The analogy between coding theory and multifractals

1998 J. Phys. A: Math. Gen. 31 5651

(http://iopscience.iop.org/0305-4470/31/26/006)

J. Phys. A: Math. Gen. **31** (1998) 5651–5660. Printed in the UK PII: S0305-4470(98)90496-6

The analogy between coding theory and multifractals

B H Lavenda†
University of Camerino, Camerino, 62032 (MC), Italy

Received 8 January 1998, in final form 28 April 1998

Abstract. An analogy exists between a coding theorem and attractors considered as multifractals. The average smallest wordlength of order τ for the coding theorem is analogous to the average smallest volume of order τ. Diminishing characteristic means for measuring the radii of spheres containing points on an attractor decrease the dimensionality which is always less than the Hausdorff dimension in non-uniform systems. The Rényi entropy is the lower bound to the average largest wordlength, and the logarithm of the reciprocal of the average smallest volume of order τ. The asymptotic distributions for minimum length and volume are derived.

1. Introduction

The Rényi entropy, or the entropy of order α, is an interpolation formula connecting the Shannon ($\alpha = 1$) and Hartley ($\alpha = 0$) entropies. The Shannon entropy is the lower bound to the average code length for a noiseless channel in a uniquely decipherable code (Feinstein 1958). Campbell (1965) gave a generalization to include code lengths of order $\tau = (1-\alpha)/\alpha$, where the Rényi entropy is the lower bound on the code lengths. The generalization involves a measure of the largest wordlengths. Regarding a measure of the smallest wordlengths, we shall derive an upper bound on the average smallest code length of order τ.

The Rényi entropy has also been implicated in the definition of the generalized dimension of strange attractors (Grassberger and Procaccia 1983). There is a close analogy between strange attractors and multifractals (Schroeder 1991). Multifractals can be generated on a fractal support of the unit interval by removing a number of open intervals leaving behind line segments of length r_i separated by holes. Each length r_i has a probability p_i, and after an infinite number of iterations we are left with specks of dust having given probabilities. These specks, or density of points, are what characterize strange attractors. More specifically, we go to some point on or near the attractor and ask for the number of points on the orbit within a distance r of this point as we make the radius r of the sphere enclosing these points which are small in comparison with the size of the attractor (Abarbanel 1996). The average volume of spheres of order τ containing the density of points, in the limit as the radius shrinks to zero, will play the same role as the average smallest code length of order τ in the coding theorem. Since this weighted mean is analogous to that of the mean code length of order τ, we are able to apply the same approach to derive upper bounds on the average smallest volumes, or equivalently, lower bounds on the negative of their logarithms. Thus, by a mere change in terminology, the generalized coding theorem is applicable to strange attractors when considered as multifractals.

† E-mail address: lavenda@camserv.unicam.it

0305-4470/98/265651+10$19.50

2. Entropies, inequalities and monotonic functions

The set of probabilities $p_1, p_2, \ldots, p_N$ can either stand for the probabilities of N input symbols $x_1, x_2, \ldots, x_N$ that are to be encoded, or the probabilities for fractal segments of lengths $r_1, r_2, \ldots, r_N$. A message of N symbols will contain on average $p_1 N$ occurrences of the first symbol, $p_2 N$ occurrences of the second and so on. The probability of this particular message will be

$$\tilde{p}^N = \prod_{i=1}^{N} p_i^{p_i N}$$

where $\tilde{p}$ is the geometric mean of the probabilities. The Shannon entropy is defined as the logarithm of the reciprocal of the geometric mean of the probabilities of a typically long sequence (Shannon and Weaver 1959)

$$H_1(p) = \log(1/\tilde{p}) = -\sum_{i=1}^{N} p_i \log p_i. \tag{1}$$

The probabilities $p_i > 0$ form a complete distribution, $\sum_{i=1}^N p_i = 1$, and the logarithm is to the base $d > 1$. The reciprocal of the geometric mean is a large number representing the 'thermodynamic' probability of a given state (Lavenda 1991). The geometric mean of the probabilities, $\tilde{p}$, is Schur-convex because its logarithm, $\sum_{i=1}^N p_i \log p_i$, is Schur-convex. The negative of the latter is the Shannon entropy and it is Schur-concave.

Based upon the concavity of the Shannon entropy, (1), it follows that if $q_i > 0$ and $q_i \neq p_i$

$$-\sum_{i=1}^{N} p_i \log\left(\frac{p_i}{q_i}\right) \leqslant 0. \tag{2}$$

This is known as Shannon's inequality (Aczél and Daróczy 1975), and provides the motivation for using majorization as a means of deriving many different forms of inequalities.

In a state of uniform probability, $p_i = 1/N$, and Shannon's entropy reduces to Hartley's entropy

$$H_0 = \log N \tag{3}$$

who believed that entropy should depend on the number of different kinds of objects, not on their probabilities. Rényi proposed:

$$H_\alpha = \frac{1}{1-\alpha} \log \sum_{i=1}^{N} p^\alpha \tag{4}$$

as an interpolating formula which spans the gamut between the Shannon (1) and Hartley (3) entropies. In the limit as $\alpha \to 1$, the Rényi entropy (4) transforms into the Shannon entropy (1), while in the limit as $\alpha \to 0$, it becomes the Hartley entropy (3).

The components of a vector $x = (x_1, x_2, \ldots, x_N)$ are said to be 'less spread out' or 'more nearly equal' than those of a vector $y = (y_1, y_2, \ldots, y_N)$ if x is majorized by y: $x \prec y$. Interest in majorization arises by asking for conditions on x and y in order that the inequality $\sum_{i=1}^N \varphi(x_i) \leqslant \sum_{i=1}^N \varphi(y_i)$ holds for all convex functions φ. A necessary and sufficient condition for the inequality to hold for all convex functions φ is that x be majorized by y (Hardy *et al* 1952). More specifically, we may say that x is weakly majorized by y, which we will denote by $x \prec_{\rm w} y$, if $\sum_{i=1}^{L} x_{[i]} \leqslant \sum_{i=1}^{L} y_{[i]}$, for $L = 1, \ldots, N$, where the $x_{[i]}$ are ordered from largest to smallest, $x_{[1]} \geqslant \cdots \geqslant x_{[N]}$ (Marshall and Olkin 1979).

Since the probabilities $p_i > 0$, and form a complete distribution, it follows that $(1/N, \ldots, 1/N) \prec p$. If φ is any Schur-convex function then $\varphi(p) \geqslant \varphi(1/N)$. The Shannon entropy being Schur-concave provides a measure of the degree of uniformity of the distribution; it is bounded above by the Hartley entropy, $H_1 \leqslant H_0$. Moreover, since the Rényi entropy (4) is Schur-concave on the open interval $\alpha \in (0, 1)$, it follows that

$$H_1 \leqslant H_\alpha \leqslant H_0. \tag{5}$$

The Rényi entropy is a monotonically decreasing function of α on the open interval $(0, 1)$. The first inequality in (5) is Jensen's inequality, $(\boldsymbol{E} p^r)^{1/r} \leqslant \exp(\boldsymbol{E} \log p)$ for $r = -(1-\alpha)$, and $\boldsymbol{E}$ denotes expectation. The larger the Rényi entropy, H_α, the more uniform the distribution is.

The Shannon entropy (1) is the logarithm of the reciprocal of the weighted mean

$$M_\varphi(p) = \varphi^{-1}\left(\sum_{i=1}^{N} p_i \varphi(p_i) \Big/ \sum_{i=1}^{N} p_i\right)$$

for $\varphi(p) = \log p$. For purposes of characterization, we admit the possibility of an incomplete distribution, $\sum_{i=1}^{N} p_i \leqslant 1$. Incomplete distributions are necessary in order to reinstate the correct functional form in the entropy expressions so that properties such as additivity, homogeneity, and maximality become manifest (Aczél and Daróczy 1975). The obvious conditions to impose on the function φ are that it is continuous and strictly monotonic, thereby possessing an inverse φ^{-1} which satisfies the same conditions (Hardy *et al* 1952). Consequently, φ has to be a logarithm, $\log p$, or a power, p^γ. These functions are the only two continuous and strictly monotonic functions whose weighted means satisfy $\exp(\log p) = p$ and $M_\gamma(p) = p$ for $N = 1$, and are first-order homogeneous functions: $\exp(\sum_{i=1}^{N} \lambda p_i \log \lambda p_i / \sum_{i=1}^{N} \lambda p_i) = \lambda \exp(\sum_{i=1}^{N} p_i \log p_i / \sum_{i=1}^{N} p_i)$ and $M_\gamma(\lambda p) = \lambda M_\gamma(p)$. Furthermore, if we invoke the condition of 'stability' (Aczél and Daróczy 1975)

$$\lim_{p_2 \to 0} M_\gamma(p_1, p_2) = p_1$$

then $\gamma + 1 > 0$, or setting $\gamma = \alpha - 1$ requires $\alpha > 0$. Although incomplete distributions are useful for entropy characterization, we will henceforth only deal with complete distributions.

In an analogous way that the Rényi entropy (4) is an interpolation formula connecting the Shannon, $\log(1/\tilde{p})$, and Hartley, $\log N$, entropies (5), the weighted mean

$$M_{-(1-\alpha)}(p) = \left(\sum_{i=1}^{N} p_i^\alpha\right)^{-1/(1-\alpha)} \tag{6}$$

connects the geometric mean of the probabilities and the state of uniform probability:

$$\frac{1}{N} \leqslant \left(\sum_{i=1}^{N} p_i^\alpha\right)^{-1/(1-\alpha)} \leqslant \tilde{p}. \tag{7}$$

Hence, the weighted mean (6) is a monotonically increasing function of α on the open interval $(0, 1)$, and is Schur-convex. Its negative can be interpreted as an entropy reduction which is Schur-concave. It constitutes a multidimensional generalization of the one-dimensional expression for strictly stable distributions (Lavenda 1995). This is still another reason for restricting the characteristic exponent α to lie in the open interval $(0, 1)$.

The weighted mean (6) is comparable with

$$M_{-(1-\alpha)}(q) = \left(\sum_{i=1}^{N} p_i q_i^{-(1-\alpha)}\right)^{-1/(1-\alpha)} \tag{8}$$

which lies between the geometric mean, $\tilde{q} = \prod_{i=1}^{N} q_i^{p_i}$, and the harmonic mean, $\hat{q} = (\sum_{i=1}^{N} p_i/q_i)^{-1}$:

$$\hat{q} \leqslant \left(\sum_{i=1}^{N} p_i q_i^{-(1-\alpha)}\right)^{-1/(1-\alpha)} \leqslant \tilde{q}.$$

Consequently, the weighted mean (8) is also a monotonically increasing function of α on the open interval $(0, 1)$. However, unlike (6) which is Schur-convex, (8) is Schur-concave for fixed p.

Since $1/N \prec p$, the uniform state is majorized by p, and (6) is Schur-convex, it follows that

$$\left(\sum_{i=1}^{N} p_i^{\alpha}\right)^{-1/(1-\alpha)} - \frac{1}{N} \geqslant 0. \tag{9}$$

Likewise the probabilities q also majorize the uniform state $1/N \prec q$, so that

$$\frac{1}{N} - \left(\sum_{i=1}^{N} p_i q_i^{-(1-\alpha)}\right)^{-1/(1-\alpha)} \geqslant 0 \tag{10}$$

because (8) is Schur-concave for fixed p and $\alpha < 1$. Adding inequalities (9) and (10) gives

$$\left(\sum_{i=1}^{N} p_i^{\alpha}\right)^{-1/(1-\alpha)} \geqslant \left(\sum_{i=1}^{N} p_i q_i^{-(1-\alpha)}\right)^{-1/(1-\alpha)} \tag{11}$$

showing that $q \prec p$. Shannon's inequality (2) is contained in (11) as a limiting case, namely as $\alpha \to 1$.

Finally, an additional inequality will be of use when discussing characteristic average dimensions. From the fact that $M_r(a) < M_s(a)$ for $r < s$, unless the numbers a are all equal (Hardy *et al* 1952), we have the additional inequality $(\sum_{i=1}^{N} p_i q_i^{-(1-\alpha)})^{-1/(1-\alpha)} \geqslant (\sum_{i=1}^{N} p_i q_i^{-\tau})^{-1/\tau}$ since $\alpha < 1$. In view of inequality (11) we now have

$$\left(\sum_{i=1}^{N} p_i^{\alpha}\right)^{-1/(1-\alpha)} \geqslant \left(\sum_{i=1}^{N} p_i q_i^{-\tau}\right)^{-1/\tau} \tag{12}$$

where $\tau = (1-\alpha)/\alpha$. Inequality (12) could have been derived directly from the Hölder inequality:

$$\left(\sum_{i=1}^{N} x_i^{a}\right)^{1/a} \left(\sum_{i=1}^{N} y_i^{b}\right)^{1/b} \leqslant \sum_{i=1}^{N} x_i y_i \tag{13}$$

with $x_i = p_i^{1/\tau}$, $y_i = p_i^{-1/\tau} q_i$, the exponents $a = (1-\alpha)$ and $b = -\tau$ satisfy the condition $1/a + 1/b = 1$, and $\sum_{i=1}^{N} q_i \leqslant 1$. The coding and multifractal theorems attribute a physical significance to the weighted mean on the right-hand side of (12).

3. A coding theorem revisited

The coding theorem for a noiseless channel is based on the Kraft inequality:

$$\sum_{i=1}^{N} d^{-\kappa_i} \leqslant 1. \tag{14}$$

This is a necessary and sufficient condition for the existence of a unique and decipherable code (Feinstein 1958). The κ_i are the lengths of the code words, N is the number of input symbols, and the alphabet contains $d > 1$ symbols. If the wordlengths κ_i are to satisfy (14) we must have

$$d^{-\kappa_i} \leqslant N^{-1}$$

for at least one i, and consequently for the largest wordlength (Campbell 1965).

The length of a word is the number of symbols it contains. Dividing the number of different words of length κ_i that can be formed from d symbols, d^{κ_i}, by total number of words, $n = \sum_{i=1}^{N} d^{\kappa_i}$, gives the probability $q_i = d^{\kappa_i}/n$ of a word of length κ_i. As a measure of the average code length of order τ we choose the function:

$$\mathcal{L}(\tau) := -\frac{1}{\tau} \log\left(\sum_{i-1}^{N} p_i d^{-\kappa_i \tau}\right). \tag{15}$$

The justification for calling (15) an average code length is based on the extreme limits of (15) obtained by letting τ tend to zero and infinity. In the limit as $\tau \to 0$ ($\alpha \to 1$), the measure of the wordlength of order zero is proportional to the weighted mean

$$\mathcal{L}(0) = \lim_{\tau \to 0} \mathcal{L}(\tau) = \sum_{i=1}^{N} p_i \kappa_i = \bar{\kappa}$$

while, in the limit as $\tau \to \infty$, the entire sum reduces to a single term (Hardy *et al* 1952)

$$\mathcal{L}(\infty) = \lim_{\tau \to \infty} \log M_{-\tau}(d^{\kappa}) = \kappa_{\min} \tag{16}$$

where $\kappa_{\min}$ is the smallest of the numbers $\kappa_1, \ldots, \kappa_N$. The measure of the average smallest wordlength of order τ is a monotonic decreasing function of τ. $\mathcal{L}(0)$ is the usual measure of the mean length, while $\mathcal{L}(\infty)$ is the measure that would be employed to measure the smallest wordlengths. The exponent τ is therefore related to how lengths are measured: the larger the value of τ, the more weight is given to smaller values of κ_i.

Inequality (12) places upper limits on the average and smallest wordlengths. In the limit as $\tau \to 0$, ($\alpha \to 1$), inequality (12) is

$$\log n - H_1 \geqslant \bar{\kappa}. \tag{17}$$

The geometric–harmonic mean inequality,

$$\sum_{i=1}^{N} d^{-\kappa_i} \geqslant N d^{-\bar{\kappa}}$$

and Kraft inequality (14), provide a lower bound to the average wordlength, $\bar{\kappa} \geqslant \log N$. Inserting this into (17) leads to

$$\log n \geqslant H_0 + H_1$$

placing an upper bound on the sum of the Shannon and Hartley entropies. In the opposite limit as $\tau \to 0$, ($\alpha \to 0$), inequality (12) becomes

$$\log n - H_0 \geqslant \kappa_{\min}.$$

The equality sign in (12) is achieved when

$$d^{\kappa_i}/n = p_i^\alpha \Big/ \sum_{i=1}^{N} p_i^\alpha$$

or

$$\kappa_i = \log n + \alpha \log p_i - \log\left(\sum_{i=1}^{N} p_i^\alpha\right).$$

However, because of the intermediary inequality (11), $d^\kappa/n \prec p$ so that the optimal value of κ_i is

$$\kappa_i = \log n + \log p_i.$$

As the probability of a wordlength tends to zero, the number of different words of all lengths must increase so as to render their product, $p_i n$, essentially constant. The same requirement is used to derive the limit laws of probability theory.

Campbell's (1965) use of the Hölder inequality, in which he compared the probabilities $d^{-\kappa_i}$ with p_i, further required the Kraft inequality (14). Whereas Campbell obtained a lower bound to the average *largest* code length of order τ, we obtain an upper bound to the average *smallest* code length of the same order.

Because of the Kraft inequality (14), we must consider the weak majorization $d^{-\kappa} \prec_{\mathrm{w}} p$ that drops the equality constraint on the total sum (Marshall and Olkin 1979). As before, we obtain the same string of inequalities:

$$\left(\sum_{i=1}^{N} p_i^\alpha\right)^{-1/(1-\alpha)} \geqslant \left(\sum_{i=1}^{N} p_i d^{\kappa_i(1-\alpha)}\right)^{-1/(1-\alpha)} \geqslant \left(\sum_{i=1}^{N} p_i d^{\kappa_i \tau}\right)^{-1/\tau}. \tag{18}$$

Equality between the first and last weighted means in (18) is obtained when (Campbell 1965)

$$d^{-\kappa_i} = p_i^\alpha \Big/ \sum_{i=1}^{N} p_i^\alpha$$

or

$$\kappa_i = -\alpha \log p_i + \log\left(\sum_{i=1}^{N} p_i^\alpha\right). \tag{19}$$

From this Campbell concluded that as $p_i \to 0$, the optimum value of κ_i is asymptotically equal to $-\alpha \log p_i$. The minimization of the average code length, subject to the Kraft inequality (14), leads to the result that the best code length of probability p_i is asymptotic to $-\log p_i$ (Feinstein 1958). Large wordlengths have small probabilities. The optimum value of κ_i being asymptotic to $-\alpha \log p_i$, would be an improvement, since the wordlength would be smaller than $-\log p_i$ because $\alpha < 1$ (Campbell 1965). However, the first inequality (18) is only satisfied if $d^{-\kappa_i} = p_i$, or equivalently, $\kappa_i = -\log p_i$.

Since the Rényi entropy is intermediate between the Hartley and Shannon entropies, and in view of inequalities (17) and (16) we have proved the following theorem.

Theorem 1. The average smallest code length of order τ, can never be greater than the difference between the logarithm of the total number of words and the Rényi entropy:

$$\log n - H_\alpha \geqslant \mathcal{L}(\tau).$$

The importance of the Rényi, in comparison with the Shannon entropy, lies in the versatility of how we measure lengths, or, for that matter, any extensive quantity. If the conventional measure of *mean* length is used, the Shannon entropy is its lower bound. However, if we are interested in intermediate measures of length, lying between maximum or minimum lengths, then the Rényi entropy of order $\alpha = (1+\tau)^{-1}$ determines the lower (upper) bound on the largest (smallest) weighted mean of order τ.

4. A coding theorem for strange attractors

The Hausdorff dimension of N equal-length pieces of size r is $Nr^{D_H} = 1$. If the fractal consists of N different lengths, r_i, the Hausdorff dimension is determined by the condition:

$$\sum_{i=1}^{N} r_i^{D_H} = 1. \tag{20}$$

If each segment r_i occurs with probability p_i the generator requires two exponents, one for the support of the fractal, $\tilde{D}_\alpha$, and one for the probabilities, α. The condition which generalizes $Nr^{D_H} =$ constant, can be written as (Halsey *et al* 1986)

$$\sum_{i=1}^{N} p_i^{\alpha} r_i^{\tilde{D}_\alpha(1-\alpha)} = 1. \tag{21}$$

The generalized dimension, $\tilde{D}_\alpha$, coincides with the Hausdorff dimension, D_H, for $\alpha = 0$. Expression (21) appears somewhat similar to the weighted mean (8) when we identify $r_i^{\tilde{D}_\alpha}$ with the probabilities q_i. However, in that case, q_i cannot depend upon the exponent α so that $\tilde{D}_\alpha$ must coincide with the Hausdorff dimension, D_H. This allows us to identify the probabilities q_i with $r_i^{D_H}$ which forms a complete set on account of (20). Rather, what changes is the dimension of the embedded object which is determined by the exponent α.

p-majorization, $r^{D_H}\mathrm{p} \succ 1/N$, with respect to a set of probabilities $p_i > 0$ for every i, implies (Marshall and Olkin 1979)

$$\sum_{i=1}^{N} p_i r_i^{-D_H(1-\alpha)} \geqslant N^{1-\alpha} \tag{22}$$

for all convex functions $q_i^{-(1-\alpha)}$ where $\alpha < 1$. The two intervals of the uniform Cantor set have equal probabilities, and (22) reduces to an equality $r^{-D_H} = N$. In the limit $\alpha \to 1$, (22) reduces to the identity $\sum_{i=1}^{N} p_i = 1$, while in the limit as $\alpha \to 0$, inequality (22) expresses the fact that the harmonic mean $\hat{q} = 1/\sum_{i=1}^{N} p_i/q_i$ is always less than the arithmetic mean, $\sum_{i=1}^{N} q_i/N$ unless all the q_i's are equal.

We are interested in the way the number of points, or specks of dust, within a sphere of radius r_i, scales as the radius shrinks to zero. We shall see that the volume of the sphere behaves as $\check{r}^{D_\alpha}$, where $\check{r}$ is some mean radius, and D_α is the dimensionality of the embedded object. This is not the same as the $\tilde{D}_\alpha$ appearing in (21) because we have insisted that the probabilities, q_i, be independent of the exponent α.

In the same way we introduced a measure of the length of a code, we need a measure of volume. Since we are now concerned with probabilities which are power laws and not exponentials, we define the mean volume of order τ as:

$$\mathcal{V}(\tau) = \left(\sum_{i=1}^{N} p_i r_i^{-D_H\tau}\right)^{-1/\tau}. \tag{23}$$

This mean volume is a decreasing function of τ:

$$\mathcal{V}(0) = \tilde{r}^{D_H} > \mathcal{V}(1) = \widehat{r^{D_H}} > \mathcal{V}(\infty) = r_{\min}^{D_H} \tag{24}$$

where $\tilde{r} = \prod_{i=1}^{N} r_i^{p_i}$ is the geometric mean radius, $\widehat{r^{D_H}} = (\sum_{i=1}^{N} p_i/r_i^{D_H})^{-1}$ is the harmonic mean volume, and (Hardy *et al* 1952)

$$\lim_{\tau\to\infty} \mathcal{V}(\tau) = r_{\min}^{D_H}$$

is the minimum volume. The larger the value of τ, the more weight is given to the smaller values of r_i.

On the strength of inequality (12), the logarithm of the reciprocal of the measure of volume of order τ has a lower bound

$$H_\alpha(p) < \frac{1}{\tau} \log\left(\sum_{i=1}^{N} p_i r_i^{-D_H\tau}\right) \qquad (0 < \tau < \infty) \tag{25}$$

given by the Rényi entropy. This can be verified from the Hölder inequality (13) by setting $x_i = p_i^{1/\tau}$, $y_i = p_i^{-1/\tau} r_i^{D_H}$, with exponents $a = (1-\alpha)$ and $b = -\tau$, independently of whether the probabilities $r_i^{D_H}$ form a complete distribution or not.

Inequality (25) leads to a sense of reduced dimensionality of the attractor in comparison with the Hausdorff dimension of the fractal support in which it is embedded. The dimensionality of the attractor is governed by the order τ. For instance, in the limit as $\tau \to 0$, (25) becomes

$$D_H \geqslant \frac{\log(1/\tilde{p})}{\log(1/\tilde{r})} = \frac{H_1(p)}{\log(1/\tilde{r})} = D_1 \tag{26}$$

which is the 'information' dimension because $H_1(p)$ is the Shannon entropy. The $\tau \to 0$ limit selects out the geometric mean of the radii as the characteristic average distance. For equal probabilities, the Shannon entropy becomes the Hartley entropy and uniform radii would lead to the equality in (26). The crucial point is that we cannot vary the entropy in the numerator of (26) without a corresponding variation in how we measure distance in the denominator. In other words, every entropy function will have a corresponding characteristic measure of distance.

The order $\tau = 1$ selects the harmonic mean, with dimension

$$D_{1/2} = \log\sum_{i=1}^{N} p_i^{1/2} \Big/ \log\left(1\Big/ \sqrt[D_H]{\widehat{r^{D_H}}}\right).$$

Increasing values of τ, or decreasing values of α, give more weight to smaller characteristic distances with decreasing dimensionality. In the limit as $\tau \to \infty$ the dimension

$$D_0 = \frac{\log N}{\log(1/r_{\min})}$$

is reached. This is the smallest embedding dimension possible. It would appear that the Hartley entropy would require uniform probability which, in turn, would imply uniform length. However, the Hartley entropy results from taking the limit $\alpha \to 0$ and not from the assumption of uniform probability. Hence, D_0 cannot be associated with the Hausdorff dimension without imposing the condition of uniform length and probability.

Introducing the arithmetic–geometric inequality, $\sum_{i=1}^{N} p_i^2 > \prod_{i=1}^{N} p_i^{p_i}$, unless all the p_i's are equal, into the numerator of the information dimension (26) would likewise necessitate the use of the inequality $\sum_{i=1}^{N} p_i r_i > \prod_{i=1}^{N} r_i^{p_i}$ in the denominator. There would then result

$$D_2 = \frac{\log\sum_{i=1}^{N} p_i^2}{\log(\sum_{i=1}^{N} p_i r_i)} \tag{27}$$

which is known as the 'correlation' dimension (Grassberger and Procaccia 1983). However, this is not a true dimension because the measure of (largest) volume of order $\tau = 1$ is $\overline{r^{D_H}} \neq \bar{r}^{D_H}$. Furthermore, the subscript 2 has nothing to do with the value of the characteristic exponent α which must lie in the open interval $(0, 1)$.

Thus we have proved the following.

Theorem 2. With $\alpha = (1+\tau)^{-1}$ and $\alpha \in (0, 1)$, the negative of the logarithm of the average volume of order τ cannot be less than the Rényi entropy. The order τ determines the dimensionality of the multifractal, or attractor, and it cannot be greater than the Hausdorff dimension in which it is embedded. Equality occurs only in the limit of uniform probabilities and lengths.

Certainly, there are probabilistic overtones in so far as the various means that occur at different orders, corresponding to different entropies, should correspond to the most probable value of the quantity measured (Lavenda 1998).

5. Extreme value distributions for wordlengths and fractal volumes

Consider a stationary sequence of wordlengths $\{K_N; N > 2\}$. We want to determine the distribution of the largest value $\check{K} = \max_{1\leqslant i\leqslant N} K_i$ for large values of N. We can then derive the distribution for the smallest wordlength by the symmetry principle for extreme value distributions (Gumbel 1958).

The sequence K_i constitutes a set of independent and identically distributed (i.i.d.) random variables. The probability for any wordlength K_i being greater than κ is exponential

$$\Pr(K_i > \kappa) = \bar{F}(\kappa) = d^{-\kappa} \tag{28}$$

where $\bar{F}(\kappa) = 1 - F(\kappa)$ is the tail of the distribution $F(\kappa)$. The probability that the largest value $\check{K}$ will not exceed κ is therefore

$$\Pr(\check{K} \leqslant \kappa) = F^N(\kappa) = [1 - d^{-\kappa}]^N. \tag{29}$$

Because of the i.i.d. property, the probability that the largest value will not exceed κ is the product of each of the individual probabilities of not exceeding κ. Using a method developed by Cramér (Gumbel 1958), we introduce the mode of the largest value $\tilde{\kappa} = \log N$, which is also the characteristic largest value (Gumbel 1958). Dividing and multiplying the last term in (29) by N gives

$$\Pr(\check{K} \leqslant \kappa) = F^N(\kappa) = \left[1 - \frac{d^{-(\kappa-\tilde{\kappa})}}{N}\right]^N.$$

The probability distribution $\check{G}(\kappa)$ becomes

$$\check{G}(\kappa) = d^{-d^{-(\kappa-\tilde{\kappa})}} \tag{30}$$

for large N. This asymptotic distribution is known as the double exponential distribution. On the strength of the symmetry principle (Gumbel 1958), relating the largest and smallest values, we may immediately write down the distribution $\hat{G}(\kappa)$ for the smallest wordlength as

$$\hat{G}(\kappa) = 1 - d^{d^{(\kappa-\tilde{\kappa})}}. \tag{31}$$

The double exponential distributions (30) and (31) have support on the open interval $(-\infty, \infty)$. However, a necessary and sufficient condition for the existence of a uniform code with a constant wordlength κ is $\kappa \geqslant \log N$, which is easily derived from the Kraft inequality (14). The same type of condition applies to the testing of materials where the characteristic largest value $\log N$ would be analogous to the lower bound on the measured strength of a material. This condition would make the Weibull distribution for the smallest value the only acceptable asymptotic distribution for the smallest value. However, it is often preferable to use the double exponential distribution for the smallest value (30) if the location parameter $\log N$ is large enough since the probability of a value below the *a priori* lower bound is negligible (Leadbetter *et al* 1983).

Turning to multifractals, consider a sequence of i.i.d. random radii $\{R_i; N \geqslant 2\}$. Their common initial probability $\Pr(R_i \leqslant r) = F(r) = r^D/v$, where v is the total fractal volume, and $D > 0$, but otherwise arbitrary. We are interested in the distribution of the smallest radius $\hat{R} = \min_{1\leqslant i\leqslant N} R_i$. The probability that the smallest radius will exceed r is the product of the probabilities of each individual value exceeding r or

$$\Pr(\hat{R} > r) = \bar{F}^N(r) = \left[1 - \frac{r^D}{v}\right]^N .$$

Writing $v = \bar{v}N$, the asymptotic probability, $\hat{G}(r)$, for the smallest radius becomes for large N

$$\hat{G}(r) = 1 - d^{-(r^D/\bar{v})}. \tag{32}$$

This is the distribution for nearest neighbour, and it belongs to the the class of Weibull distributions for the smallest value.

Acknowledgments

BHL was sponsored, in part, by MURST (60%) and CNR.

References

Abarbanel H D I 1996 *Analysis of Observed Chaotic Data* (New York: Springer)

Aczél J and Daróczy Z 1975 *On Measures of Information and Their Characterizations* (New York: Academic)

Campbell L L 1965 *Information and Control* **8** 423–9

Feinstein A 1958 *Foundations of Information Theory* (New York: McGraw-Hill).

Grassberger P and Procaccia I 1983 *Phys. Rev. Lett.* **50** 346–8

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

Halsey T C, Jensen M H, Kadanoff L P, Procaccia I and Shraiman B I 1986 *Phys. Rev.* A **33** 1141–51

Hardy G H, Littlewood J E and Pólya G 1952 *Inequalities* 2nd edn (Cambridge: Cambridge University Press)

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

——1995 *Thermodynamics of Extremes* (Chichester: Albion)

——1998 The relation between measures of information and error laws, submitted

Leadbetter M R, Lindgren G and Rootzén H 1982 *Extremes and Related Properties of Random Sequences and Processes* (New York: Springer)

Marshall A W and Olkin I 1979 *Inequalities: Theory of Majorization and its Applications* (New York: Academic)

Schroeder M 1991 *Fractals, Chaos, Power Laws* (New York: Freeman)

Shannon C E and Weaver W 1959 *The Mathematical Theory of Communication* (Urbana, IL: University of Illinois Press)