}$$

In Section 3 we showed that in the asymptotic limit of an infinitely large population the cumulative distribution function of $Y_{(r)}$ is

$$\Pr(Y_{(r)} \leq y) = \int_0^{N(y)} \frac{t^{r-1}}{\Gamma(r)} e^{-t} dt \tag{48}$$

In the same limit, the probability that there are r or more values of the Y_i less than or equal to y is

$$\Pr(N(y) \geq r) = \sum_{i=r}^{\infty} \frac{[N(y)]^i}{i!} e^{-N(y)} = 1 - \sum_{i=0}^{r} \frac{[N(y)]^i}{i!} e^{-N(y)} \tag{49}$$

Introducing (48) and (49) into (47) give the analytical identity[5]

[5] The asymptotic relation for $n \to \infty$ can be derived from the equivalence between the cumulative distributions of the binomial and the beta

$$\sum_{k=r}^{n} \binom{n}{k} p^k (1-p)^{n-k} = \frac{\int_0^p t^{r-1}(1-t)^{n-r}\, dt}{B(r, n-r+1)}$$

In the limits that $n \to \infty$ and $p \to 0$, such that the product np is moderate, the left side transforms into the tail of the Poisson distribution, while the right side becomes an incomplete gamma function.

$$\Pr(N(y) \geq r) = 1 - \sum_{j=0}^{r-1} \frac{[N(y)]^j}{j!} e^{-N(y)} = \int_0^{N(y)} \frac{t^{r-1}}{(r-1)!} e^{-t} dt$$

which can easily be verified by integrating by parts. The last integral is simply

$$\Pr(Y_{(r)} \leq y) = \int_0^y \frac{[N(u)]^{r-1}}{(r-1)!} e^{-N(u)} N'(u)\, du$$

Now according to (31), $N(y)$ is the number of the Y_i less than or equal to y which coincides with the entropy reduction $S_0 - S(y)$ from the bottom. Hence, the statistical equivalence principle (47) can be expressed as

$$\Pr(S_0 - S(y) \geq r) = \Pr(Y_{(r)} \leq y) \tag{50}$$

Analogously, in order for $X_{(r)} \geq x$ there cannot be more than r of the X_i that are less than or equal to x. The statistical equivalence principle is now given by

$$\sum_{j=0}^{r} \frac{[N(x)]^{j-1}}{j!} e^{-N(x)} = \int_{N(x)}^{\infty} \frac{t^{r-1}}{\Gamma(r)} e^{-t} dt$$

or equivalently in terms of the entropy reduction from the top (34) as

$$\Pr(S_\infty - S(x) \leq r) = \Pr(X_{(r)} \geq x) \tag{51}$$

The statistical equivalence principles (50) and (51) assert that the specification of intervals for a fixed number of events, and the number of events occurring in fixed intervals, are statistically equivalent.

The entropy of a classical perfect gas is of the form $-n \log n$, while for an ideal degenerate gas it varies as n itself (Lavenda, 1995). For inhomogeneous systems like fractals and clusters, the magnitude of the entropy reduction also varies as n. This number can also increase or decrease with the independent variable according to expressions (30) and (33), respectively.

Let Y denote the distance between a particle placed at the origin and the nearest neighbor. The probability $g(y)\, dy$ that the nearest neighbor occurs between y and $y + dy$ is the product of the probability that no particle occurs within a distance y and the probability $Dy^{D-1}/y_0^D\, dy$ that a particle exists between y and $y + dy$ in an arbitrary space of dimension D, where y_0 is the radius of the entire hypersphere. Consequently, $g(y)$ must satisfy the functional relation

$$g(y) = \left[1 - \int_0^y g(s)\, ds\right] D \frac{y^{D-1}}{y_0^D} \tag{52}$$

The functional relation may be solved by first differentiating and then integrating with respect to y (Chandrasekhar, 1943). The result is

$$g(y) = D \frac{y^{D-1}}{y_0^D} e^{-(y/y_0)^D} \tag{53}$$

which is the nearest neighbor probability density in a space of D dimensions. The nearest neighbor probability distribution is a special case of the Weibull distribution for the smallest value. It can be generalized to r nearest neighbors by simply employing the results of Section 3 for order-statistics (Lavenda, 1995).

The functional relation (52) reduces to

$$g(y) \to D \frac{y^{D-1}}{y_0^D} \qquad \text{as} \quad y \to 0$$

This is confirmed by the solution (53) we have found. In this limit we have

$$\Pr(Y \leq y) = \left(\frac{y}{y_0}\right)^D$$

which is the negative of the entropy reduction (30) for a single particle—the nearest neighbor.

Now let $g(x)\,dx$ denote the probability that the *furthest* neighbor to a particle occurs between x and $x + dx$. This is equal to the probability of there being no particles further than $x + dx$ and the probability $Bx_0^B/x^{B+1}dx$ that the furthest particle will be found between x and $x + dx$. By the conservation of probability $[1 - \int_x^\infty g(s)\,ds] = \int_{x_0}^x g(s)\,ds$, where the lower cutoff x_0 can be thought of as the radius of the excluded volume about the central particle, it follows that $g(x)$ must satisfy the functional relation

$$g(x) = \int_{x_0}^x g(s)\,ds\, \frac{Bx_0^B}{x^{B+1}} \tag{54}$$

Rearranging and differentiating (54) leads to

$$\frac{d}{dx}\left[\frac{g(x)x^{B+1}}{Bx_0^B}\right] = \frac{Bx_0^B}{x^{B+1}} \frac{g(x)x^{B+1}}{Bx_0^B}$$

This can be integrated to give

$$g(x) = \frac{Bx_0^D}{x^{B+1}} e^{-(x_0/x)^B} \tag{55}$$

which will be easily recognized as Fréchet's probability density for the largest value.

On the strength of the conservation of probability, the functional relation (54) tends to

$$g(x) \to \frac{Bx_0^D}{x^{B+1}} \qquad \text{as} \quad x \to \infty$$

This is substantiated by the solution we have found, (55), and for large x

$$\Pr(X > x) = \left(\frac{x_0}{x}\right)^B$$

which is equal to the negative of the entropy reduction (33) for a single particle—the furthest neighbor.

ACKNOWLEDGMENTS

This work was supported in part by the EU Human Capital and Mobility Programme, Accordo Programma MICA-ENEA, MURST, and CNR.

REFERENCES

Castillo, E. (1988). *Extreme Value Theory in Engineering*, Academic Press, San Diego, California.

Chandrasekhar, S. (1943). *Reviews of Modern Physics*, **15**, 1 [reprinted in *Noise and Stochastic Processes*, N. Wax, ed., Dover, New York (1954), pp. 3–91].

Cramér, H. (1946). *Mathematical Methods of Statistics*, Princeton University Press, Princeton, New Jersey, §28.6.

De Finetti, B. (1970). *Theory of Probability*, Wiley, Chichester, England, Vol. 2, pp. 100–101.

Feller, W. (1971). *An Introduction to Probability Theory and Its Applications*, Wiley, New York, Vol. II, p. 439.

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.

Hastings, H. M., and Sugihara, G. (1993). *Fractals: A User's Guide for the Natural Sciences*, Oxford University Press, Oxford.

Kendall, M. G., and Stuart, A. (1969). *Advanced Theory of Statistics*, Vol. 1, 3rd ed., Griffin, London, Chapter 14.

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

Landau, L. D., and Lifshitz, L. D. (1938). *Statistical Physics*, 1st ed., Clarendon Press, Oxford, §37.

Lavenda, B. H. (1991). *Statistical Physics: A Probabilistic Approach*, Wiley-Interscience, New York, Chapter 1.

Lavenda, B. H. (1995). *Thermodynamics of Extremes*, Albion, Chichester, England.

Lavenda, B. H., and Florio, A. (1992). *International Journal of Theoretical Physics*, **31**, 1455.

Mendelbrot, B. B. (1982). *The Fractal Geometry of Nature*, Freeman, New York.

Sexl, R., and Sexl, H. (1979). *White Dwarfs–Black Holes*, Academic Press, New York, p. 68.

Zipf, G. K. (1949). *Human Behavior and the Principle of Least Effort*, Addison-Wesley, Cambridge, Massachusetts.