ESTIMATING THE HAUSDORFF DIMENSION

by

Russell Lynn Reeve

Dissertation submitted to the Faculty of the
Virginia Polytechnic Institute and State University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Statistics

APPROVED:

G. R. Terrell, Chairman

I. J. Good

M. V. Day

M. R. Reynolds

D. R. Jensen

E. P. Smith

July, 1990

Blacksburg, Virginia

ESTIMATING THE HAUSDORFF DIMENSION

by

Russell Lynn Reeve
Committee Chairman: George R. Terrell
Statistics

(ABSTRACT)

The use of fractals in fields such as molecular biology, epidemiology, landscape, ecology, geology, physics, etc., is becoming more common. In order to use fractals to model many phenomena, the researcher requires the knowledge of the fractal, or Hausdorff-Besicovitch, dimension. However, no statistical properties of the usual estimator, the entropy estimator, are known. In addition, the entropy estimator is biased high when an inefficient net is used.

This dissertation develops a new estimator, the relative entropy estimator, which is asymptotically unbiased and is consistent. The estimator is asymptotically normal, and asymptotic confidence intervals are presented. An estimate of the variance of the estimator is given which does not depend on the dimension, or its estimate, using an occupancy model. The exact distribution of the estimator is also derived.

Applications of the theory to various fields are presented. For example, I find that from the point of view of dimension, the logarithms of stock prices behave consistently with the classical Brownian function. Also, the relative entropy estimator gives a more realistic estimate of the dimension of surface terrain than an ad hoc estimate found in the literature. The Hausdorff dimensions of nursery-grown tree roots were estimated, and it was found that the dimension is related to the probability of the tree's survival when the tree is planted in the wild. The dimensions of Julia sets and of the Hénon attractor were also investigated.

A computer program for calculating the estimates is included.

To my mother and father

ACKNOWLEDGEMENTS

The author would like to express his sincere appreciation and gratitude to:

Dr. George Terrell, his advisor, for inspiring this research and for his encouragement and support; his constant guidance has been invaluable.

Dr. Donald Jensen, for his willingness to discuss issues and problems related to this research as well as other topics.

Dr. Eric Smith, for contributing his expertise to this research, his many useful comments and for his collaboration on the paper on neutrality of traits.

Dr. Raymond Myers, for his guidance as a friend and teacher.

Drs. Martin Day, I. J. Good, and Marion Reynolds for several useful conversations.

Finally, the author owes special thanks to his wife, Vanita, for her encouragement, patience and help in preparing this manuscript.

TABLE OF CONTENTS

Chapter 1
Introduction

1.1 Background on Hausdorff dimension

The author will first discuss what the Hausdorff dimension is, and what it tries to accomplish. The Hausdorff dimension is also known as the fractal dimension, since it is of interest primarily for fractals. The term fractal is loosely defined, but the usual definition is that it is a set with noninteger dimension (Mandelbrot 1977, p. 295). For non-fractals, the fractal dimension coincides with the topological dimension, and so adds no new information. Most of the background information that will be presented can be found in Benoit B. Mandelbrot's book (1977). The latest edition of the book is Mandelbrot (1982).

Before proceeding, some notation must be defined. The limit infimum will be denoted by $\underline{\lim}$. The limit superior will be denoted by $\overline{\lim}$.

Hausdorff dimensions are intimately related to Hausdorff measures (but neither have anything to do with Hausdorff spaces.) From Rogers (1970) the Hausdorff measure of a set A is defined to be

$$h_d(A) = \lim_{\epsilon \to 0} \inf_{\{C_i\}} \{ \sum (\text{diam } C_i)^d : \bigcup C_i \supseteq A, \text{ diam } C_i \leq \epsilon \}$$

where diam C_i is the diameter of C_i. For a quick review of Hausdorff measures, consult Billingsley (1986), pp. 247-258. Note that Billingsley (1986) normalizes the Hausdorff measure to look like a Lebesgue measure (using volumes of balls).

Suppose $A \subseteq \mathbf{R}^E$, E-dimensional Euclidean space. Then if d=E, $h_d(\cdot)$ is just a scaled Lebesgue measure. So, one can think of this as generalizing the Lebesgue measure. We will be looking at the situation where d is not necessarily equal to E.

To show that $h_E(\cdot)$ is a scaled version of Lebesgue measure, we merely note that $h_E(\cdot)$ is translation invariant, since

$$h_E(A+x) = \lim_{\epsilon \to 0} \inf_{\{C_i+x\}} \{\sum (\text{diam } (C_i+x))^E : \bigcup (C_i+x) \supseteq A+x, \text{ diam } C_i \leq \epsilon\} = h_E(A).$$

And therefore, $h_E(\cdot) = c\lambda(\cdot)$ for some $c > 0$ and Lebesgue measure $\lambda(\cdot)$, by Billingsley (1986), p. 180, problem 12.1. This is not the entire proof, but is the gist of the argument showing that $h_d(\cdot) = c\lambda(\cdot)$. Also, $h_0(\cdot)$ is counting measure.

The theory of dimensions is classical. See, for instance, Ryszard (1978) and Hurewicz (1948). The Hausdorff dimension is just one of many possible definitions of dimension. Informally, the Hausdorff dimension is the number d so that

$$0 < h_d(A) < \infty, \ A \neq \emptyset.$$

It is not hard to show that d is unique (Billingsley (1986), p. 258, problem 19.11). The dimension of A, dim A, is more formally defined as

$$\dim A = \inf \{d : h_d(A)=0\} = \sup\{d : h_d(A)=\infty\}.$$

Now, if $A \subseteq B$, then $\dim A \leq \dim B$. Also, if $A = \bigcup_{i=1}^{\infty} A_i$, then $\dim A = \sup_i \dim A_i$. For the proof, see Billingsley (1960). Also, see Billingsley (1961). Note, however, that the Hausdorff dimension is not necessarily an integer. The Cantor ternary set, for instance, has dimension $\log 2/\log 3 = 0.63093\cdots$; cf. Mandelbrot (1977, 1982), Chorin (1982a). Mandelbrot defines a fractal as a set whose dimension is noninteger. Applications of the Hausdorff dimension will be discussed later.

Another dimension, the entropy (sometimes mistakenly called capacity) dimension, is also in common usage. The entropy dimension was first proposed by Kolmogorov and Tihomorov (1959-1961) and is given by

$$\dim_E A = \underline{\lim}_{\epsilon \to 0} \log \tilde{N}_\epsilon(A) \Big/ \log(1/\epsilon)$$

where $\tilde{N}_\epsilon(A)$ is the number of balls of radius ϵ needed to cover A. Note that they use *radius* in their definition, whereas most of the literature uses ϵ for the *diameter* of the covering balls.

Kolmogorov and Tihomorov also propose another dimension, called the capacity dimension, which they define as

$$\dim_C A = \underline{\lim}_{\epsilon \to 0} \log M_\epsilon(A) \Big/ \log(1/\epsilon)$$

where M_ϵ is the largest number of points so that their mutual distances exceed 2ϵ. This is equivalent to packing in as many spheres as possible. On the line, $M_\epsilon = \tilde{N}_\epsilon$. We will define $N_\epsilon(A)$ to be the number of balls of *diameter* ϵ necessary to cover A, and redefine the entropy dimension as

$$\dim_E A = \underline{\lim}_{\epsilon \to 0} \log N_\epsilon(A) \Big/ \log(1/\epsilon).$$

Then it is easy to see that $\dim A \leq \dim_E A$. To show this, let $\dim_E A < \infty$, take $d > \dim_E A$, and define $h_d^*(A)$ by

$$h_d^*(A) = \underline{\lim}_{\epsilon \to 0} \inf_{\{C_i\}} \{\sum (\operatorname{diam} C_i)^d : \bigcup C_i \supseteq A, \operatorname{diam} C_i = \epsilon\}.$$

Then clearly $h_d(A) \leq h_d^*(A) = \underline{\lim}_{\epsilon \to 0} N_\epsilon(A)\epsilon^d$. For ϵ small, $N_\epsilon(A)\epsilon^d = c_\epsilon$. So,

$$\frac{\log c_\epsilon}{\log \epsilon} = d - \frac{\log N_\epsilon(A)}{\log \epsilon} \to d - \dim_E A > 0.$$

Therefore,

$$c_\epsilon < \epsilon^{d - \dim_E A} \to 0.$$

Hence, we have $h_d^*(A) = 0$, which implies $h_d(A) = 0$. Therefore, $h_d(A)$ is finite, and it may be possible to reduce d further and still maintain the finiteness of $h_d(A)$. Hence we conclude that

$$\dim A \le \dim_E A.$$

As you can see from the above discussion, it is very likely that $\dim A = \dim_E A$ for some A. In fact, it is known that for many sets the two dimensions do coincide. However, it is also known that the two dimensions are different. Farmer, Ott, and Yorke (1983) (footnote of p. 158) give an example of a set where the two dimensions are different. For the most part, the entropy dimension is used to estimate the Hausdorff dimension.

A set is said to be self-similar if it can be decomposed into parts which are each a scaled version of the entire set (rotations are allowed.) Many sets exhibit this characteristic, at least stochastically. For example, the interior of a square is self-similar because each quarter looks exactly like the entire set. A stochastic set is (stochastically) self-similar if it can be decomposed into parts, each of which has the same distribution as the full set (except for scaling.) An example of this sort is Brownian motion; *e.g.*, a function B(t) so that

$$B(t) \sim N(0, t\sigma^2).$$

This is self-similar since $\mathcal{L}(B(rt))=N(0, rt\sigma^2)=\mathcal{L}(\sqrt{r}B(t))$. ($\mathcal{L}(X)$ is the law of X.)

The similarity dimension uses this concept of self-similar sets to construct a dimension very similar to the entropy dimension, $\dim_E A$. It is best described using an example. Take a line segment, and break it into thirds (Fig. 1.1). Replace the middle segment with two equal length segments to produce Fig. 1.2. Note that the total length is now 4/3 the original length. Repeat this procedure for each of the 4 segments in the object to get Fig. 1.3. Continue in this manner to the limit. The curve obtained is known as the Koch curve; it has infinite length. It's self-similarity dimension is

$$\dim_S A = \frac{\log 4}{\log 3} = 1.2619\cdots.$$

Notice that every time you reduce the segments by a third, you get 4 times as many segments. Using the notation introduced in the discussion of entropy dimension,

$$N_\epsilon = 4^l$$
$$\epsilon = (1/3)^l,$$

for l=1, 2, $\cdots$.

Using the entropy dimension idea,

$$\dim_S A = \frac{\log N_\epsilon}{\log(1/\epsilon)} = \frac{\log 4}{\log 3}.$$

The self-similarity dimension is very hard to develop carefully, and it is not worth the effort for the purposes of this paper. The best way to understand it is to look at several examples. See Mandelbrot (1977, 1982), for instance. (Mandelbrot never does formally define self-similarity dimension, either.) Mandelbrot states that for self-similar sets, dim A = $\dim_E$ A = $\dim_S$ A. The self-similarity dimension really rests on a scaling argument. Using the same argument as for the Koch curve, you can show that the self-similarity dimension of the Cantor ternary set is log 2/log 3.

Another measure of dimensionality is Lyapunov numbers; these are frequently seen in the chaos literature. Consider a p-dimensional (in the sense of linear algebra) map F, where

$$x_{n+1}=F(x_n), \qquad x_n, x_{n+1} \in \mathbb{R}^p.$$

Let $J(x)$ be the Jacobian of F, $J(x)=(\partial F/\partial x)$. Define

$$J_n=J(x_n)J(x_{n-1})\cdots J(x_2)J(x_1)$$

and let $j_1(n) \geq j_2(n) \geq \cdots \geq j_{p-1}(n) \geq j_p(n)$ be the magnitudes of the eigenvalues of J_n. Define the Lyapunov numbers by

$$\lambda_i=\lim_{n\to\infty}\left[j_i(n)\right]^{1/n}.$$

The Lyapunov exponent is $\log \lambda_i$. The Lyapunov dimension is then defined as

$$\dim_L A=1+\frac{\log(\lambda_1\cdots\lambda_k)}{\log(1/\lambda_{k+1})}$$

where k is the largest value for which $\lambda_1\lambda_2\cdots\lambda_k \geq 1$. If $\lambda_1 < 1$, define $\dim_L A=0$. The Lyapunov dimension has been shown to be a lower bound on the fractal dimension (Kaplan and Yorke 1978). The Lyapunov dimension is often used with strange attractors. A strange attractor is a set that $F(x_n)$ converges to, but consists of more than just a single, or fixed, point. See the applications of the theory to strange attractors in Chapter 3 for more details. From Farmer, Ott, and York (1983), we get Table 1.1.

Table 1.1 lists the names of the most commonly used dimensions, and their generic names. The fractal dimensions are a measure of how large a set is, and are most commonly applied to fractals. The dimensions of the natural measure are more common in nonlinear dynamics. These dimensions weight a part of the set of interest by the frequency with which that part is occupied.

Lyapunov dimensions are used to determine the amount of chaos and mixing in a strange attractor. They appear to lead to more efficient estimators of the dimension of a strange attractor than the classical ones, possibly because of the semimetric properties of the Lyapunov dimension versus the purely topological properties of the other definitions, such as the entropy and capacity dimensions.

Notice in Table 1.1 the absence of the ϵ-capacity estimator of Kolmogorov and Tihomorov. In fact, the author has not seen it used anywhere. It is mentioned here only to try to clear up the nomenclature confusion.

Towards the end of Farmer, et al. (1983) many methods of computing the Hausdorff dimension are presented; one method uses the entropy dimension, another the Lyapunov dimension. Note that they mistakenly call the entropy dimension the capacity dimension, and this error has continued into more recent literature; *e.g.*, Saupe (1987). For more information about dimensions, (especially those that haven't been mentioned in this discussion,) see Farmer, et al. (1983).

If the set is scaled to fit into the unit cube (but no further), the Hausdorff dimension is a measure of its size. (This interpretation breaks down for sets that have fewer than E points, E being the dimension of the space the set in embedded in. But such sets are uninteresting, and it will be assumed that all sets encountered have more than E points.) For example, the Cantor set (dimension=0.63) is smaller than a line segment (dimension=1) but larger than a countable collection of points (dimension=0). Another interpretation of the Hausdorff dimension is the roughness of the set (within its topological dimension). Consider a Brownian function (motion); *e.g.*, a random function $B_H(t)$ so that $B_H(t)$ is a Gaussian process with stationary increments and variance

$$E[B_H(t_2) - B_H(t_1)]^2 \sim \sigma^2 \mid t_2 - t_1 \mid^{2H}. \tag{1.1a}$$

For H=1/2 you have the classical Brownian function with independent increments. Now,

(Mandelbrot 1977)

$$\mathrm{Cov}(B_H(s),\ B_H(t+s)-B_H(s)) = \frac{\sigma^2}{2}\left[(t+s)^{2H} - s^{2H} - t^{2H}\right]. \tag{1.1b}$$

Notice that for $H < 1/2$ we get a negative covariance, making for a particularly bumpy function. For $H > 1/2$ we get a positive covariance, making the function smoother. So, as $H \downarrow 0$, $B_H(t)$ becomes bumpier. Also, the Hausdorff dimension of the graph of $B_H(t)$ is $d = 2 - H$ a.s. (Orey 1970). Notice that for $H = 1/2$, (ordinary Brownian motion), $d = 3/2$. Incidentally, the dimension of the level set of $B_H(t)$ (i.e., $\{t \mid B_H(t) = 0\}$) is $d = 1 - H$ a.s. (Marcus 1976). So the dimension can also be interpreted as a measure of bumpiness (within a fixed topological dimension). For a more rigorous discussion, the reader may consult Mandelbrot and Van Ness (1968), Mandelbrot (1971), Voss (1988), and Saupe (1988).

1.2. Paradox

The author will now discuss an apparent paradox. Consider the set of rational numbers in the unit interval [0, 1]. The rational numbers have dimension 0, being countable. However, every method of calculating its dimension, including the one the author will propose in Chapter 2, gives a value of 1. How can this paradox be reconciled?

For practical purposes, the rational numbers are equivalent to the interval [0, 1], since one cannot distinguish the rational numbers from [0, 1] by visualizing the set. This suggests that, in practice, one should estimate the dimension of the closure of a set rather than that of a set itself. The closure of a set is what we think of when we picture a set.

Further, given a sample of points, there is no empirical method of discerning whether these observations come from the rational numbers or from the real numbers. One then assumes that the observations come from the real numbers. This is an assumption made not only in the estimation of dimension, but in all applied mathematical work. In fact, any observation we make about the physical world must come from the rational numbers, not from the real numbers. But to make our models, we assume these observations come from the closure of the rational numbers, i.e., the real numbers.

Using the closure of a set does not present any difficulties. Sets of interest in the physical world are closed sets, and so the closure is the set itself. For instance, the Cantor ternary set is closed; trees and turbulent flows can be modeled with closed sets. Open sets do not present a problem either. Adding the limit points to an open set does not change its dimension since an open set has dimension equal to the dimension of the space in which it is embedded. Mandelbrot (1977) considers the *complements* of open sets, the complements being closed.

1.3. Literature

We will now consider estimators (in the statistical sense) of the Hausdorff dimension. The estimators we will look at will be entropy-type estimators.

First, the author will make a point about terminology. To **calculate** the dimension, we compute an analytic result for the true dimension of the set. This has been done only for some special sets (mainly self-similar sets). The (ternary) Cantor set is one of the sets whose dimension has been calculated. To **approximate** the dimension, we approximate the set with another set whose dimension has been calculated. This is chiefly what has been done in the past, usually using computer simulations. To **estimate** the dimension, we obtain a statistical estimate of the dimension using data. The literature is not too precise on terminology.

Farmer, et al. (1983) discuss one of the most common estimators of the Hausdorff dimension, the common entropy estimator. Let N be the number of cubes with a data point in it, and ϵ the length of a side of the cubes. The estimator $\hat{\mathrm{d}}$ is then given by

$$\hat{\mathrm{d}} = \frac{\log \mathrm{N}}{\log (1/\epsilon)}.$$

No properties of $\hat{d}$ are discussed, nor even hinted at. To show how much remains to be done, Chorin (1982a) uses the same technique, and Clarke (1986) uses a small variant on the method, substituting areas of prisms for number of cubes. Still, no properties are mentioned, though he does caution against using the estimate if the sample size is too small. Saupe (1987) uses entropy dimension, but modifies it to speed up convergence. He comes up with a simplification of what will be proposed. Again, no properties are looked at.

An investigation into methods of calculating dimension and their properties is needed, as very little has been done so far. Some notable work in estimating dimension was performed by Cutler and Dawson (1989). Cutler and Dawson look at distributions that are regular: Let $B_\epsilon(x)$ be a ball of radius ϵ centered at x. Then a set S is regular if we have the relationship

$$\underline{\lim}\, \frac{h_d(S \cap B_\epsilon(x))}{(2\epsilon)^d} = \overline{\lim}\, \frac{h_d(S \cap B_\epsilon(x))}{(2\epsilon)^d} = \pi^{d/2} 2^{-2} / \Gamma(1 + d/2)$$

where all limits are taken with $\epsilon \rightarrow 0^+$. Under this assumption, the dimension of a set must be an integer greater than 0 (Cutler and Dawson 1989, Theorem 2.3). The authors considered the reciprocal of the dimension and estimate it with the estimator $l_n(x) = \log_e 2\rho_n(x) \,/\, \log 1/n$, where $\rho_n(x) = \min_{1 \leq i \leq n} \| X_i - x \|$, $\{X_i\}_{i=1}^{\infty}$ being i.i.d. They show this estimator to converge to the reciprocal of the dimension with probability one, and asymptotically a linear transformation of $l_n(x)$ has an extreme value distribution. To get confidence intervals and (equivalently) tests of hypotheses, the authors compute the average $\bar{l} = k^{-1} \sum_{i=1}^{k} l_i$, and appeal to the Central Limit Theorem to obtain asymptotic normality. Hypothesis tests are then straightforward Z-tests.

Some comparisons of the work of Cutler and Dawson (1989) with the work found in Chapter 2 of this paper are in order. Cutler and Dawson's work precludes the estimating of dimensions that are fractions, such as estimating the dimension of the Cantor set, for example. The estimator of Chapter 2 of this paper, however, allows for the estimation of fractional dimensions. Both papers eventually appeal to Central Limit Theorem results, and so use normal-theory confidence statements. In both papers, it is seen that the precision of the estimate decreases as the dimension increases (check the variance equations in Chapter 2; see also Cutler and Dawson 1989, p. 144). Both methods are computationally expensive. Cutler and Dawson's approach requires the computation of many norms from many different points in space. The approach in this paper requires the construction of a tree data structure which must be completely traversed. A comparison of computation times of these two procedures would be interesting. Both procedures require an immense amount of data to get good estimates. (However, the same can be said for certain problems in spectral estimation (Ramsey 1989).) For example, in Cutler and Dawson's (1989) Case 1 (simulation study), using samples of size $t = 1890$ resulted in quite poor results, with 34% of the confidence intervals generated covering at least 2 integers.

Another paper to look at the statistical properties of estimating the Hausdorff dimension was Denker and Keller (1986). The authors use U-statistics to test hypotheses and to estimate quantities of interest. Denker and Keller use the U-statistics to test features of dynamical systems, such as testing the symmetry of an invariant distribution. To estimate the dimension, Denker and Keller look at a quantity called the correlation exponent. They state that for higher-dimensional systems, the correlation exponent agrees with the fractal dimension to be studied. For a discussion of the correlation expoenent, see Denker and Keller (1986), p. 70. Using the theory of U-statistics, they construct an estimator and find its standard error.

Some applications of the Hausdorff dimension to the physical sciences will be presented to illustrate the uses of this concept and to motivate research into its statistical properties. Gardner, et al. (1987) looked at how human interference affects landscape patterns at different scales. An important measure of the clustering of the trees was the fractal dimension of the tree cluster boundaries. Along similar lines, Krummel, et al. (1987) investigated the effects of human farming activity in the Natchez Quadrangle on the dimensions of the boundaries of tree clusters. They found different dimensions for large clusters than for small clusters, indicating a difference in effect across scales.

Morse, et al. (1985) looked at the fractal dimension of the surface of vegetation to get some predictions of the relative densities of arthropods living on the plants. They predicted that small arthropods would have a higher density than large arthropods because of the increased surface area available to them.

Lewis and Rees (1985) looked at how the dimension of a segment of protein affects its function. Allen, et al. (1982) looked at the relationship between dimension and protein conformation. In this paper, the authors presented some estimates (with sampling errors) for several paramagnetic and diamagnetic proteins. Schaffer and Kot (1985a) looked at epidemic data of measles for New York and Baltimore, attempting to explain what appears to be a noisy time series as a deterministic system with little noise. Some preliminary calculations determined that the dimension of the epidemic in state space was less than three. Schaffer and Kot (1985a) thought it looked 2-dimensional. Further plotting of the data using Poincaré sections suggested that the data were approximately 1-dimensional.

Schaffer (1984) considered returns of lynx trapping in Canada. Here, as in the previous paper, he concluded that the data lie essentially on a 1-dimensional attractor (in state space), with some noise. Noticing how closely the dimension of the lynx series is to the dimension of the measles epidemic series, he went on to say: "Why this should be the case, and whether or not this property is shared by other ecological systems are interesting questions." However, his analysis was mostly heuristic and graphical. It would be nice to have some numerical evidence to buttress his argument.

Goldberger, et al. (1984) discussed various arrhythmias of the heart, and exhorted their colleagues to look at some of the nonlinearities in the heart system. Possibly by looking at the dimension, one could detect various problems, and determine how to get the heart back to its usual state. From the description in the article, it appears that higher dimensional EEG signals correspond to a sicker heart.

Schaffer and Kot (1985b) argued convincingly that there is good reason to believe that ecosystems lie on strange attractors. And one measure of "strangeness" of an attractor is its dimension. If its dimension is nearly integer valued, it is not very strange and linear (possibly stochastic) equations should model the system of interest adequately.

Mandelbrot (1982) argued that the dimension of a tree should be looked at. He went on to hypothesize that vines should have low dimensions, while trees with a lot of foliage should have higher dimensions. The dimension should be related to a plant's function in the ecosystem.

As an aside, this might be one way of quantifying variability and how the environment affects variability. Lewontin (1974) says that organisms living on the edge of their ecosystem tend to be less variable than those in the interior. He gives some evidence involving flies supporting this statement. Would this statement apply to the dimension of trees as well? Maybe trees in near desert conditions have much less variability in dimension than trees in better growing regions, both within and between species. The author suspects that trees in arid regions have lower dimension, as well as less variability in dimension. (Cacti seem to be particularly low-dimensional, ignoring the thorns.)

It is believed that evolving systems tend to get more complex over time; *e.g.*, the number of species at a given time seems to grow approximately linearly. Do trees exhibit this same kind of evolution? Is this true of biological systems in general? One measure of complexity is the dimension. It is reasonable to suppose that the growth in complexity of trees in drier areas would be slowed by the hostile environment. Is there a difference in complexity? The author speculates that answering these questions would help to resolve some of the fundamental issues of species fitness.

Mandelbrot (1982) also discussed lung air pipes and the cardiovascular system, which appear to be space filling fractals, not unlike polymers. He stated that the dimension for the cardiovascular system should be between 2 and 3, but did not get more specific. However, the air pipes should have dimension $D \approx 3$ (Mandelbrot 1982, p. 157). This would be interesting to research, and it might reveal some of the similarities or differences between the two systems.

The Hausdorff dimension has also been used to solve problems in turbulence. It has been known that highly stretched vorticity collects itself into a volume much smaller than the volume available to it. Mandelbrot (1982) believes that the vorticity collects itself into a set of dimension $D \approx 2.5$. Chorin (1981) stated: "We find that the vorticity stretches unevenly . . . that [the vorticity] approximates a set with Hausdorff dimension $\sim$ 2.5. *From the point of view of applications, this may be our most important result,*" (italics added). He goes on to say (p. 860) "The Hausdorff dimension of the stretched vorticity set plays a major role in a numerical method used to solve problems involving inhomogeneous turbulence." A relation of interest here relates energy E to dimension D:

$$E(r)=\text{constant} \times r^{\beta}$$

where r is the scale ("structure function"), and β is the inertia range exponent. Chorin (1981, p. 857) presents the heuristically developed relation

$$\beta = [2+(3-D)]/3.$$

In Chorin (1982b), he again looks at a vortex and displays some pictures of it (pp. 526, 527).

It's clear that the β above would be interesting for the study of turbulence. Why isn't it calculated directly? Apparently, this is quite difficult. Hentschel and Procaccia (1982) stated "Attempts to calculate $[-\beta]$ on the basis of the Navier-Stokes equation, or even to prove $[-\beta] \neq 0$, have been elusive so far." The paper focused mainly on what D should be, without producing any data (computer simulations or otherwise.)

Voss, Laibowitz, and Allessundrini (1982) studied the boundaries of percolation clusters of gold, and found that the boundary of all clusters have dimension D=2, while individual boundaries have dimension $D \approx 1.9$. (They presented pictures of the clusters.) Along a similar line of investigation, Kapitalnik and Peutscher (1982), studying percolation clusters of lead, found that the dimension of an infinite cluster is $D=1.9 \pm 0.02$, while the backbone has dimension $D=1.65 \pm 0.05$.

Notice the similarity in the dimension between the clusters. Is this because both gold and lead are heavy metals? Would other metals have different dimensions? Passamante, Hediger, and Gollub (1989) presented an algorithm for estimating the dimension of a strange attractor with measurement error.

One may want to consult Halsey, Meakin, and Procaccia (1986) or Amitrano, Coniglio and diLiberto (1986) to see how the fractal dimension relates to growth models. Another paper that deals with this topic is Ott, Sauer, and York (1989).

Hausdorff dimension has also been used in the geophysical sciences. Rothrock and Thorndike (1980) found that the bottom of sea ice looks like a fractal. However, they smoothed the data (to eliminate measurement error), and wound up with a dimension of 2.

As you can see, the applications of Hausdorff dimension are potentially far-reaching, and may provide fresh insight into many perplexing problems. However, the state of the art has not advanced very much over the last 12 years. The statistical methods used then are still being used today, but little is known about any of their properties, and few have mentioned their statistical properties. However, Clarke (1986) did urge against the use of his estimator when the sample size is too "small". Clearly, some research along these lines is desirable. It would also be desirable to investigate the methodology of estimating the dimension. In the past, if an estimator didn't work well, another definition of dimension was put forward, and estimated in an obvious manner.

1.4. Goal

The goal of this dissertation is to investigate "good" methods of estimating the Hausdorff dimension of a set. In this dissertation, a competitor to the current estimator (the slope of the regression line regressing log N_ϵ on $\log(1/\epsilon)$) will be developed. The statistical properties of the estimator will also be developed. Finally, the author will apply the new technique to some processes that have appeared previously in the literature, and compare the results.

The rest of this dissertation is organized as follows: In Section 2.1, some preliminary and theoretical background is studied to determine what estimator to consider. Also, the properties of the estimator are studied when the set is known. In Section 2.2, the extrapolation method is selected. In Section 2.3, statistical properties of the estimator are studied. In Section 2.4, comparisons between the new estimator and the entropy estimator are made. Also, some properties of the entropy estimator are discussed. In Section 2.5, averaging over space is investigated to reduce variability. The reason for not averaging over the sample is also discussed. Section 2.6 establishes consistency of the new estimator. Chaper 3 is devoted to applications.

Table 1.1. **Various dimensions in common use (from Farmer, et al. (1983)). The table lists the dimensions in common use and their generic names.**

Name of dimension	Generic name
Entropy (alias capacity)	fractal
Hausdorff	
Information	dimension of the
$\mathfrak{I}$—capacity	natural measure
$\mathfrak{I}$—Hausdorff	
Pointwise	
Hausdorff-of-the-core	
Lyapunov	Lyapunov

Figure 1.1. The beginning of a Koch curve.

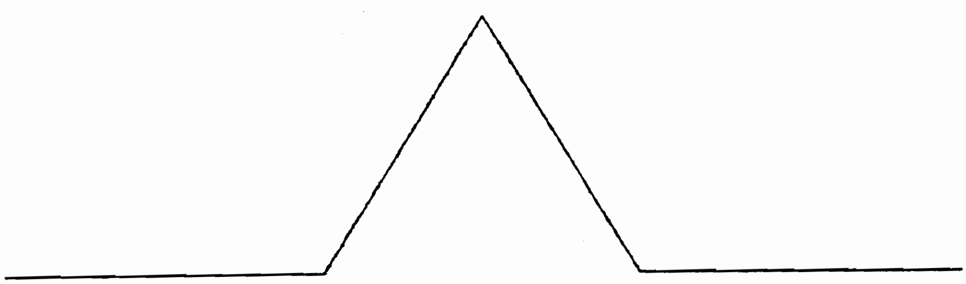

Figure 1.2. The curve after one step.

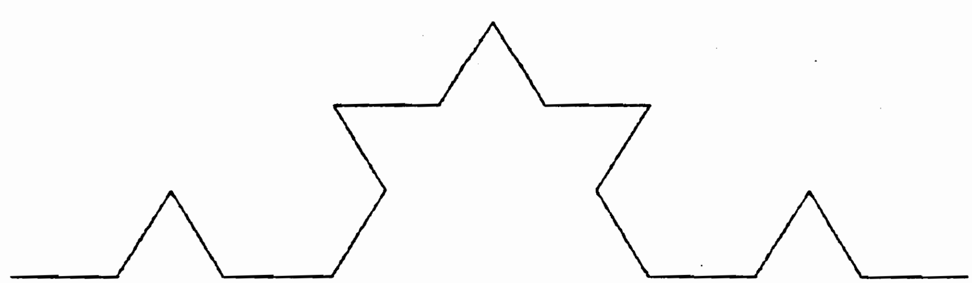

Figure 1.3. The curve after two steps. The Koch curve is obtained by taking this curve to the limit.

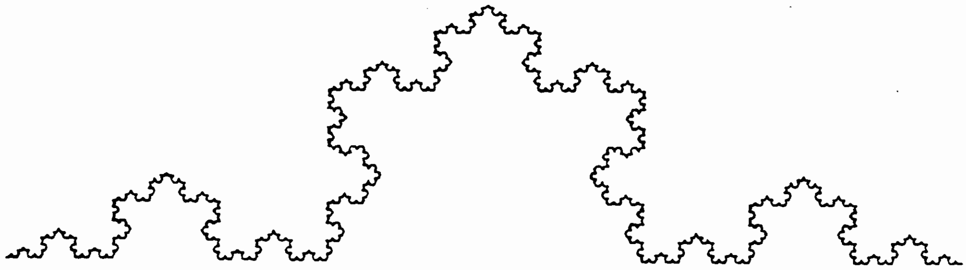

Figure 1.4. The Koch curve after many steps.

Chapter 2
Estimating the Dimension

2.1. Preliminaries

We will now investigate methods of calculating the Hausdorff dimension of compact sets, in $\mathbf{R}^E$. While it is not necessary for a set to be compact for its dimension to be calculated, sets in the real world almost always will be; and if the sets are stored in a computer, they always will be compact. We will make use of nets (see Definition 2.1), particularly the dyadic net because of its convenience with computers.

For a totally bounded set A, let $N_\epsilon(A)$ denote the smallest number of spheres (balls) of diameter ϵ needed to cover A. Define the **ϵ-entropy** of A to be

$$H_\epsilon(A)=\log_2 N_\epsilon(A).$$

(See Hawkes (1974). Also see Kolmogorov and Tihomorov (1961)). Note: Hawkes uses radius here, but changes the symbols a little so that the calculated number is the same. We shall need the following:

Definition 2.1. A **net** is a collection $\mathcal{C}$ of sets so that

(i) $\mathcal{C}$ is countable

(ii) $A, B \in \mathcal{C} \Rightarrow A \subseteq B$, $A \supseteq B$, or $A \cap B = \emptyset$

(iii) Given x and $\epsilon > 0$ $\exists$ $A \in \mathcal{C}$ so that $x \in A$ and diam $A \leq \epsilon$.

(iv) $\emptyset \in \mathcal{C}$.

Condition (iv) above is included so that we can generate a σ-field from the net. One might want to think of a net as a (graph theoretic) tree.

Definition 2.2. Let I be an open set. The **upper** and **lower relative entropy** with respect to

A is, respectively,

$$\overline{H}_0^A(I)=\overline{\lim_{\epsilon\to 0}}\left[H_\epsilon(A)-H_\epsilon(A\cap I)\right]$$

$$\underline{H}_0^A(I)=\underline{\lim}_{\epsilon\to 0}\left[H_\epsilon(A)-H_\epsilon(A\cap I)\right]$$

Definition 2.3. Let $\mathcal{C}$ be a class of sets. If $\overline{H}_0^A(C)-\underline{H}_0^A(C)$ is uniformly bounded as C varies over $\mathcal{C}$, then $\overline{H}_0^A$ is **$\mathcal{C}$-regular entropy.**

Note: From now on, all logarithms will be calculated to the base 2 unless specified otherwise.

By $\mathcal{U}_p$, we shall mean the net of p-adic cylinders (squares). So $\mathcal{U}_p$ is the net consisting of sets of the form $u_{n,k}=[p^{-kn}, p^{-(k+1)n})$, p an integer. The set $u_{n,k}$ (as k varies) is a set of order n. The dyadic net is the case p=2. A slight modification of a theorem by Hawkes (1974) is presented next.

Theorem 2.4. Let $A \subseteq \mathbf{R}^d$ be closed, and suppose it has $\mathcal{U}_p$-regular entropy. If $\lim_{\epsilon\to 0} N_\epsilon(A\cap D)/N_\epsilon(A) > 0$, or even $\underline{\lim}_{\epsilon\to 0} N_\epsilon(A\cap D)/N_\epsilon(A) > 0$, and

$$D \subseteq \{\, x \mid \lim_{n\to\infty}\left[\overline{H}_0^A(u_n(x))/n\right] = \alpha \log p \,\},$$

or even

$$D \subseteq \{\, x \mid \underline{\lim}_{n\to\infty}\left[\overline{H}_0^A(u_n(x))/n\right] = \alpha \log p \,\},$$

then

$$\dim D = \alpha,$$

where $u_n(x)$ is the unique set in $\mathcal{U}_p$ of order n that contains x.

For the proof of this theorem, we need to construct a measure $l(\cdot)$. This construction can be found in Hawkes (1974), Theorem 3.2, and is repeated here. Let $D \in \mathcal{U}_p$, and define $m_\epsilon(D) = N_\epsilon(A\cap D)/N_\epsilon(A)$. Since $\mathcal{U}_p$ is countable, a diagonal argument shows that we can find a sequence $\{\delta_k\}$ tending to zero such that

$$l(D) = \lim_{k \to \infty} m_{\delta_k}(D)$$

exists for each $D \in \mathcal{U}_p$. We can now extend $l(\cdot)$ to a Borel measure supported by A.

Proof. See Hawkes (1974), Theorem 3.3 and note that

$$\lim_{\epsilon \to 0} \frac{N_\epsilon(A \cap D)}{N_\epsilon(A)} \leq l(D) \tag{2.1}$$

□

The Cantor set, obtained by deleting s−r open intervals from each closed interval (of the r possible) and repeating to the limit, will be denoted by $C_{r,s}$. The usual Cantor ternary set has notation $C_{2,3}$. As an application of Theorem 2.4, it will be shown that dim $C_{r,s}$= log r/log s (Good 1941, p. 200, Eggleston 1949; for easier reading, see, for example Mandelbrot 1977, 1982; Chorin 1982a).

Lemma 2.5. Let $\mathcal{U}$ be the set of s-adic cylinders in $\mathbb{R}$ that contain some point of $C_{r,s}$. Then $C_{r,s}$ has $\mathcal{U}$-regular entropy.

Proof. Pick $u \in \mathcal{U}$. Suppose it is of order t. Then

$$H_\epsilon(C_{r,s}) - H_\epsilon(C_{r,s} \cap u) = \log(r^{k_\epsilon}) - \log(r^{k_\epsilon - t}),$$

where $2\epsilon \in [s^{-k_\epsilon}, s^{-(k_\epsilon - 1)})$ determines k_ϵ, and $k_\epsilon \geq t$. Hence,

$$\overline{H}_0^{C_{r,s}}(u) = \overline{\lim_{\epsilon \to 0}}\, t \log r = t \log r = \underline{\lim}_{\epsilon \to 0}\, t \log r = \underline{H}_0^{C_{r,s}}(u). \tag{2.2}$$

Hence the difference is zero. □

Lemma 2.6. Let l be the measure on $C_{r,s}$ constructed as in Hawkes (1974) Theorem 3.2.

Then $l(C_{r,s}) > 0$.

Proof. From Lemma 2.5, the D of Theorem 2.4 is $C_{r,s}$. Further, $A=C_{r,s}$. From Theorem 2.4, it is clear that $l(C_{r,s}) > 0$. □

Theorem 2.7. $\dim C_{r,s} = \log r / \log s$.

Proof. Let $x \in C_{r,s}$. Then, from equation (2.2),

$$\overline{H}_0^{C_{r,s}}(u_n(x)) \equiv \log r = \alpha \log s$$

for $\alpha = \log r / \log s$. We quote Lemmas 2.5 and 2.6 to fulfill the hypotheses of Theorem 2.4. Hence, $\dim C_{r,s}=\alpha$. □

If we assume the hypotheses of Hawkes (1974) Theorem 3.3 to hold, where we use that part of the dyadic net that contains A, then we can calculate dim A by using

$$\dim A = \lim_{n\to\infty} \overline{H}_0^A(u_n(x))/n$$

for any $x \in A$. So, let us calculate everything of interest with the dyadic net.

Lemma 2.8. Let $A \subseteq \mathbf{R}^d$ be a Borel set. Let k_ϵ be the unique integer so that $2^{-k_\epsilon} \le \epsilon/\sqrt{d} \le 2^{-(k_\epsilon-1)}$. Define $\eta_\epsilon(A)$ to be the number of sets in $\mathcal{U}_2$ of order k_ϵ needed to cover A. Then

$$N_\epsilon(A) \le \eta_\epsilon(A) \le 3^d 2^{d(d+1)} N_\epsilon(A).$$

Proof. Since the diameter of a cube of side 2^{-k_ϵ} is $2^{-k_\epsilon}\sqrt{d} \le \epsilon$, the left inequality is clear. The right inequality is shown in the proof of Rogers (1970) Theorem 49. □

Note: k_ϵ *as defined above will be used from now on.*

Consider $\hat{m}_\epsilon(I)=\eta_\epsilon(A\cap I)/\eta_\epsilon(A)$. From Lemma 2.8, we get

$$b_u^{-1}\frac{N_\epsilon(A\cap I)}{N_\epsilon(A)} \le \hat{m}_\epsilon(I) \le b_u\frac{N_\epsilon(A\cap I)}{N_\epsilon(A)},$$

where $b_u=3^d 2^{d(d+1)}$, or

$$b_u^{-1}m_\epsilon(I) \le \hat{m}_\epsilon(I) \le b_u m_\epsilon(I), \tag{2.3}$$

where $m_\epsilon(I)=N_\epsilon(A\cap I)/N_\epsilon(A)$. By the proof of Hawkes' Theorem 3.2, there exists $\{\delta_k\}_{k=1}^{\infty}$ so that

$$\lim_{k\to\infty} m_{\delta_k}(I)=l(I)$$

exists for all $I\in\mathcal{U}_p$. So define

$$\bar{\hat{l}}(I)=\overline{\lim_{k\to\infty}}\ \hat{m}_{\delta_k}(I)$$

$$\underline{\hat{l}}(I)=\underline{\lim}_{k\to\infty}\ \hat{m}_{\delta_k}(I).$$

Let $\hat{l}^*$ denote either $\bar{\hat{l}}(I)$ or $\underline{\hat{l}}(I)$. So, from equation (2.3),

$$b_u^{-1}l(I) \le \hat{l}^* \le b_u l(I). \tag{2.4}$$

Extending $\hat{l}^*$ to a Borel measure, it is clear that (2.4) holds for all Borel sets. So, we see that $\hat{l}^*$ is an equivalent measure to l.

Definition 2.9. The **ϵ-net entropy** is $Q_\epsilon(A)=\log\eta_\epsilon(A)$.

Letting $K = \log b_u = d(d+1) + d\log 3$, Lemma 2.8 gives

$$H_\epsilon(A) \le Q_\epsilon(A) \le H_\epsilon(A)+K. \tag{2.5}$$

Therefore,

$$H_\epsilon(A)-H_\epsilon(A\cap I)-K \le Q_\epsilon(A)-Q_\epsilon(A\cap I) \le H_\epsilon(A)-H_\epsilon(A\cap I)+K$$

Define

$$\overline{Q}_0^A(I)=\varlimsup_{\epsilon\to 0}\left[Q_\epsilon(A)-Q_\epsilon(A\cap I)\right].$$

Similarly, define $\underline{Q}_0^A(I)$ as the limit infimum. Then it is clear that

$$\overline{H}_0^A(I)-K \le \overline{Q}_0^A(I) \le \overline{Q}_0^A(I)+K$$

and

$$\underline{H}_0^A(I)-K \le \underline{Q}_0^A(I) \le \underline{Q}_0^A(I)+K.$$

This leads immediately to the conclusion:

Lemma 2.9. A has a $\mathfrak{C}$-regular entropy if and only if the difference $\overline{Q}_0^A(I)-\underline{Q}_0^A(I)$ is uniformly bounded as I varies over $\mathfrak{C}$.

Using Lemma 2.9, we can now calculate everything using the $\mathfrak{U}_2$ net.

Theorem 2.10. Let $A \subseteq \mathbb{R}^d$ be closed with $\mathfrak{U}_2$-regular entropy. Let $\hat{l}^*$ be constructed as above. Then if

$$D \subseteq \{\, x \mid \lim_{n\to\infty}\left[\overline{Q}_0^A(u_n(x))/n\right] = \alpha \log p \,\},$$

or even

$$D \subseteq \{\, x \mid \varliminf_{n\to\infty}\left[\overline{Q}_0^A(u_n(x))/n\right] = \alpha \log p \,\},$$

then we have dim D $= \alpha$, whenever D is a Borel set with $\hat{l}^*(D) > 0$.

Proof. By equivalence of measures, $\hat{l}^*(D) > 0 \Rightarrow l(D) > 0$. From above,

$$\frac{\overline{H}_0^A(I)-K}{n} \le \frac{\overline{Q}_0^A(u_n(x))}{n} \le \frac{\overline{H}_0^A(I)+K}{n}$$

Therefore, $\overline{Q}_0^A(u_n(x))$ has the same limit as $\overline{H}_0^A(u_n(x))$. Now we need only invoke Theorem 2.4. □

Corollary 2.11. Define $\underline{\Gamma}(D) = \underline{\lim}_{\epsilon\to 0}\ \hat{m}_\epsilon(I)$. Extend $\underline{\Gamma}$ to the Borel sets. If $\underline{\Gamma}(D) > 0$ and the other hypotheses of Theorem 2.10 hold, then again dim D = α.

Proof. $0 < \underline{\Gamma}(D) < \hat{\Gamma}^*(D)$. □

The author will now make a few comments on Theorem 2.10. The requirement that $\hat{\Gamma}^*(D) > 0$ is just to make sure the set doesn't get "too small". In most applications, D=A=$\{x : \lim_{n\to\infty} \left[\overline{Q}_0^A(u_n(x))/n\right] = \alpha\}$ and $\hat{\Gamma}^*(D) > 0$ will be satisfied automatically, assuming D isn't countable or otherwise "small".

We have shown above that everything we want to do can now be done using the $\mathcal{U}_2$ net. Of course, the $\mathcal{U}_2$ net is convenient for computer applications.

In our applications of Theorem 2.10, we will assume that $\hat{\Gamma}^*(D) > 0$ since this condition cannot be checked. Also, let D=A. Hence, for each $x \in A$,

$$\dim A = \alpha = \underline{\lim}_{n\to\infty}\ n^{-1}\ \overline{\lim}_{\epsilon\to 0}\left[Q_\epsilon(A) - Q_\epsilon(A \cap u_n(x))\right].$$

Let $\rho = 2^{-n}$. The equation now looks like

$$\alpha = \underline{\lim}_{\rho\to 0}\left[\log(1/\rho)\right]^{-1}\ \overline{\lim}_{\epsilon\to 0}\left[Q_\epsilon(A) - Q_\epsilon(A \cap u_{-\log(1/\rho)}(x))\right].$$

If we assume that $\underline{\lim}$ and $\overline{\lim}$ are ordinary limits, the equation above simplifies to

$$\alpha = \lim_{\rho\to 0}\left[\log(1/\rho)\right]^{-1}\ \lim_{\epsilon\to 0} K^*_{\rho,\epsilon}(A),$$

where $K^*_{\rho,\epsilon}(A) = \left[Q_\epsilon(A) - Q_\epsilon(A \cap u_{-\log(1/\rho)}(x))\right]$. Let

$$J(\rho)= \lim_{\epsilon\to 0} K^*_{\rho,\epsilon}(A)/\log(1/\rho).$$

For each ρ=1/2, 1/4, 1/8, $\cdots$, one can approximate $J(\rho)$ using some extrapolation procedure to calculate $\lim_{\rho\to 0} J(\rho)=\dim A=\alpha$.

Some details that have to be taken care of will now be discussed. Let x_j be the j-th coordinate of $x \in \mathbf{R}^E$. Then the j-th coordinate of $u_n(x)$ is a number m_j such that

$$2^{-n}m_j \le x_j < 2^{-n}(m_j+1),$$

or $m_j=\lfloor x_j 2^n \rfloor$. Let ϵ_0 be the smallest distance between points in the sample. Then the order of the cylinders must be no larger than $n_0=\lfloor -\log \epsilon_0 \rfloor$.

Let us explore the procedure outlined above to see some problems, and possibly find some ways of improving the algorithm.

Example 2.12a. Let A=[0, 1/3]$\cup$[2/3, 1]. Pick n, k so that $n < k$; $\epsilon=2^{-k}$. Then $\eta_\epsilon(A)=2(\lfloor 2^k/3 \rfloor+1)$. So,

$$Q_\epsilon(A) = 1 + \log\,(\lfloor 2^k/3 \rfloor+1) \approx 1 + \log(2^{k-\log 3})$$

and hence

$$Q_0^A(u_n(0)) \approx 1 + \log(\, 2^{k-\log 3}) - (k-n) = 1 - \log 3 + n$$
$$= n - 0.585 \quad \text{for all } \epsilon > 0.$$

For each n, our estimate of dim A is

$$\frac{\overline{Q}_0^A(u_n(x))}{n} = 1 - \frac{0.585}{n}$$

These values are listed in Table 2.1 below. This converges quite slowly. If we could take into account the remainder term $-0.585/n$, we would have a more accurate way of approximating dim A.

A more detailed analysis follows. Suppose $\eta_\epsilon(A)=f_k 2^k$, $(\epsilon=2^{-k})$, $\eta_\epsilon(A \cap u_n(x))=g^x_{kn}2^k$. Then

$$Q_0^A(u_n(x))=\log(f_k/g^x_{kn})$$

and

$$\dim A=\lim_{n\to\infty} n^{-1} \overline{\lim_{k\to\infty}} \log \frac{f_k}{g^x_{kn}}$$

Example 2.12b. Again, let A be as in Example 2.12a. Then $f_k \approx 2/3$, $g^0_{kn}=2^{-n}$. So we compute

$$\lim_{n\to\infty} n^{-1} \overline{\lim_{k\to\infty}} \log \frac{f_k}{g^x_{kn}} = \lim_{n\to\infty} n^{-1}[n+(1-\log 3)]= \lim_{n\to\infty}\left[1+\frac{1-\log 3}{n}\right].$$

Again, this converges, but of order $1/n$.

Example 2.13. Let $A=[0, 1]$. Then $f_k=1$, $g^0_{nk}=2^{-n}$. We have

$$n^{-n} \overline{\lim_{k\to\infty}} \log \frac{f_k}{g^0_{nk}} = n\ (\tfrac{1}{n}) = 1 \text{ for all n.}$$

In this situation, we get very quick convergence. The hole in Example 2.12 slows down the convergence.

To save time and ink, we shall denote $n^{-1}\lim_{k\to\infty}\log(f_k/g^x_{nk})$ by $J_x(n)$.

Example 2.12c. We could speed things up a bit if $g^x_{nk} \approx (2/3)2^{-n}$, for some x. Then our net entropy would act as in Example 2.13, and we would have instant convergence. However, the best we could hope for is some x near 1/3 or 2/3 chosen to make g^x_{nk} as near 2/3 as possible. Consider $x=1/3$. On average, 1/3 will be right in the middle of $u_n(1/3)$. Therefore, $\bar{g}^{1/3}_{nk}=1/2$, where the bar denotes average over n. So, using this point we have

$$J_{1/3}(n) = n^{-1}\overline{\lim_{k\to\infty}} \frac{2/3}{2^{-n}/2} = n^{-1}[n+(2-\log 3)] = 1+0.415/n.$$

We have over-corrected. The difference between 2.12b and 2.12c is that in 2.12c, we chose a point in the least dense part of A, whereas in 2.12b we chose a point in the most dense part.

Notice that Example 2.13 converged immediately because every point is as crowded as any other. This leads me to conclude that we could average several estimates, using points x in various places. If we averaged $J_0(n)$ and $J_{1/3}(n)$ we get

$$\bar{J}(n) = 1 - \frac{0.17}{n}.$$

For n=10, the error is now −0.017 instead of errors of −0.0585 or 0.0415 as before. Further, if $J_x(n)$ were to be calculated at more values of x, the averaging should reduce the error even more.

Now let's take a look at a general set, A. Define, for some $x \in A$, $A_n = u_n(x) \cap A$. Suppose

$$\eta_k(A) \equiv \eta_\epsilon(A) = f_k 2^{k\alpha}, \qquad \eta_k(A_n) = g_{nk} 2^{(k-n)\alpha}, \qquad (\epsilon = 2^{-k})$$

where $f_k \approx f$, $g_{nk} \approx (1+\delta) f_k$ for large k, n, and $\alpha = \dim A$. Then

$$\begin{aligned} H_0^A(u_n(x)) &= \log \eta_k(A) - \log \eta_k(A_n) \\ &\approx f + k\alpha - f - (1+\delta) - k\alpha + n\alpha, \end{aligned}$$

and hence

$$n^{-1} H_0^A(u_n(x)) \approx \alpha - \frac{1+\delta}{n} \tag{2.6}$$

For x in a dense part of A, $\delta > 0$, and for x in a sparse part of A, $\delta < 0$. So, $n^{-1} H_0^A(u_n(x))$ converges to α at a rate of $1/n$.

Equation (2.6) gives two suggestions for getting at α:

- Regress $\mathbf{H}_0^{\mathbf{A}}(\mathbf{u}_n(\mathbf{x}))$ onto $\mathbf{n}$, using **slope** as the estimate
- Regress $\mathbf{n}^{-1}\mathbf{H}_0^{\mathbf{A}}(\mathbf{u}_n(\mathbf{x}))$ onto $\mathbf{n}^{-1}$, using **intercept** as the estimate

Now, what does $\log \eta_k(A)$, as a function of k, look like? We have

$$\alpha \le k^{-1}\log \eta_k(A) \;\Rightarrow\; 2^{k\alpha} \le \eta_k(A)$$

since entropy dimension bounds Hausdorff dimension. So $\eta_k(A)$ should be close to $2^{\alpha k}$. But it could be much larger if the net was inefficient for covering A. In fact, $\eta_k(A)$ could be up to b_u times as big as $2^{\alpha k}$. Hence,

$$2^{\alpha k} \le \eta_k(A) \le b_u 2^{\alpha k}$$

or

$$\alpha k \le \log \eta_k(A) \le K + \alpha k.$$

Now consider A_n; define α_n=dim $A_n \le \alpha$ for all n. Also, $\alpha_n \ge \alpha_{n+1}$ for n=1,2 3, $\cdots$. Scaling each $u_n(x)$ up to the unit cube, and using the above inequality, we have

$$\alpha_n(k-n) \le \log \eta_k(A_n) \le K + \alpha_n(k-n), \quad k \ge n$$

So the relative entropy satisfies (after simplifying)

$$\begin{aligned}(\alpha-\alpha_n)k + \alpha_n n - K &\le \log \eta_k(A) - \log \eta_k(A_n) \\ &\le (\alpha-\alpha_n)k + \alpha_n n + K. \end{aligned} \tag{2.7}$$

Hence,

$$\frac{(\alpha-\alpha_n)k-K}{n} + \alpha_k \le n^{-1}\log \frac{\eta_k(A)}{\eta_k(A_n)} \le \frac{(\alpha-\alpha_n)k+K}{n} + \alpha_k$$

Let's inspect equation (2.7). For large k, fixed n, it looks like

$$n^{-1}\log(\eta_k(A)/\eta_k(A_n)) = \frac{(\alpha-\alpha_n)k}{n} + O(1) \tag{2.8}$$

From (2.8) we infer that x should be chosen so that $\alpha_n=\alpha$ for each n. (This can be accomplished.) Otherwise,

$$\overline{\lim_{k\to\infty}} \log \frac{\eta_k(A)}{\eta_k(A_n)} = \infty.$$

If $\alpha_n=\alpha$ for all n, then (2.8) is close to a constant and we can evaluate

$$J_n^*(k) \equiv \sup_{r \geq k} n^{-1}\log \frac{\eta_r(A)}{\eta_r(A_n)}, \quad k=1,2,3,\cdots$$

and take limits of $J_n^*(k)$ as $k\to\infty$, using an extrapolation technique, such as Richardson (look in index of Johnson and Reisz 1982, for instance), L_1-, or L_2-regression using the intercept as the limit.

In summary, it is proposed that we evalute

$$K(n,k) \equiv n^{-1}\log \frac{\eta_k(A)}{\eta_k(A_n)}$$

for each n, k. Find $J_n^*(k)$ by $J_n^*(k)=\sup_{l \geq k} K(n,l)$. Find J(n) by extrapolating $J_n^*(k)$ to the limit as $k\to\infty$. Then extrapolate J(n) to the limit α by a regression technique.

2.2. Finding Limits

The author considered three extrapolation techniques: Richardson, L_1-, or L_2-regression. From Monte Carlo experiments with randomly generated samples on sets, he has found that Richardson extrapolation is the best method for finding the inside limit. Therefore, find

$$J(n) = \lim_{k \to \infty} K(n,k)$$

by using Richardson extrapolation. Then find the estimate of the dimension $\hat{\alpha}$ by performing an L_2 regression of J(n) onto n and using the slope as the estimate. See Appendix A for the details.

2.3. Properties of $\hat{\alpha}$.

Although the author has been unable to develop the analytic properties of $\hat{\alpha}$, he will look at the simple estimator

$$\hat{\alpha}_{RE} = n^{-1} \log \frac{\hat{\eta}_k(A)}{\hat{\eta}_k(A_n)}$$

for large k and n. Looking at the simple estimator $\hat{\alpha}_{RE}$ might give some clue as to how $\hat{\alpha}$ behaves. The estimator $\hat{\alpha}$ should behave similarly to the simple estimator since it is a linear combination of the simple estimators.

The estimator $\hat{\alpha}_{RE}$ above is a local estimator, A_n depending on the particular data point x chosen. One would like a more universal estimator, depending on all the data. A natural way to accomplish this is to average the simple estimators, as discussed in Example 2.13 above. However, one must not average over the data, as this introduces a positive bias to the estimate (see section 2.4 below); instead, one should average over *area*. The averaged estimator, which will be denoted $\hat{\alpha}_{\overline{RE}}$, is more complicated. However, being an average, it should exhibit more stability than the unaveraged version. Hence, its properties will be investigated. But it is derived from the simple relative entropy estimators; and so its properties will be derived from those of the simple relative entropy estimator. Thus, we need to investigate the relative entropy estimator in more depth. Section 2.4 will deal with the averaged entropy estimator.

The expression above for $\hat{\alpha}_{RE}$ can be rewritten in a more convenient form:

$$\hat{\alpha}_{RE} = n^{-1} \log\left(\frac{\hat{\eta}_k(A_n^c \cap A)}{\hat{\eta}_k(A_n)} + 1 \right)$$

2.3.1. *Joint distribution of* $\hat{\eta}_k(A_n^c \cap A)$ and $\hat{\eta}_k(A_n)$. Consider a partiton of $\overline{\mathfrak{U}}_k$, the sets of order k, into classes I(1), I(2), ⋯, I(α), containing respectively b_1, b_2, ⋯, b_α sets with nonzero probability p_1, p_2, ⋯, p_α of receiving an observation. Let X denote a single observation that lies in one of the order k sets. So, $P(X \in I(g))=b_g p_g$.

Define $\mathbf{V}_{t'} = \left[V_{t'1},\cdots,V_{t'\alpha}\right]'$ to be the number of sets in classes $\left[I(1),\cdots,I(\alpha)\right]'$ that are occupied after taking t' observations. Supposing that $\mathbf{i} = \left[i_1,\cdots,i_\alpha\right]'$ sets have already been occupied, then

$$P(\mathbf{V}_{t'}=\mathbf{j} \mid \mathbf{i}) = \sum_{\mathbf{h}=\mathbf{0}}^{\mathbf{j}-\mathbf{i}} (-1)^{\sum_1^\alpha (j_{g'}-i_{g'}-h_{g'})} \prod_{g=1}^{\alpha} \binom{b_g-i_g}{b_g-j_g,\ h_g,\ j_g-i_g-h_g}$$

$$\times \left(\sum_{g''=1}^{\alpha} (h_{g''}-i_{g''})p_{g''} \right)^{t'} \tag{2.9}$$

where $\mathbf{j}=[j_1,\ldots,j_\alpha]'$. (See Johnson and Kotz 1977, equation 3.88, p. 153).

In our situation, we have $\mathbf{V}_t=\left[V_{t1}, V_{t2}\right]'=\left[\eta_k(A_n^c \cap A), \eta_k(A_n)\right]'$. Also, i=[0,1]. So (2.9) reduces to

$$P(\mathbf{V}_t=\mathbf{j}) = \sum_{h_1=0}^{j_1} \sum_{h_2=0}^{j_2} (-1)^{j_1+j_2-h_1-h_2-1} \binom{b_1}{b_1-j_1,\ h_1,\ j_1-h_1} \binom{b_2}{b_2-j_2,\ h_2,\ j_2-h_2}$$

$$\times \left(h_1p_1 + h_2p_2 - p_2\right)^{t-1} \tag{2.10}$$

Call $S \equiv \{\gamma \mid \gamma=x/y,\ x \in \{1,\ldots,b_1\},\ y \in \{1,\ldots,b_2\}\}$ the support of the ratio $\rho \equiv V_{t1}/V_{t2}$. Then, using (2.10), we obtain, for $\gamma \in S$,

$$P(\rho=\gamma) = \gamma^{-2}\sum_{j_1=1}^{b_1} \binom{b_1}{j_1}\binom{b_2-1}{j_1\gamma^{-1}-1} \sum_{h_1=0}^{j_1} (-1)^{j_1-h_1}$$

$$\sum_{h_2=0}^{j_2\gamma^{-1}-1} (-1)^{j_1\gamma^{-1}-h_2-1}\binom{j_1\gamma^{-1}-1}{h_2}\left(h_1p_1 + h_2p_2 - p_2\right)^{t-1}.$$

Then $P(K(n,k)=\theta) = P(\log(1+\rho)=\theta) = P(\rho=2^{\theta}-1)$ for $\theta \in \log(1+S)$. These expressions are unwieldy, and so approximations will be very useful.

Let $m_j \equiv V_{\infty j}$, j=1,2. Suppose that $\lim_{n\to\infty} n^{-1} \lim_{k\to\infty} \log(1+m_1/m_2) = \alpha$. Further, suppose that

$$t/m_j \to \infty$$
$$m_j \exp(-t/m_j) \to \infty$$

for j=1,2. In this situation, V_{tj} → Normal in distribution (Johnson and Kotz 1977). Johnson and Kotz (1977) give formulae for the number of empty sets, $X_j=m_j-V_{tj}$. Let t_1 be the number of observations that fall in $A_n^c \cap A$. Let t_2 be the number of observations that fall in A_n. So $t_1+t_2=t$. From Johnson and Kotz (1977), p. 114, we have

$$\mu_j = E(X_j \mid t_j) = m_j\,(1 - m_j^{-1})^{t_j}$$

$$\sigma_j^2 = \mathrm{Var}(X_j \mid t_j) = m_j\,(m_j-1)\,(1 - 2m_j^{-1})^{t_j} + \mu_j - \mu_j^2.$$

Now, $t_j \sim \mathrm{Bin}(t, m_j/m)$, where $m \equiv m_1 + m_2$. We now calculate

$$E(X_j) = E_{t_j}[\,E(X_j \mid t_j)\,]$$

$$\mathrm{Var}(X_j) = E_{t_j}[\,\mathrm{Var}(X_j \mid t_j)\,] + \mathrm{Var}_{t_j}[\,E(X_j \mid t_j)\,].$$

Now,

$$E_{t_j}[\ E(X_j \mid t_j)\ ] = m_j E[(1\text{-}m_j^{-1})^{t_j}] = m_j(1\text{-}m^{-1})^t.$$

Also,

$$\mathrm{Var}_{t_j}[\ E(X_j \mid t_j)\ ] = E_{t_j}[\ E(X_j \mid t_j)^2\ ] - \{E_{t_j}[\ E(X_j \mid t_j)\ ]\}^2.$$

The first term on the right is eventually going to cancel out, so we don't need to calculate it. And

$$E_{t_j}[\ \mathrm{Var}(X_j \mid t_j)\ ] = m_j(m_j-1)E[(1-2m_j^{-1})^{t_j}] + m_j E[(1-m_j^{-1})^{t_j}]$$

$$- E_{t_j}[\ E(X_j \mid t_j)^2\ ]$$

$$= m_j(m_j-1)(1-2m^{-1})^t + m_j(1-m^{-1})^t - E_{t_j}[\ E(X_j \mid t_j)^2\ ].$$

Hence,

$$E(X_j) = m_j(1-m^{-1})^t \tag{2.11a}$$

$$\mathrm{Var}(X_j) = m_j(m_j-1)(1-2m^{-1})^t + m_j(1-m^{-1})^t - m_j^2(1-m^{-1})^{2t} \tag{2.11b}$$

Further, from Johnson and Kotz (1977), p. 154,

$$\mathrm{Cov}(X_1, X_2) = m_1 m_2[(1-2m^{-1})^t - (1-m^{-1})^{2t}]. \tag{2.11c}$$

From these expressions, we see that

$$E(V_{tj}) = m_j[1 - (1-m^{-1})^t]$$

$$\mathrm{Var}(V_{tj}) = \mathrm{Var}(X_j)$$

$$\text{Cov}(V_{t1}, V_{t2}) = \text{Cov}(X_1, X_2).$$

Below, it will be shown that (V_{t1}, V_{t2}) is asymptotically bivariate normal. This fact will be used for now. See Appendix C for a proof of this.

2.3.2. *Distribution of* $\hat{\alpha}_{RE}$. Now consider

$$\log \frac{V_{t1}+V_{t2}}{V_{t2}} = \log (V_{t1}+V_{t2}) - \log (V_{t2}).$$

Suppose a random variable $X > 0$ is approximately normal. For instance, the log-normal distribution is approximately normal if the mean is much larger than the standard deviation. Then it is well known that log X is approximately N(log[E(X)], Var(X)/(E(X) $\log_e 2)^2$), obtained from a first order Taylor series expansion of log x. Using this fact, we see that

$$\log \frac{V_{t1}+V_{t2}}{V_{t2}} \sim N(\mu_{RE}, n^2\sigma^2_{RE}),$$

where

$$\mu_{RE} = \log\{m[1-(1-m^{-1})^t]\} - \log\{m_2[1-(1-m^{-1})^t]\} = \log m - \log m_2 \quad (2.12a)$$

and

$$n^2(\log_e 2)^2\sigma^2_{RE} = \frac{\text{Var}(V_{t1}+V_{t2})}{[E(V_{t1}+V_{t2})]^2} + \frac{\text{Var}(V_{t2})}{[E(V_{t2})]^2} - 2\,\frac{\text{Cov}(V_{t1}, V_{t2})+\text{Var}(V_{t2})}{E(V_{t1}+V_{t2})E(V_{t2})} \quad (2.12b)$$

From Monte Carlo experiments, this approximation works very well, even for sample sizes as low as 50.

These expressions need to be in terms of $\alpha = \dim A$. Given $\epsilon > 0$, there exist n, k such that

$$| \ n^{-1}\log[\ (m_1+m_2)/m_2\ ] - \alpha \ | \ \leq \epsilon.$$

Therefore,

$$2^{n(\alpha-\epsilon)} \leq \frac{m_1+m_2}{m_2} \leq 2^{n(\alpha+\epsilon)}.$$

Now, let's suppose that

$$\alpha - \epsilon \leq k^{-1}\log(m_1 + m_2) \leq \alpha + \epsilon,$$

i.e., that entropy and Hausdorff dimensions coincide, a very common assumption. This assumption will make the entropy estimator look better as compared to the relative estimator, since all properties will be derived under conditions used to justify the entropy estimator. (See almost any paper on fractals, for instance.) So,

$$2^{(\alpha-\epsilon)k} \leq m_1 + m_2 \leq 2^{(\alpha+\epsilon)k}$$

Some arithmetic yields

$$4^{-n\epsilon}2^{(\alpha-\epsilon)k} \leq m_2 \leq 4^{n\epsilon}2^{(\alpha+\epsilon)k}$$

and

$$(2^{n(\alpha-\epsilon)}-1)4^{-n\epsilon}2^{(\alpha-\epsilon)k} \leq m_1 \leq (2^{n(\alpha+\epsilon)}-1)4^{n\epsilon}2^{(\alpha+\epsilon)k}$$

To get an expression in terms of α, let us suppose that m_j is approximately the geometric mean of its bounds. Note that this is the same as taking the arithmetic mean of the logs of the bounds. For large n and k, and small ϵ, we then get

$$(m_1+m_2)/m_2 \approx 2^{\alpha n} \tag{2.13a}$$

$$(m_1+m_2) \approx 2^{\alpha k} \tag{2.13b}$$

$$m_1 \approx 2^{\alpha k} \tag{2.13c}$$

$$m_2 \approx 2^{\alpha(k-n)} \tag{2.13c}$$

Recall that

$$\hat{\alpha}_{RE} = n^{-1} \log\left(\frac{V_{t1}+V_{t2}}{V_{t2}}\right).$$

We see that

$$E(\hat{\alpha}_{RE}) \sim n^{-1}[\alpha k-(k-n)\alpha] = \alpha.$$

Now, the variance can be obtained without using the dimension, but simply by estimating m, m_1, and m_2, as is explained below. So, the approximate (1-γ)100% confidence interval for α is

$$\hat{\alpha}_{RE} \pm z_{\gamma/2}\, \hat{\sigma}_{RE} \tag{2.14}$$

where z_γ is the point so that $P(Z > z_\gamma)=\gamma$ for a standard normal variate Z.

2.3.3. *Estimating σ_{RE}.* We are going to estimate m, m_1, and m_2 for use in equations (2.11a, b, and c), and then substitute those results into equation (2.12b) to get the estimate. This estimation procedure is independent of α. Recall that

$$V_{t1}=\eta_k(A_n^c \cap A)$$

$$V_{t2}=\eta_k(A_n)$$

and t_1 observations fall in $A_n^c \cap A$, and t_2 observations fall in A_n, $t=t_1+t_2$. Conditioning on t_1 (or equivalently t_2), V_{t1} and V_{t2} are independent, and so we can maximize the likelihood of V_{t1} without regard to V_{t2}, and vice versa.

The likelihood of m_j is

$$L(m_j) = \binom{m_j}{V_{tj}} V_{tj} \cdot \text{constant}$$

(Johnson and Kotz 1977, p. 137). The ratio of likelihoods then turns out to be

$$R(m_j) = \frac{L(m_j+1)}{L(m_j)} = \frac{m_j+1}{m_j+1-V_{tj}} \left(\frac{m_j}{m_j+1}\right)^{t_j} \qquad (2.15)$$

We can perform a linear search to maximize equation (2.15), using the algorithm outlined below.

1. If $R(V_{tj}) < 1$, assign $\hat{m}_j = V_{tj}$.
2. If $R(V_{tj}) \geq 1$, increase $\hat{m}_j$ until $R(\hat{m}_j - 1) > 1 \geq R(\hat{m}_j)$.

2.4. Asymptotic Properties

Now, the author will investigate the asymptotic properties of the relative entropy estimator $\hat{\alpha}_{RE} \equiv n^{-1} \log[(V_1+V_2)/V_2)]$. Since we need a fairly large sample to get a reasonable estimate, this does not seem to be an unreasonable line of inquiry.

The characteristics of $\hat{\alpha}_{RE}$ will be compared to that of the entropy estimator $\hat{\alpha}_E \equiv k^{-1} \log(V_1+V_2)$. Again, let $tm_j^{-1} \to \infty$ (j=1, 2) which implies $tm^{-1} \to \infty$; and $m_j \exp(-tm_j^{-1}) \to \infty$ (j=1, 2) and $m \exp(-tm^{-1}) \to \infty$. So we still have approximate normality. Then, we obtain

$$\text{Var}(V_1+V_2) \approx m\, e^{-t/m} (1 - e^{-t/m})$$

$$\text{Var}(V_2) \approx m_2\, e^{-t/m} (1 - e^{-t/m})$$

$$E(V_1+V_2) \approx m\,(1 - e^{-t/m})$$

$$E(V_2) \approx m_2\,(1 - e^{-t/m})$$

$$Cov(V_1, V_2) \approx (m - m_2)\, m_2\, [e^{-2t/m} - e^{-2t/m}] \approx 0.$$

So, by substituting these expressions into equation (2.12b), we obtain (after simplifying)

$$n^2\,(\log_e 2)^2\, Var(\hat{\alpha}_{RE}) \approx \frac{\exp(-t/m)}{m(\,1 - \exp(-t/m)\,)}\left(\frac{m}{m_2}-1\right).$$

Also,

$$k^2\,(\log_e 2)^2\, Var(\hat{\alpha}_E) \approx \frac{\exp(-t/m)}{m(\,1 - \exp(-t/m)\,)} \equiv \beta^2.$$

Finally, we obtain

$$Var(\hat{\alpha}_{RE}) \approx \beta^2\left(\frac{m}{m_2}-1\right)/(n \log_e 2)^2$$

$$Var(\hat{\alpha}_E) \approx \beta^2/(k \log_e 2)^2$$

Therefore, looking at efficiency

$$e_{var}(\,\hat{\alpha}_{RE}, \hat{\alpha}_E\,) = \frac{Var(\hat{\alpha}_E)}{Var(\hat{\alpha}_{RE})} \approx \left(\frac{n}{k}\right)^2\left(\frac{m_2}{m_1}\right)$$

Using the usual approximations $m_2 \approx 2^{\alpha(k-n)}$, and $m_1 \approx 2^{\alpha k}$, we obtain

$$e_{var}(\,\hat{\alpha}_{RE}, \hat{\alpha}_E\,) \approx \left(\frac{n}{k}\right)^2 2^{-\alpha n}. \qquad (2.16)$$

Notice that the efficiency here (looking strictly at variance) is always less than one. So,

$Var(\hat{\alpha}_{RE}) > Var(\hat{\alpha}_E)$. However, there is a difference in biases. To see this, the above equations give

$$E(V_i) \approx m_i \, (1 - \exp(-t/m) \,), \;\; i=1, 2.$$

Therefore,

$$E(\hat{\alpha}_{RE}) \approx n^{-1} [\log m + \log m_2]$$
$$E(\hat{\alpha}_E) \approx k^{-1} [\log m + \log (1 - \exp(-t/m))].$$

At this point, we will make a more realistic assumption for the relationship between m_j and α. From the author's observations of sets in relation to the net, he has found that

$$m_2 \approx (1+c)2^{\alpha(k-n)},$$
$$m_1 \approx (1+c)2^{\alpha k},$$

for some $c > 0$ (depending on the set and its relation to the net placement). This works much better. With this approximation, one sees that equation (2.16) above still holds. For the means, we get

$$E(\hat{\alpha}_E) \sim \alpha + k^{-1}[\log(1+c) + \log\{1-\exp(-t2^{-\alpha k}/(c+1))\}$$

$$E(\hat{\alpha}_{RE}) \sim \alpha.$$

One can see that there is a complicated bias for the entropy estimator. The bias of the entropy estimator tends to increase as c increases and decrease as α and k increase. However, the function does have bumps in it.

This bias term might aid in estimating α using the entropy estimator. Instead of the model

$$E(V_1+V_2) = k\alpha + \text{constant},$$

we might amend it to

$$E(V_1+V_2) = k\alpha + [\ \log(1+c) + \log\{1-\exp(-t2^{-\alpha k}/(c+1))\},$$

and use nonlinear regression to solve for c and α. However, we will not look at this possibility here.

The efficiencies of the estimators $\hat{\alpha}_{RE}$ and $\hat{\alpha}_{E}$ will now be compared in terms of mean squared error (mse). Substituting the new m_1 and m_2 into the above equations for variance, we define

$$s \equiv \exp(-t2^{-\alpha k}/(c+1)),$$

and then obtain

$$\mathrm{Var}(\hat{\alpha}_E) \approx \frac{s}{(c+1)2^{\alpha k}(1-s)(k\ \log_e 2)^2}$$

$$\mathrm{Var}(\hat{\alpha}_{RE}) \approx k^{-2}s\left(\frac{2^{\alpha n}-1}{(c+1)\ 2^{\alpha k}\ (1-s)\ (\log_e 2)^2}\right)\left(\frac{k}{n}\right)^2.$$

And so,

$$e(\hat{\alpha}_{RE}, \hat{\alpha}_E) \approx \frac{(2^{\alpha n}-1)\left(\frac{k}{n}\right)^2}{\left\{1 + [\log([c+1][1\text{-}s])]^2\ \frac{(c+1)2^{\alpha k}(1-s)(\log_e 2)^2}{s}\right\}} \tag{2.17}$$

This equation has proven itself to be quite resistant to analytic investigation into its behavior. The best method of investigation is to look at plots of it as a function of c and α for various k and n. In the figures that follow, this has been done.

In Figure 2.1 are three contour plots of the square root of the efficiency, $[e(\hat{\alpha}_E, \hat{\alpha}_E)]^{1/2}$, plotted as a function of α and k, holding c and n fixed. In the plots, n=6 throughout. Pictures with other values for n look similar, and so are not shown. Figure 2.1(a) has c=0; figure 2.1(b) has c=1; and finally figure 2.1(c) has c=2. In all three figures, the horizontal axis has k ranging between 7 and 14, while the vertical axis has α ranging between 0.30 and 2.25 (very small to fairly large sets). One notices that the efficiency is bigger than one ($\hat{\alpha}_E$ better) in the high α range, and in the slanted region of the low α range. The slanted region has bumps in it, much like the Piedmont Mountains of the southeastern United States, flattening out as we travel east. This behavior is characteristic of all the plots.

Why are there bubbles in the efficiency plot? If one looks at the plot of $[\text{mse}(\hat{\alpha}_{RE})]^{1/2}$, it is fairly flat and nearly independent of k. Of course, α has a large influence. However, looking at the contour plot of $[\text{mse}(\hat{\alpha}_E)]^{1/2}$, we see that it has regions where it does particularly well.

Let us look at $\text{mse}(\hat{\alpha}_E)$ in more detail. If we define

$$z \equiv t/m$$

we can rewrite the $\text{mse}(\hat{\alpha}_E)$ as

$$k^2 \ \text{mse}(\hat{\alpha}_E) = \frac{e^{-z}}{1-e^{-1}} \frac{z}{(\log_e 2)^2 t} + [\ \log(1-e^{-z}) + \log(c+1)\ ]^2.$$

Recall that $z \to \infty$. So, for large z,

$$k^2 \ \text{mse}(\hat{\alpha}_E) \approx [\ \log(1-e^{-z}) + \log(c+1)\ ]^2.$$

So, the mse is almost purely bias squared.

The next logical question is, "Where is the bias approximately zero?" Define

$$\beta \equiv (c+1)z \approx t2^{-\alpha k}.$$

We do this to separate c from the rest of the expression (β is independent of c.) Then the bias is

$$k^{-1}[\log(c+1) + \log(1-\exp(-\beta/(c+1)))].$$

Setting the bias to zero, we get

$$-\log(c+1) = \log(1-\exp(-\beta/(c+1))).$$

Solving for β yields

$$\beta = (c+1)\log_e\left(\frac{c+1}{c}\right).$$

The graph of $\beta=\beta(c)$ is shown in Fig. 2.2(a). It is seen that β and c have an inverse relationship. In Fig. 2.2(b), k times the bias is plotted as a function of c for various values of β. There it is seen that the bias increases in both β and c. It is also seen that for β fixed, the bias is concave in c. Further, one can see where the bias vanishes.

If we replace β by what it denotes, we can solve for k. We have

$$k=\alpha^{-1}[\log t - \log\{(c+1)\log_e((c+1)/c)\}]. \tag{2.18}$$

So, if one knows approximately what α should be (and one often does), and if one can find c (which one can, approximately, as below), then one should be able to find the k to make the bias nearly zero, by using equation (2.18) above.

To find c, recall that

$$m_1+m_2 \approx (c+1)2^{\alpha k}.$$

So, in the linear regression, regressing $\log(V_1+V_2)$ onto k, the slope is $\hat{\alpha}$, but the intercept is approximately $\log(c+1)$. Then, just solve for c:

$$c \approx e^{\text{intercept}}-1.$$

Now, the variance of $\hat{\alpha}_{RE}$ will be investigated. But first, some simplifying notation is introduced. Define

$$\begin{aligned} E(V_1+V_2) &\equiv E \qquad & E(V_2) &\equiv E_2 \\ Var(V_1+V_2) &\equiv \Sigma \qquad & Var(V_2) &\equiv \Sigma_2 \end{aligned}$$
$$Cov(V_1, V_2) \equiv C.$$

In this new notation, we then have

$$Var(\hat{\alpha}_{RE}) \approx [n \log_e 2]^{-2}\left[\frac{\Sigma}{E^2} + \frac{\Sigma_2}{E_2^2} - 2\,\frac{C+\Sigma_2}{E\,E_2}\right] \tag{2.19a}$$

Now,

$$\frac{\Sigma}{\Sigma_2} = \frac{m(m-1)(1-2/m)^t + m(1-1/m)^t - m^2(1-1/m)^{2t}}{m_2(m_2-1)(1-2/m)^t + m_2(1-1/m)^t - m_2^2(1-1/m)^{2t}}$$

$$\approx \frac{(c+1)\,2^{\alpha k}\,e^{-z} - (c+1)\,2^{\alpha k}\,e^{-z}}{(c+1)\,2^{\alpha(k-n)}\,e^{-z} - (c+1)\,2^{\alpha(k-n)}\,e^{-2z}} = 2^{\alpha n}$$

Also,

$$E(V_1+V_2) = E = (c+1)\ 2^{\alpha k}\ [1-(1-1/m)^t]$$

$$E(V_2) = E_2 = (c+1)\ 2^{\alpha(k-n)}\ [1-(1-1/m)^t]$$

Hence,

$$\frac{\Sigma}{E^2} \div \frac{\Sigma_2}{E_2^2} = \frac{\Sigma}{\Sigma_2}\cdot\frac{E_2^2}{E^2} \approx 2^{\alpha n}\left(\frac{2^{\alpha(k-n)}}{2^{\alpha k}}\right) = 2^{-\alpha n}.$$

So, in (2.19a) the first term in the brackets is dominated by the second term. Investigating the third term,

$$R \equiv \frac{\Sigma_2}{E_2^2} \div \frac{2(C+\Sigma_2)}{E\ E_2} = \frac{1}{2}\frac{\Sigma_2}{C+\Sigma_2}\cdot\frac{E}{E_2} \approx \frac{1}{2}\frac{\Sigma_2}{C+\Sigma_2}\ 2^{\alpha n}$$

$$=(1/2)\ 2^{\alpha n}\ (1 + C/\Sigma_2)^{-1}.$$

Now, since the correlation between V_1 and V_2 lies between -1 and 0, it is easy to see that

$$-2^{\alpha n/2}\Sigma_2 \le C \le 0. \tag{2.19b}$$

Therefore,

$$|R| \ge (1/2)\ 2^{\alpha n}\ |\ 1 - 2^{\alpha n/2}\ | = \frac{(1/2)\ 2^{\alpha n}}{|2^{\alpha n/2}-1|} \ge (1/2)\ 2^{\alpha n}.$$

So, again, the second term dominates, although by less than in the previous instance.

From the above discussion, it is clear that the second term, Σ_2/E_2^2, explains the variability in $\hat{\alpha}_{RE}$. If we could reduce the second term, then $\hat{\alpha}_{RE}$ would be much more competitive. A good way to do this is, of course, to form an average for the V_2.

From (2.19b) it is clear that $C/\Sigma_2 \to 0$ as n, $k \to \infty$, and hence the correlation between V_1 and V_2 tends to zero as well, being bounded above in magnitude by $|C/\Sigma_2|$.

Hence, V_1 and V_2 are asymptotically uncorrelated. This fact will be used in Appendix C.

2.5. Stability by Averaging.

Earlier, we have seen that if we reduce Σ_2/E_2^2, then we would reduce $\mathrm{Var}(\hat{\alpha}_{RE})$ quite a bit. There are two ways of doing this: (i) Increase E_2 by decreasing n (but this would increase $\mathrm{Var}(\hat{\alpha}_{RE})$ by its influence on other terms); (ii) Decrease Σ_2. Option (i) clearly is not viable. So we must be content with option (ii).

Recall that $V_2=\eta_k(A \cap u_n(x))$, which depends on a *single* data point; so V_2 is really a local value. There is no reason not to average over more data points. Let us develop this idea.

We have a sample $(X_1, \ldots, X_t)$ of data points from the set of interest. Now, $A \cap u_n(x_i)$ can refer to only a few distinct sets, S_j, $1 \le j \le b$. So, $S_j=A \cap u_n(x_i)$ for some i, and the number of distinct $A \cap u_n(x_i)$ is b. Within each set j, t_j observations fell, and these t_j observations occupied V_{2j} k-order net sets.

It is tempting to compute the average $\overline{\log V_2}$ as

$$\overline{\log V_2} = t^{-1} \sum_{i=1}^{t} \log \eta_k(A \cap u_n(x_i)) = t^{-1} \sum_{j=1}^{b} t_j \log V_{2j}.$$

However, this *won't work*; the average obtained is biased high because t_j is positively correlated with V_{2j}. One should not average over the sample; one must average over space instead:

$$\overline{\log V_2} = b^{-1} \sum_{j=1}^{t} \log V_{2j}.$$

The mean of this is

$$E(\overline{\log V_2}) = \overline{\log m_2} + \log[1-(1-1/m)^t],$$

where $\overline{\log\ m_2}=b^{-1}\sum_{j=1}^{b}\log\ m_{2j}$. Now, to get the variance, let $E_{2j}=E(V_{2j})$, $\Sigma_{2j}=\mathrm{Var}\ V_{2j}$, and $C_{ij}=\mathrm{Cov}(V_{2j}, V_{2i})$.

We will estimate α with

$$\hat{\alpha}_{\overline{RE}} = b^{-1}\sum_{j=1}^{b}\log(\ \sum V_{2j}/V_{2j}) = \log(\sum V_{2j}) - \overline{\log\ V_2}$$

$$= [\log(\sum E_{2j}) - \overline{\log\ E_2}] + b^{-1}(\log_e 2)^{-1}\sum_j (V_{2j}-E_{2j})\Big(\frac{1}{\sum E_{2j}} - \frac{1}{E_{2j}}\Big).$$

So, we obtain

$$E(\hat{\alpha}_{\overline{RE}}) = \log(\sum E_{2j}) - \overline{\log\ E_2} = \log\ m - \overline{\log\ m_2},$$

and

$$\mathrm{Var}(\hat{\alpha}_{\overline{RE}}) = \frac{1}{(b\ \log_e)^2}\left\{ \sum_{j=1}^{b}\Sigma_{2j}\Big(\frac{1}{\sum E_{2j}} - \frac{1}{E_{2j}}\Big)^2 \right.$$

$$\left. - 2\sum_{k=1}^{b-1}\ \sum_{j=k+1}^{b} C_{kj}\Big(\frac{1}{\sum E_{2j}} - \frac{1}{E_{2j}}\Big)\Big(\frac{1}{\sum E_{2j}} - \frac{1}{E_{2k}}\Big)\right\}. \qquad (2.20)$$

We will compare this variance to that of the original relative entropy estimate $\hat{\alpha}_{RE}$. To do this, we make the simplifying assumption that $m_{2j} \approx m_2$ for all j. So, $\Sigma_{2j} \approx \Sigma_2$ for all j, $C_{ij} \approx C$ for all i,j, and $E_{2i} \approx E_2$ for all i. Hence, $E \approx bE_2$, and the expression now becomes

$$\mathrm{Var}(\hat{\alpha}_{RE}) \approx (n\ \log_e 2)^{-2}\Big(\frac{\Sigma_2}{E_2^2} - 2\,\frac{C+\Sigma_2}{E\ E_2}\Big) = \frac{1}{(n\ \log_e 2)^2}\Big(\frac{(b-2)\Sigma_2-2C}{bE_2^2}\Big).$$

We also have, simplifying (2.20) above,

$$\mathrm{Var}(\hat{\alpha}_{\overline{RE}}) = \frac{1}{(n\ \log_e)^2}\, b^{-3}\,\frac{(b-1)^2}{bE_2^2}\,[b\Sigma_2 - b(b-1)C]. \qquad (2.21)$$

The ratio of the variances is

$$\frac{\mathrm{Var}(\hat{\alpha}_{RE})}{\mathrm{Var}(\hat{\alpha}_{\overline{RE}})} \approx \frac{[(b-2)\Sigma_2 - 2C]b}{(b-1)^2[b\Sigma_2 - b(b-1)C]b^{-2}} = \left(\frac{b}{b-1}\right)^2 \frac{(b-2)\Sigma_2-2C}{\Sigma_2-(b-1)C}$$

$$= \left(\frac{b}{b-1}\right)^2 \frac{(b-2) - 2C/\Sigma_2}{1 - (b-1)C/\Sigma_2}, \quad b \geq 2.$$

Since $-1 \leq C/\Sigma_2 \leq 0$, it is seen that this ratio always exceeds 1. So, averaging does improve the estimate.

We want an equation for the efficiency of $\hat{\alpha}_{\overline{RE}}$ relative to $\hat{\alpha}_E$. This is easily accomplished by taking the ratio of the mse's to yield:

$$e(\hat{\alpha}_E, \hat{\alpha}_{\overline{RE}}) \approx \left(\frac{k}{n}\right)^2 \frac{(b-1)^2 2^{\alpha n}/b^3}{1+[(\log_e 2)\log\{(c+1)(1-s)\}]^2\left(\frac{1-s}{s}\right)(c+1)2^{\alpha k}}.$$

Recall equation (2.16) for $e(\hat{\alpha}_E, \hat{\alpha}_{RE})$. If one looks at the equation for $e(\hat{\alpha}_E, \hat{\alpha}_{\overline{RE}})$ above, it is apparent that

$$e(\hat{\alpha}_E, \hat{\alpha}_{\overline{RE}}) \approx \frac{(b-1)^2}{b^3} \frac{2^{\alpha n}}{2^{\alpha n}-1} e(\hat{\alpha}_E, \hat{\alpha}_{RE}) \approx \frac{e(\hat{\alpha}_E, \hat{\alpha}_{RE})}{b}.$$

Thus, averaging drastically improves the competitiveness of the relative entropy estimator. Contour plots of the efficiency of $\hat{\alpha}_{\overline{RE}}$ to $\hat{\alpha}_E$ are shown in Fig. 2.3. These are the same type of plots as seen in Fig. 2.1, but with averaging. Notice how averaging has reduced the region where $\hat{\alpha}_E$ does better than the relative entropy estimator.

2.6. Consistency of the Relative Entropy Estimators.

The author will now demonstrate that the relative entropy estimators are consistent. If one knew m_i for all n and k, one could calculate α. However, even in this situation, it is not guaranteed that one could calculate α using the entropy estimator (Farmer, et al. 1983). First, a lemma is needed.

Lemma 2.14. Let $\{a_{ij}\}$ be an infinite array ($i=1, 2, ..., \infty$, $j=1, 2, ..., \infty$). And let $a=\lim_{i\to\infty}\lim_{j\to\infty}a_{ij}$ (and assume all limits exist). Then, there is a function $j(i)$ so that (1) $j(i)\to\infty$, (2) $j(i)-i\to\infty$, and (3) $a_{i,j(i)}\to a$ as $i\to\infty$.

Proof. Given $\epsilon > 0$. Then there exist i^* so that for $i \geq i^*$, $|\lim_{j\to\infty}a_{ij}-a| < \epsilon/2$. Also, there exist $\tilde{j}(i)$ so that $|a_{i,\tilde{j}(i)} - \lim_{j\to\infty}a_{ij}| < \epsilon/2$. Recursively define $j(i)$ by: $j(i^*)=\tilde{j}(i^*)$, and $j(i)=\max\{\tilde{j}(i), j(i-1) + i + 1\}$, $i > i^*$. Now, since $j(i) \geq i$, $j(i)\to\infty$. Also, $j(i)-i \geq j(i-1)\to\infty$ as $i\to\infty$. And finally, $|a_{i,j(i)}-\lim_{j\to\infty}a_{ij}| < \epsilon/2$. Therefore, $|a_{i,j(i)}-a| < \epsilon/2 + \epsilon/2 = \epsilon$. □

Now to state the consistency theorem.

Theorem 2.15. Consider the simple relative entropy estimate $\hat{\alpha}_{RE}=n^{-1}\log(\,(V_1+V_2)\,/\,V_2\,)$ with no averaging. Let m_1 and m_2 be defined as before. Suppose that $m_i(1-m^{-1})^t\to 0$. Then there exists a consistent sequence of estimators for α.

Proof. Let $a_{nk}=n^{-1}\log((m_1+m_2)/m_2)$. By the lemma above, there exists $k(n)$ so that $a_{n,k(n)}\to\infty$. Now, $E(V_i)=m_i-m_i(1-m^{-1})^t\to m_i$. Also, $m_i(1-m^{-1})^t\to 0$ implies $m_i(1-2m^{-1})^t\to 0$. Therefore,

$$\text{Var } V_i = m_i^2[\,(1-2/m)^t - (1-1/m)^{2t}\,] + m_i[\,(1-1/m)^t - (1-2/m)^t]\to\infty.$$

So,

$$n^{-1}\log\frac{V_1+V_2}{V_2} \to_{\mathfrak{P}} \lim_n n^{-1}\,[\,\log m - \log m_2\,] = \alpha,$$

where $\to_{\mathfrak{P}}$ means convergence in probability. □

This theorem says that the relative entropy estimate is consistent. We can easily show that $\hat{\alpha}_{\overline{RE}}$ is consistent as well, since $\hat{\alpha}_{\overline{RE}}$ has the same mean as $\hat{\alpha}_{RE}$, and a smaller variance.

An aside: The theorem has a pretty strong assumption (although no stronger than the central limit theorem assumptions for the urn models): that $m_i(1-1/m)^t \to 0$. One does not need this assumption as the weaker condition that Var $V_i/n^2 \to 0$ will suffice.

Table 2.1. Estimates of dim([0, 1/3] ∪ [2/3, 1]) for various values of n.

n	5	6	7	8	9	10
$\overline{Q}_0^A u_n(0))/n$	0.893	0.903	0.916	0.927	0.935	0.942

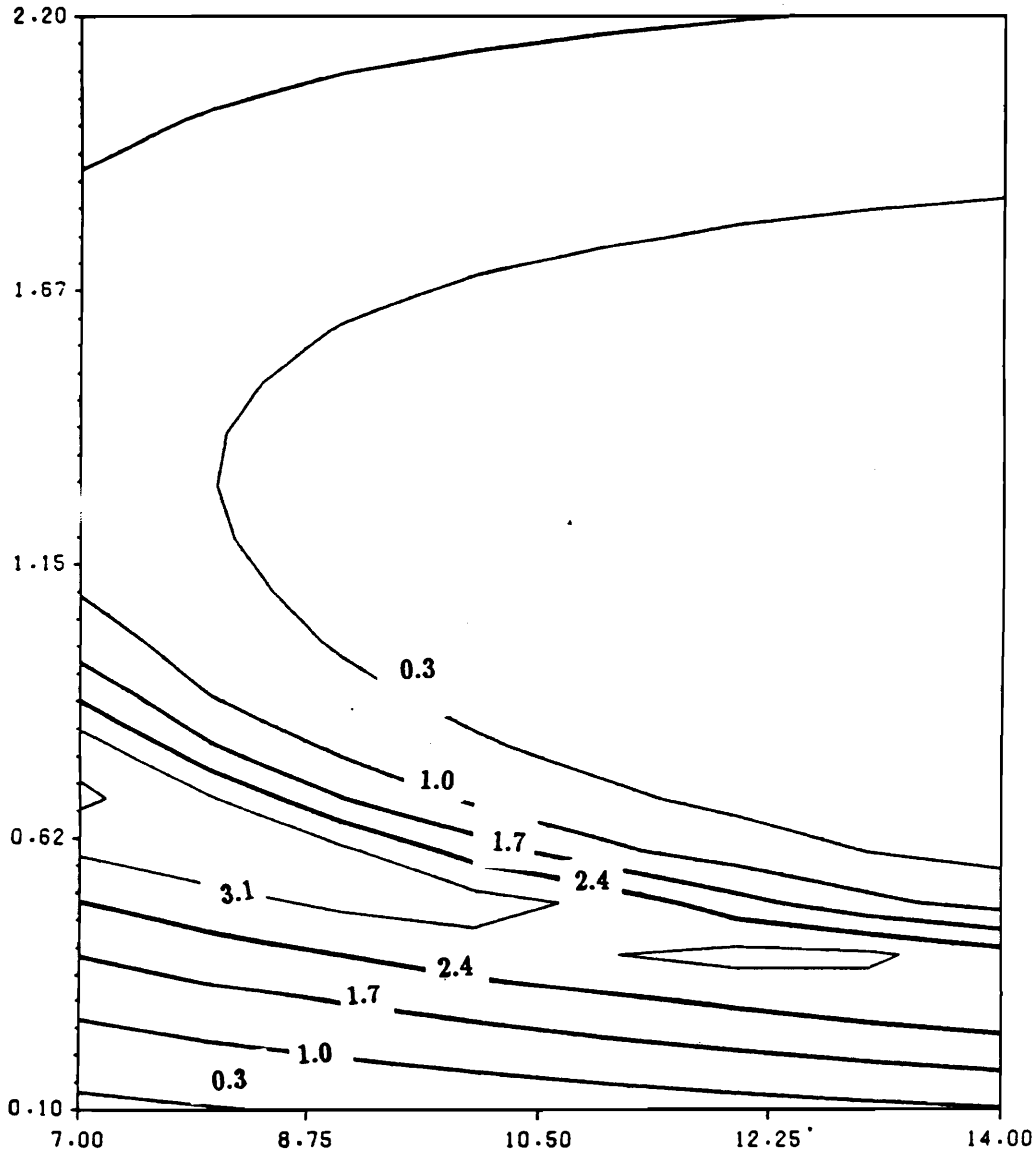

Figure 2.1. (a) Contour plots of the square root of the efficiency of the entropy estimator to the relative entropy estimator. In the figure, n=6, c = 0.

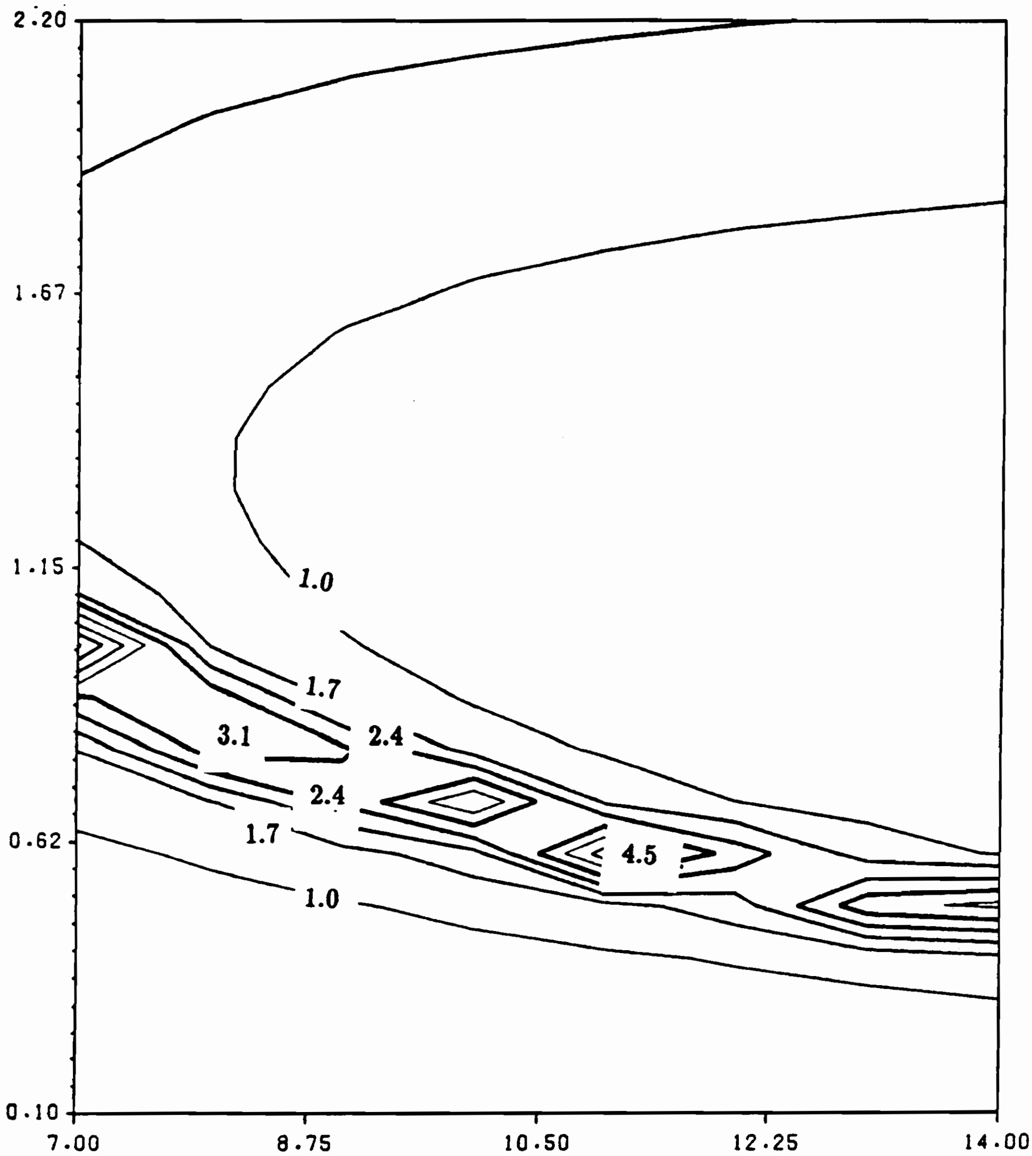

Figure 2.1. (b) Contour plots of the square root of the efficiency of the entropy estimator to the relative entropy estimator. In the figure, n=6, c = 1.

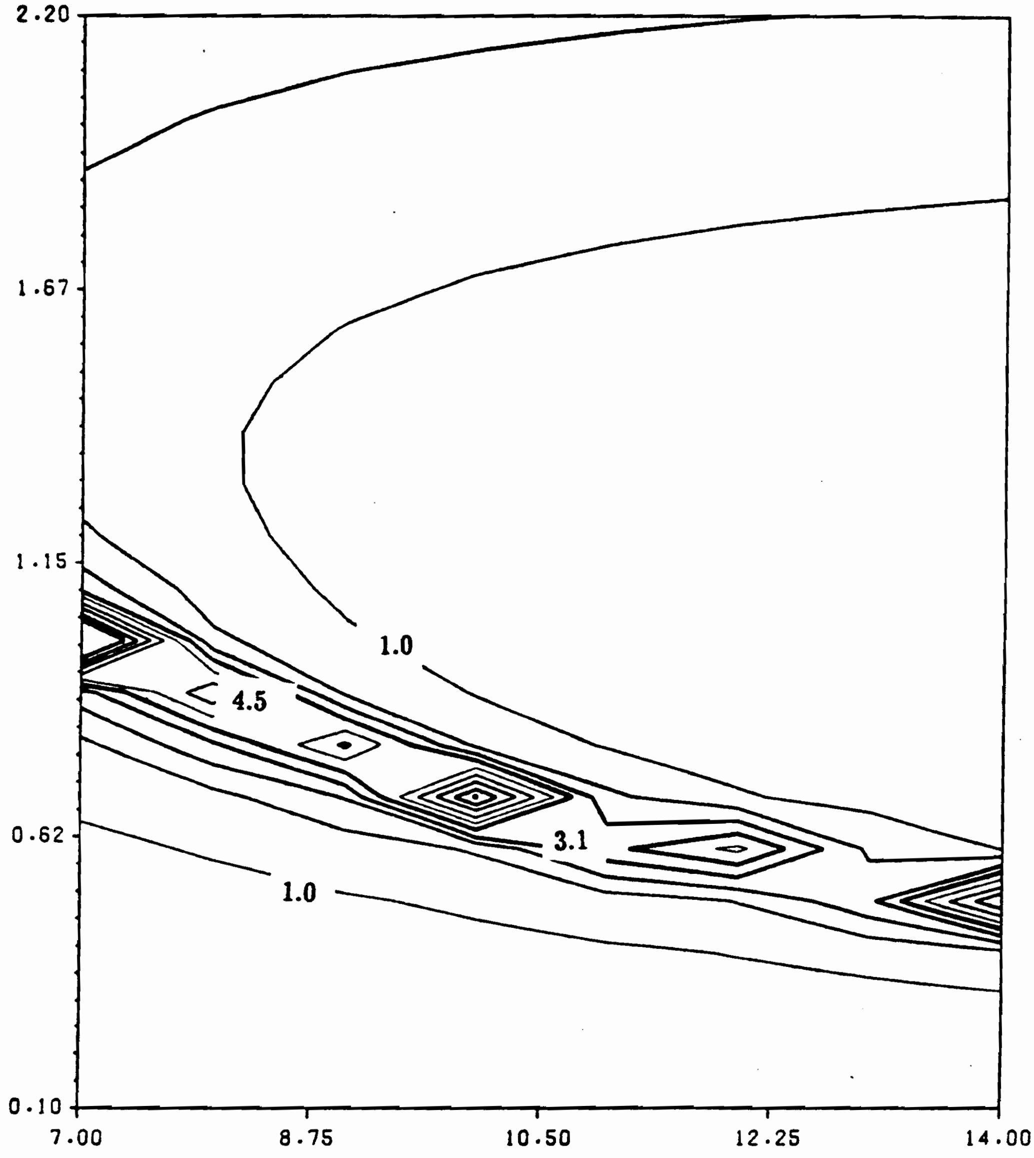

Figure 2.1. (c) Contour plots of the square root of the efficiency of the entropy estimator to the relative entropy estimator. In the figure, n=6, c = 2.

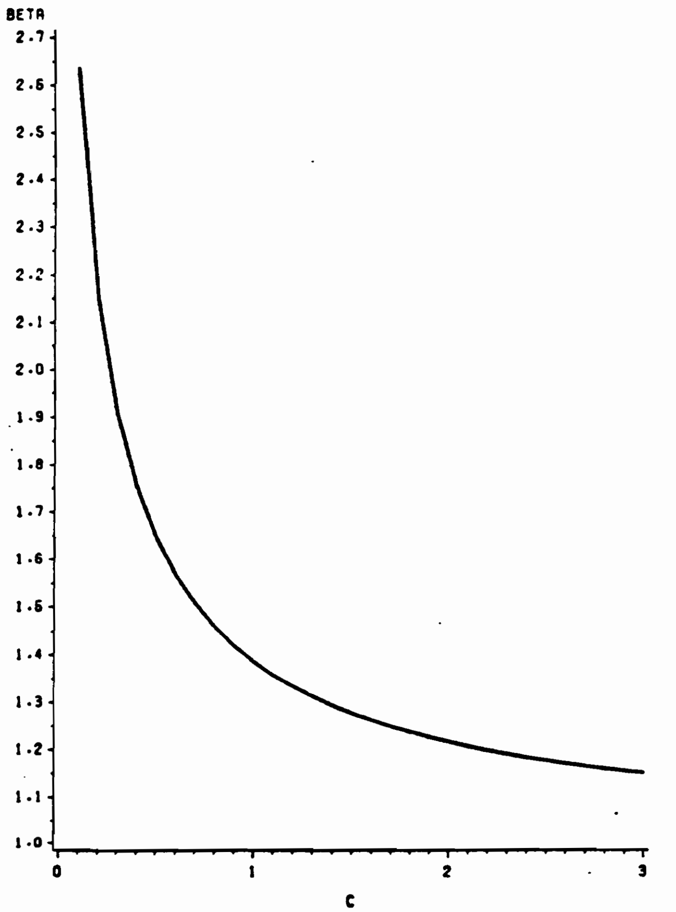

Figure 2.2. (a) Graph of $\beta=(c+1)\log_e(1+c^{-1})$.

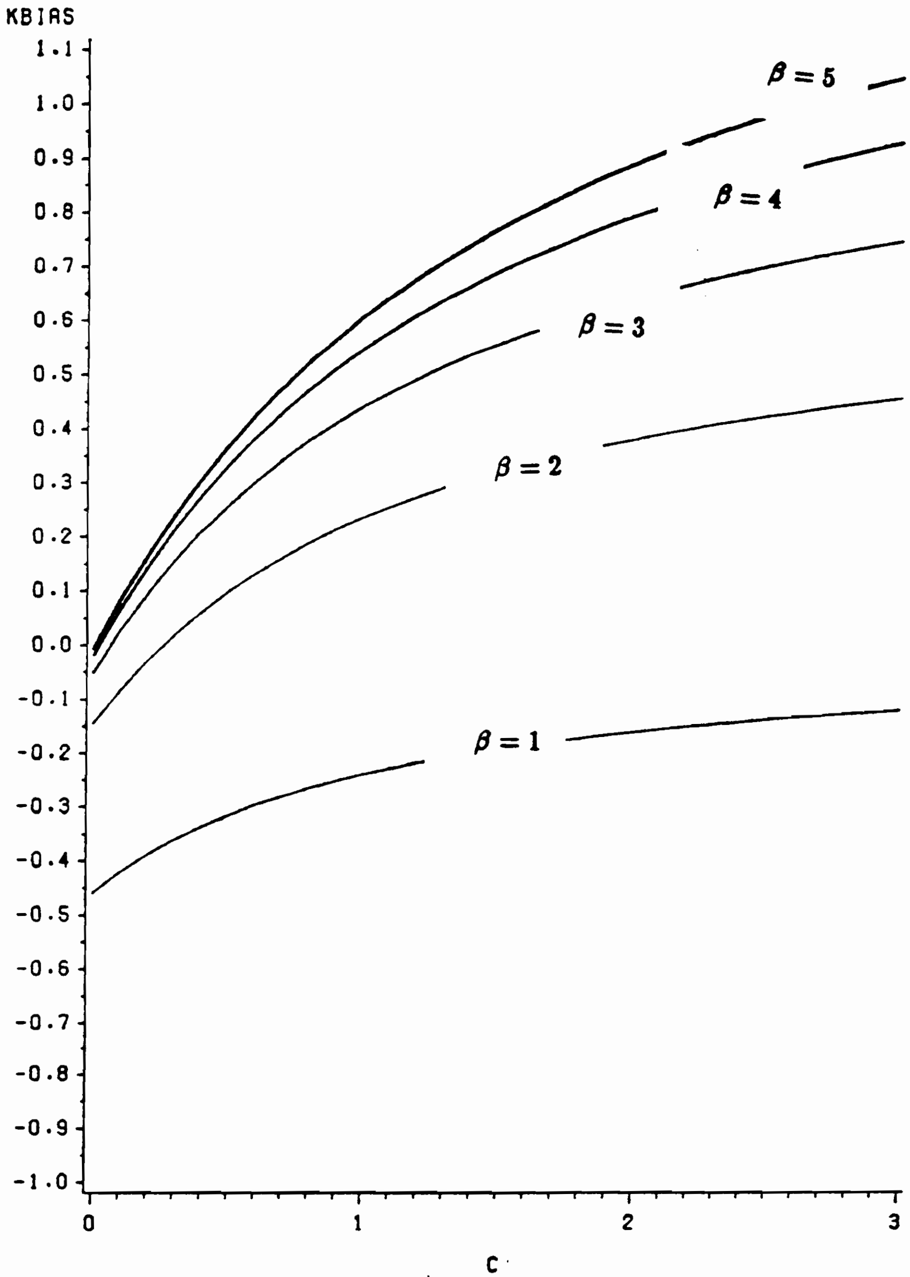

Figure 2.2 (b) Graph of k × bias for various values of β (fixed).

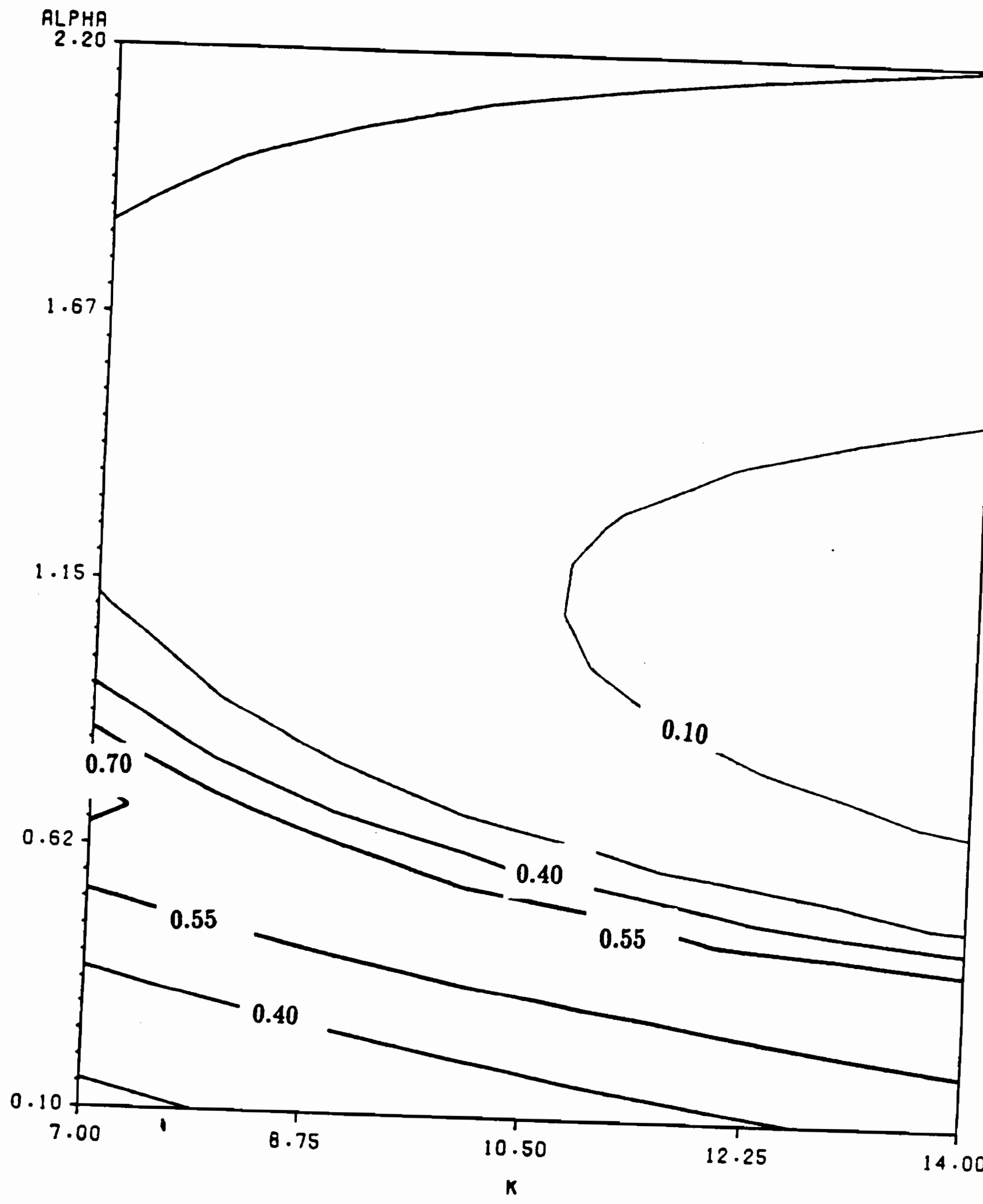

Figure 2.3. (a) Contour plots of the square root of the efficiency of the entropy estimator to the relative entropy estimator with averaging (b=40). In the figure, n=6, c = 0.

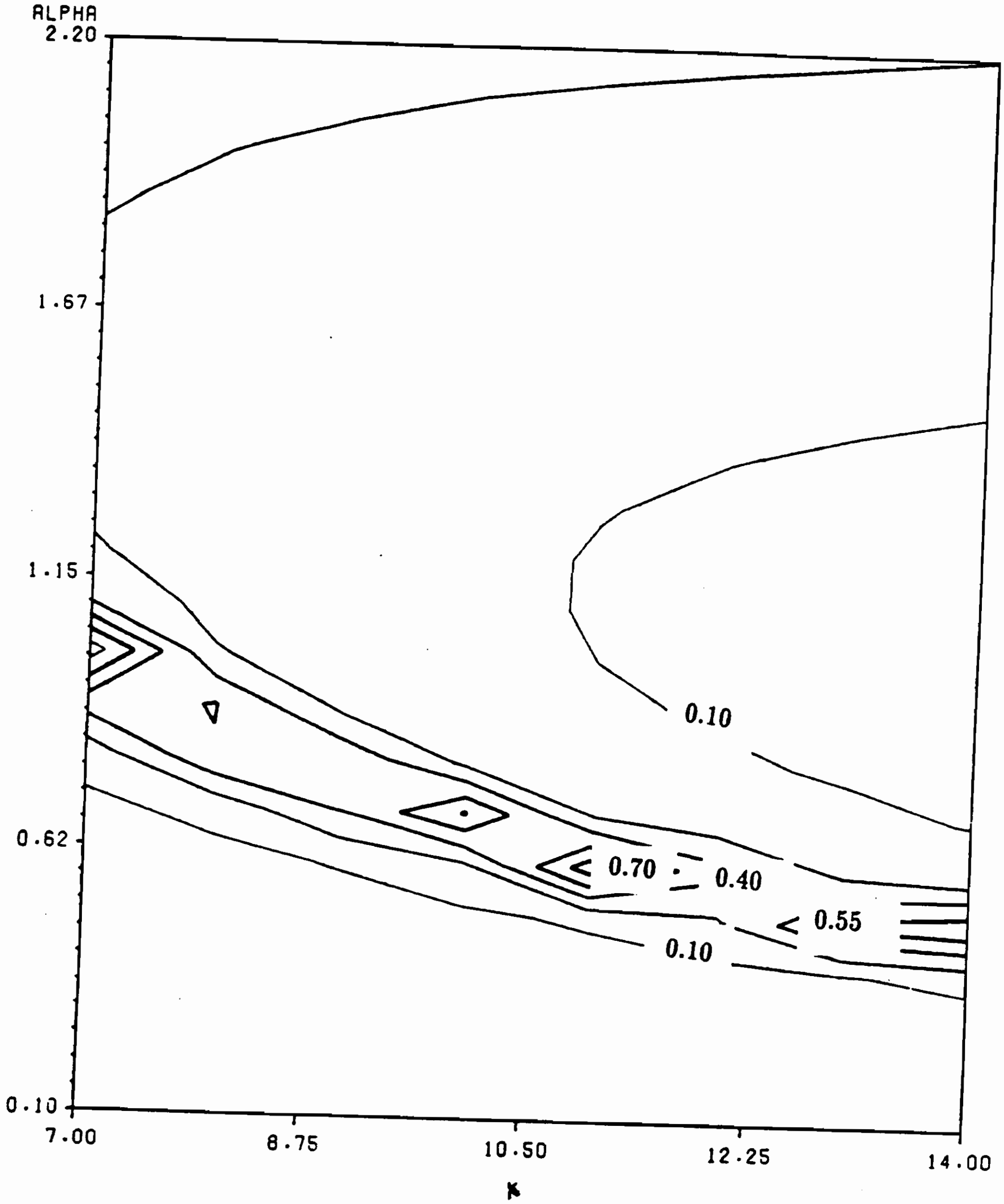

Figure 2.3. (b) Contour plots of the square root of the efficiency of the entropy estimator to the relative entropy estimator with averaging (b=40). In the figure, n=6, c = 1.

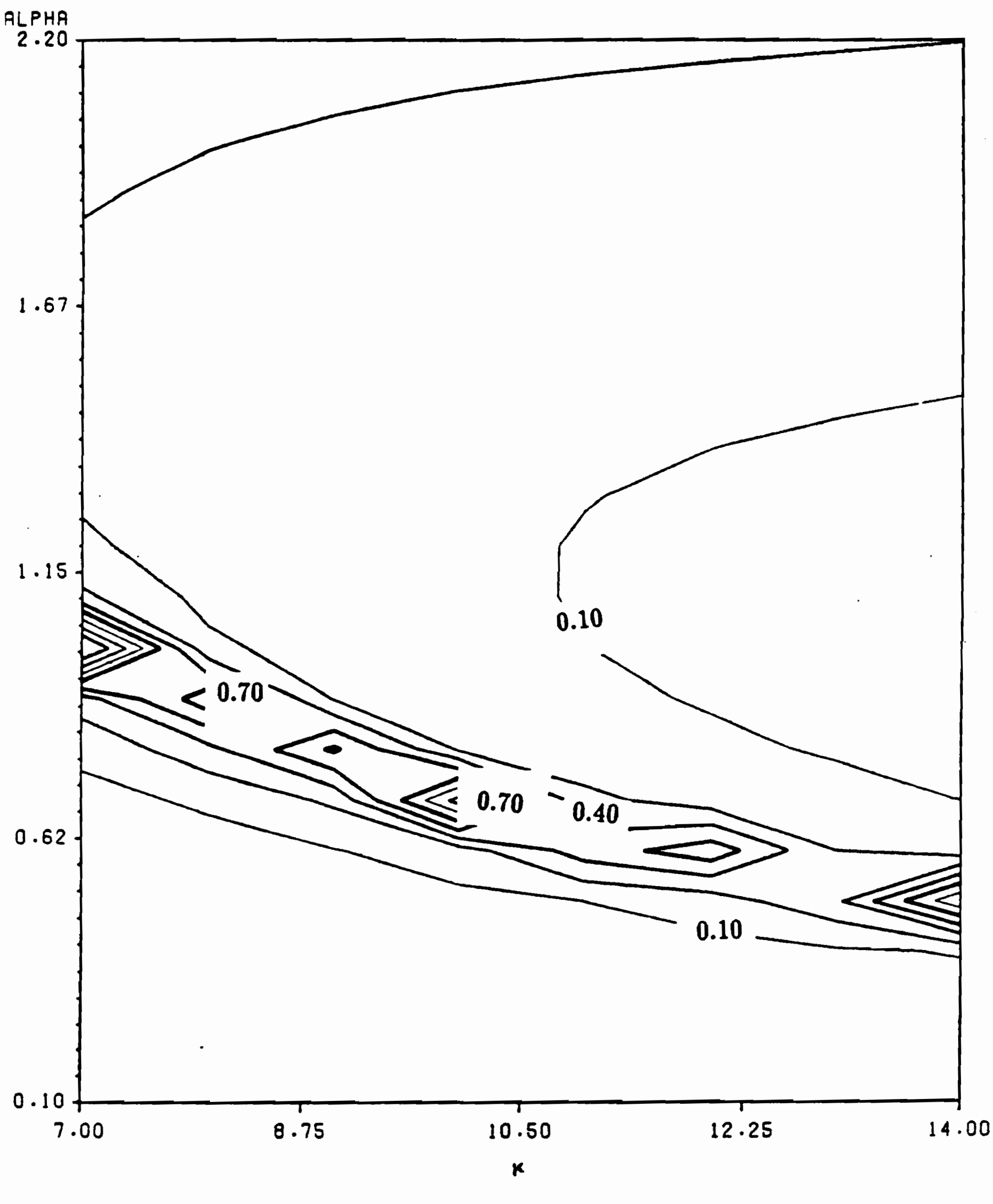

Figure 2.3. (c) Contour plots of the square root of the efficiency of the entropy estimator to the relative entropy estimator with averaging (b=40). In the figure, n=6, c = 2.

Chapter 3
Applications

In this chapter the author will present some applications of the method developed in the previous chapters. The applications will shed some new light on well-known processes. First, however, the author will make some pragmatic remarks on the estimation process.

The computations to estimate the dimension are performed using the method of Chapter 2, using a Pascal program implemented on a personal computer. The program is named Hausdorf. The source code can be found in Appendix B, along with an explanation of how it works.

Hausdorf outputs a series of simple estimates along with their estimated standard deviations, the standard deviations being computed from equation (2.12b). (Refer to the sample output in Appendix B.) Hausdorf performs a weighted regression, regressing $n\hat{\alpha}_{RE}$ onto n. The slope, of course, as in Chapter 2, is $\hat{\alpha}$. But, not all of the simple estimators will fall on the line, the departures being the most serious (and most influential) at the end of the list of simple estimators. We may get simple estimates for all n so that $2 \leq n \leq 10$, for instance. Here, the simple estimates most likely to be off the line would be the two end estimates; *e.g.*, when $n = 2$ or $n = 10$. The author's experience is that these end estimates are too small. The non-fitting points are dropped one by one, at the user's discretion, until the line fits the data. The slope at this point is $\hat{\alpha}$.

Once an estimate has been found, one ought to be sure that the line generated from it does in fact fit the data. One check of this is to make sure that the standard deviation of the slope, calculated from the actual residuals, is considerably smaller than the standard deviations printed next to the simple estimates. Also, the slope should not be far from the simple estimates it is derived from.

If a problem with the fit should develop, *e.g.*, the residuals are big or the slope is far from the simple estimates, then one of the simple estimates ought to be used. The author prefers to take the median of the estimates last printed to the screen.

To obtain standard errors, consider the simple estimator. The simple estimator has a standard error that is calculated using equation (2.12b), as mentioned earlier. The estimator actually used is based on these simple estimators. The covariance structure of these estimators is very difficult to discern. However, the variance of $\hat{\alpha}$, whether obtained using regression or finding a median, cannot be larger than the largest variance; so, the largest variance is used for the variance of $\hat{\alpha}$. This gives conservative confidence intervals and tests.

To show that this does indeed work, the author simulated a Brownian function (α = 1.5 a.s.), and took a regularly spaced sample of t = 200 points. Doing this several times, he got the following estimates, plus or minus the standard error:

1.36 ± .145	1.33 ± .110	1.27 ± .142	1.44 ± .168
1.25 ± .104	1.38 ± .136	1.39 ± .097	1.33 ± .110

The average estimate was 1.34 with a standard deviation of 0.063, well below the average conservative standard error (0.1265). The conservative error is indeed conservative. However, for sparse data sets, as in the case just simulated, it appears that there is a negative bias. For larger data sets, the bias reduces. Consider a sample size three times as large, t = 600. Performing the simulations for this situation, the author got the following:

1.51 ± .088	1.30 ± .155	1.38 ± .090
1.55 ± .069	1.29 ± .097	1.44 ± .101

The estimates have a mean of 1.41 and a standard deviation of .110. The performance is improving with the sample size.

The discussion just preceding gives some indication of how the procedure works in a moderately high-dimensional situation, and in violation of the random sample assumption. The fact that the observations are not independent and identically distributed does not seem to affect the statistical properties of the estimator seriously.

Some applications of the relative entropy estimator will now be discussed.

3.1. *Stock market.* One of the basic premises used by market analysts is that the logarithms of the stock prices follow a Brownian motion (Smith 1976, p. 628; see also Cho and Frees 1988 and O'Brien 1986). This assumption has been used to develop option pricing models; for instance, it was used in the classic paper by Black and Scholes (1973) and subsequent extensions by other authors. (See Smith (1976) for an excellent review of the literature of papers building on the Black and Scholes model.) If the Brownian motion hypothesis is true, the dimension of the graph of the price of a stock should be 1.5, as pointed out in Chapter 1.

To test the hypothesis $H = 1/2$, the author randomly selected Seamens Corporation from the American Stock Exchange. Taking the first 600 trading days following January 1, 1986, the closing price was recorded.

Using Hausdorf, the dimension is estimated to be (correcting for the negative bias) $\hat{\alpha} = 1.435$. Using the conservative standard deviation of $\hat{\sigma} = 0.07117$, we get an asymptotic 90% confidence interval of

$$(1.32,\ 1.55)$$

From this point of view, it appears that the Brownian motion hypothesis is tenable.

To get more information on the stock market, the author looked at the Dow Jones Industrial Average for 1000 trading days after January 1, 1986. From this the author got an estimate of $\hat{\alpha} = 1.41$ and a standard deviation of $\hat{\sigma} = .088$. This gives a 90% confidence interval of

$$(1.26,\ 1.55)$$

Again, the Brownian motion hypothesis appears tenable.

The level sets of a Brownian motion, i.e. $\{ t : B_{1/2}(t) = x \}$, has a dimension of 1/2 a.s., as mentioned in Chapter 1. For both series, the range was divided into 10 bands, and then the midpoint of 3 of them was randomly selected. Those levels are x = 5.832, 7.245, and 7.782. From these, the author computed estimates $\hat{\alpha}$ = .444, .5276, and .5689 respectively, with errors $\hat{\sigma}$= .229, .226, and .167. Taking an average of these values we obtain an asymptotic 90% confidence interval of

$$(.316,\ .712)$$

The Seamens stock seems to support the Brownian motion hypothesis.

So far, we have been working on the raw prices. However, most researchers tend to support the use of log-prices; i.e., taking the logarithm of the prices before analyzing them. Repeating the analysis as above on the transformed data, the author obtained, for the graph of the log-prices, the 90% confidence interval (after adjusting for the bias)

$$1.32\ \pm\ (1.645)(.0932) = (1.157,\ 1.46)$$

which does not cover 1.5 (= dim$\{ B_{1/2}(t) \}$). This leads one to suspect that the graph of the transformed data does not follow a Brownian motion. However, looking at level sets for levels x = 1.768 and 2.080 gives $\hat{\alpha}$ = .4906 and .6235 with standard deviations of $\hat{\sigma}$ = .226 and .152, respectively. This gives a confidence interval on the dimension of the level sets as

$$.557\ \pm\ (.1.645)(.136) = (.333,\ .781)$$

which does contain .5 (= dim{ t : $B_{1/2}(t)$ = x }).

The above discussion raises a doubt as to the validity of the Brownian motion on logged prices. More work on other stocks and with longer series needs to be done.

Performing the same calculations on the Dow Jones Industrial Average, the author got $\hat{\alpha}$ = .3932 and .3046 with errors of $\hat{\sigma}$ = .220 and .212 respectively. Only two estimates were calculated here since other levels produced too few data points to be used. From this he got a confidence interval for dim { t : B(t) = x } of

$$(.097, .601)$$

Therefore, this analysis supports the Brownian motion hypothesis.

Conclusion. Researchers, for good logical reasons, prefer to model log-prices with a Brownian motion. Only a few researchers have doubted this assumption (*e.g.*, Mandelbrot 1963). This analysis casts doubt upon this assumption, though it cannot be viewed as refuting it. Interestingly enough, the price itself seems to fit a Brownian motion better than the logged prices. The data seems to better support the ancient model of Bachelier (1900) than the current version that says that log-prices follow a Brownian motion. The Industrial Average, as well, seemed to conform to a Brownian motion behavior. The author did not take logarithms of this data because of its nature.

3.2. *Nonlinear Dynamics.* The dimension of the Julia set is of interest. Yet not much is known of the Julia set's dimension. This application is an attempt to investigate the dimension of two specific Julia sets using the relative entropy estimator.

The Julia set is defined by the complex map $z' = P_\mu(z) = z^2 - \mu$ for complex μ. Saupe (1987) gives some properties of Julia sets, which will be denoted by $J(P_\mu)$. Among them, he presents the fact that the Julia set is the closure of all preimages of a repelling fixed point z^* of P_μ:

$$J(P_\mu) \equiv \text{closure } \{z \in C : P_\mu^k(z)=z^* \text{ for some } k\}.$$

A repelling fixed point z^* of P_μ has the properties that $z^* = P_\mu(z^*)$ and $|P_\mu'(z^*)| > 1$. This provides a convenient way of generating the Julia set of P_μ. Pick a point z_0. The author's experience is that any nonzero point seems to work--the Julia set appears to be a strange attractor. Get the next point by finding the inverse $P_\mu^{-1}(z_0) = (z_0 + \mu)^{1/2}$. Of course, there are two roots to this inverse. If you consistently pick one root, you get a very small part of the set. Therefore, one of the roots was randomly picked for each iteration, thereby obtaining a random sample from the set, at least as random as could be obtained. Continue inverting the map iteratively to obtain more points. Saupe (1987) provides a description of an algorithm to generate the Julia set using this procedure, but complicates it to get every point from every root.

The Julia set for quadratic maps, indexed by the parameter μ, has just been described. However, one need not consider just quadratic maps; any rational or entire map will generate a Julia set as well. For the quadratic map, a result by Ruelle (1982) seems to be the only analytic result on the dimension of Julia sets:

$$\dim(J(P_\mu)) = 1 + |\mu|^2/4 \log_e 2 + \text{higher order terms}.$$

Even less is known of other Julia sets.

More information on Julia sets can be found in most books on fractals (*e.g.*, Peitgen and Richter 1986) or in the nonlinear dynamics literature (*e.g.*, Viczek 1989 and Devaney 1986).

For each of two Julia sets (identified below), the author collected four *independent* sets of 1000 data points each. He estimated the dimension for each data set and averaged them together. For a picture of the particular sets used, see Fig. 3.1(a) and 3.1(b) below. The 90% confidence intervals for the dimensions of the Julia sets are

$$P_1(z)=z^2-1\text{: } (1.318,\ 1.441)$$
$$P_{-i}(z)=z^2+i\text{: } (1.229,\ 1.438)$$

It is interesting to note that the dimension appears to change with the Julia set, and not just $|\mu|$, though this may be just sampling variation.

Note that Ruelle's formula for the dimension gives the value 1.361 (ignoring the higher order terms) for both sets, as different as they are. P_1 contains infinitely many circles, which P_{-i} is a dendrite. The value above is in the confidence interval for both P_1 and P_{-i}, suggesting that Ruelle's formula is a good approximation to the actual dimension.

Another attractor of interest in nonlinear dynamics is the Hénon attractor. The Hénon attractor lies in $\mathbb{R}^2$, and is found by iterating the map

$$\begin{aligned} x_{n+1} &= 1 + y_n + Ax_n^2 \\ y_{n+1} &= Bx_n \end{aligned}$$

(Peitgen and Saupe 1988.) The attractor with $A = 1.4$ and $B = 0.3$ seems especially popular. However, to the author's knowledge, no one has investigated this attractor's dimension. The attractor has many thin lines, each line containing many more lines within it, much like the Cantor set. See Fig. 3.3 for a picture of this Hénon attractor.

The author generated 4 sets of data, each containing 1000 data points on the attractor. He estimated the dimension and standard errors. Combining the four estimates, he got a 90% confidence interval for the dimension of the attractor of

$$(1.28,\ 1.36).$$

3.3. *Surface Terrain.* The dimension of surface terrain of various locations in the United States has been studied (Mark and Aronson 1984), in part to see if this way of looking at the surface could help researchers investigate what causes various features of terrain, and at what scale the processes of nature work; also of concern was a method of modeling errors in topographical map making. Rougher terrain creates larger errors.

For the Marshlands, Pennsylvania, quadrant, Mark and Aronson calculated an estimate of 2.7, which seems too high. Computer simulated terrain of dimension 2.7 was discernibly rougher than the real terrain used to get the estimate. The authors recognized this, but didn't know what to do about it.

The author believes that their estimator was biased high. The relative entropy estimator should overcome the bias and give a more realistic estimate.

The author did not have access to a digitized version of the maps; so he put two randomly selected lines on the topographic map and read off the elevation every eighth of an inch. Each of the lines was analyzed separately. A look at the map gives one the feeling that the dimension should be homogeneous in the entire region. The two (uncorrected) estimates and their conservative standard errors, respectively, are $\hat{\alpha}$=1.31 and 1.22, and $\hat{\sigma}$=0.122 and 0.148. This gives a conservative 90% confidence interval on the dimension of the surface of

$$(2.20,\ 2.64)$$

(adding unity to get the dimension of the surface, and correcting for the negative bias of the estimator, as shown earlier in the chapter.) Notice that this interval remains below 2.7. If you compare the graph of one of the slices to the Dow Jones Industrial Average series, but using only the first one-hundred points (to get comparable sample size), you see that the stock market appears to be rougher. Therefore, it would seem that the dimension of the slice of surface should be below that of the stock market (≈ 1.50). The estimate of the dimension of the slices is below 1.5, being 1.26.

Just for the sake of comparing one part of the Appalachian Mountains of Eastern

United States to another, the author did the same calculations for the Bluefield, West Virginia, quadrant. For Bluefield, he found the 90% confidence interval on the surface of the Earth to be

$$(2.12, 2.60).$$

This is about the same as the Marshlands region. This would suggest that the mountains do not get smoother as you travel south, but they remain essentially the same roughness. If you look at a slice of Bluefield and compare it to a slice of Marshlands, they do look like they have approximately the same roughness.

In Fig. 3.2 is displayed a slice of Marshlands, Bluefield, and first 100 data points of the Dow Jones Industrial Average series.

3.4. *Tree roots.* Seedlings are grown in large nurseries, each nursery producing millions of seedlings a year. These seedlings are then graded to determine whether to plant them in the forest or not for the purposes of revegetation. The grades range from 1 to 3, and are a criterion for determining the likelihood of a seedling to survival once planted. The grades are calculated on the basis of the width of the trunk at ground level. However, Peter Feret (1990), Forestry Department, Virginia Polytechnic Institute and State University, believes that root growth is a better predictor of a plant's ability to survive. A more developed, complicated root structure would predict a higher probability of surviving.

To investigate this theory, Feret planted several trees in thin boxes. These boxes were flat so that the roots would expand in a plane, rather than in 3-dimensions as they would in the wild. After the trees reached a specified grade, they were taken out of the box and the dirt in the roots was washed away. He then took a picture of the roots and digitized the picture.

The roots have a large body in the center, which has a dimension of 2 in the plane. However, this tap root does not give any information as to the development of the root.

Therefore, the tap root was filtered out of the pictures. The filtering preserves the character of the root, leaving the fine root strands, but deleting the large straight body.

From each grade, one tree was randomly selected, and its dimension calculated. The results are listed below (plus or minus the standard error).

Grade 1: 1.30 ± 0.070

Grade 2: 1.58 ± 0.161

Grade 3: 2

The roots of the grade 3 tree were a large mass, and the tap root was not distinguishable from the rest of the root structure. So it was assigned a dimension of 2.

One notices that the dimension increases with the grade. In other words, the dimension of the root structure does increase as the tree grows. This supports the idea that the dimension of the tree could be used for predicting tree survivability. This evidence, along with more evidence from Peter Feret (1990), Department of Forestry, Virginia Polytechnic Institute and State University, suggests that this might be a beneficial direction of investigation for foresters.

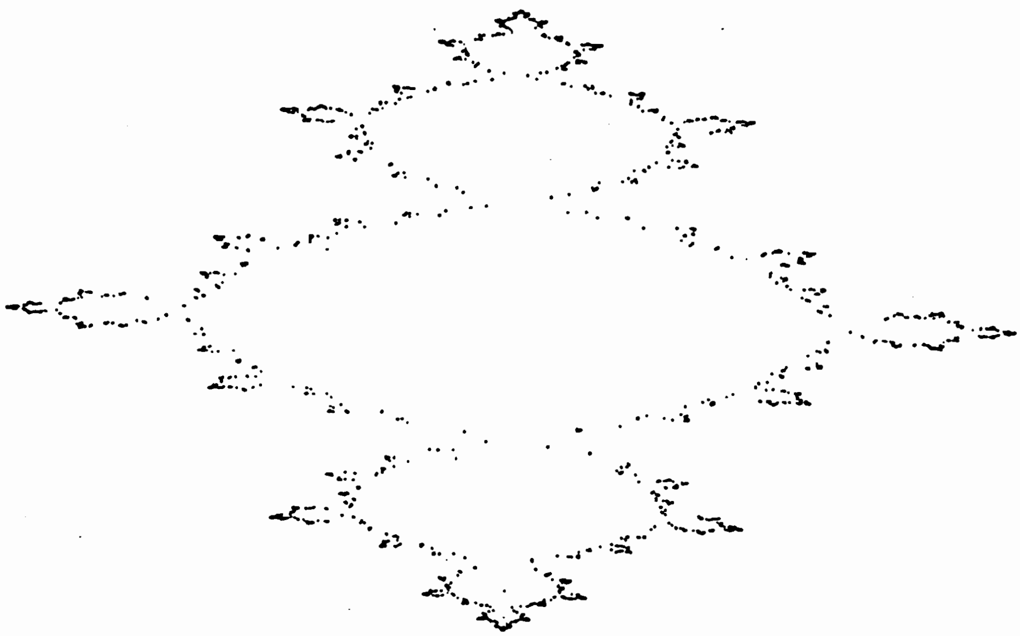

Figure 3.1. (a) This is a graphical representation of the combined data to estimate the dimension of the Julia set $J(z^2-1)$, using 4000 data points in all.

Figure 3.1. (b) This is a graphical representation of the combined data to estimate the dimension of the Julia set $J(z^2+i)$, using 4000 data points in all.

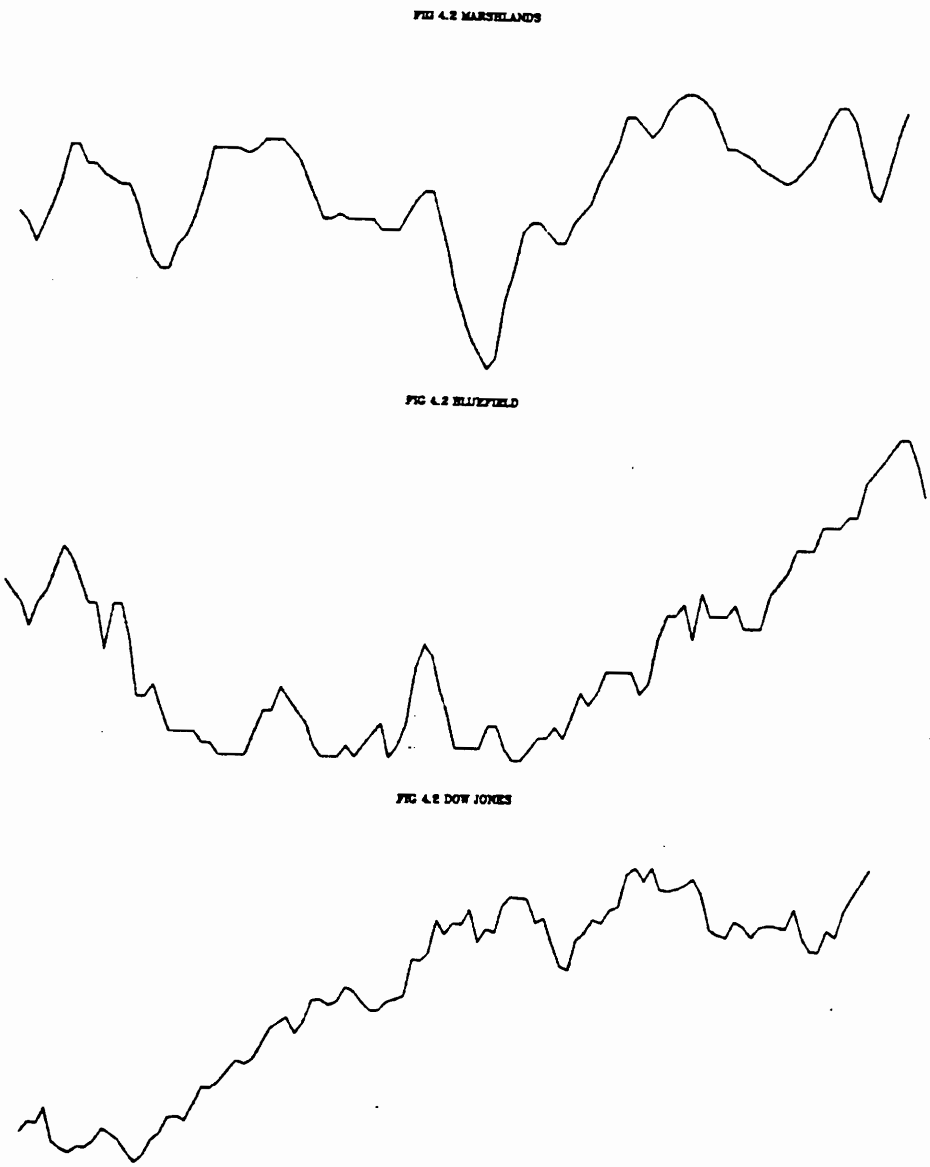

Figure 3.2. Above are three series of data. The top series is a slice of Marshlands, PA, quadrant. The middle series is a slice of Bluefield, WV, quadrant. The bottom series is the Dow Jones Industrial Average series (first 100 trading days). Notice that the Dow Jones series appears to be rougher than the other two series. Such a visual comparison is valid only for an equal number of observations in each series. Each series here contains approximately 100 data points.

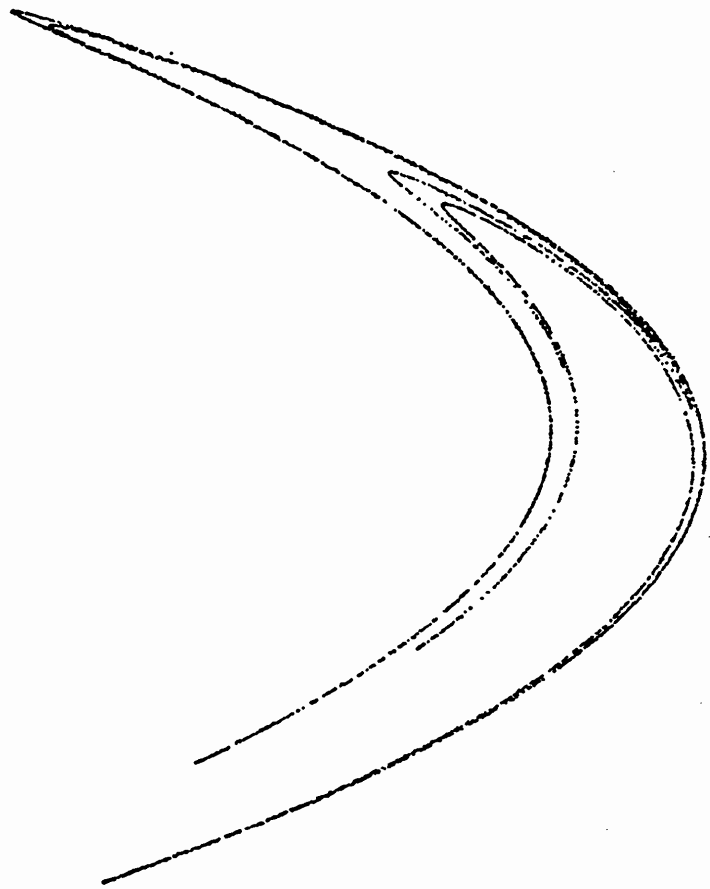

Figure 3.3. This is a graphical representation of the combined data on the Hénon attractor, using 4000 data points in all.

References

Allen, J., Colvin, J., Stinson, D., Flynn, C., Stapleton H. (1982) Protein conformation from electron spin relaxation data, *Biophysical J.*, **38**, 299-310

Amitrano, C., Coniglio, A., and diLiberto, F. (1986) Growth probability distribution in kinetic aggregation processes, *Physical Review Letters*, Vol. 57, No. 8, 1016-1019

Bachelier, L. (1900) Theory of speculation (English translation), in: P. Cootner, ed. 1964, *The Random Character of Stock Market Prices*, MIT Press, Cambridge, Mass. 17-78

Billingsley, P. (1960) Hausdorff dimension in probability theory, *Illinois J. Math.* **4**, 187-209

Billingsley, P. (1961) Hausdorff dimension in probability theory, II, *Illinois J. Math.* **5**, 291-298

Billingsley, P. (1968) *Convergence of Probability Measures*, New York: Wiley & Sons

Billingsley, P. (1986) *Probability and Measure, 2nd. edition*, New York: John Wiley & Sons

Black, F. and Scholes, M. (1973) The pricing of options and corporate liabilities, *The J. of Political Economy*, No. 3, Vol. 81, 637-654

Cho, D. C. and Frees, E. W. (1988) Estimating the volatility of discrete stock prices, *J. of Finance*, Vol. XLIII, No. 2, 451-466

Chorin, A. (1981) Estimates of intermittency, spectra, and blow-up in developed turbulence, *Comm. Pure Appl. Math.*, **34**, 853-866

Chorin, A. (1982a)Numerical estimates of hausdorff dimension, *J. of Computational Physics*, **46**, 390-396

Chorin, A. (1982b) The evolution of a turbulent vortex, *Commun. Math. Phys.*, **83**, 517-535

Clarke, K. (1986) Computation of the fractal dimension of topographic surfaces using the triangular prism surface area method, *Computer and Geosciences*, Vol. 12, No. 5, 713-722

Cutler, C., and Dawson, D. (1989) Estimation of dimension for spatially distributed data and related limit theorems, *J. of Multivarite Analysis*, **28**, 115-148

Devaney, R. L. (1986) *An Introduction to Chaotic Dynamical Systems*, Menlo Park, Ca.: Benjamin/Cummings

Denker, M., and Keller, G. (1986) Rigorous statistical procedures for data from dynamical systems, *J. of Statistical Physics*, Vol. 44, No. 1, 67-93

Eggleston, M. (1949) The fractional dimension of a set defined by decimal properties, *Quart. J. Math. Oxford Ser. 20*, 31-36

Farmer J., Ott, E., and Yorke, J. (1983) The dimension of chaotic attractors, *Physica 7D*, 153-180

Feret, P. (1990) Personal communication

Frisch, U., Sulem, P., and Nelkin, M. (1978) A simple dynamical model of intermittent fully developed turbulence, *J. Fluid Mechanics*, **87**, 719-736

Gardner, R., Milne, B. Turner, M. and O'Neill, R. (1987) Neutral modes for the analysis of broad-scale landscape pattern, *Landscape Ecology*, Vol. 1, No. 1, 19-28

Goldberger, A., Findley, L., Blackburn, M., Mandell, A. (1984) Nonlinear dynamics in heart failure: Implications of long-wavelength cardiopulmonary oscillations, *Amer. Heart J.*, **107**, 612-615

Good, I. J. (1941) The fractional dimensional theory of continued fractions, *Proc. Cambridge Phil. Soc.*, **37**, 199-228

Halse, T., Meakin, P., and Procaccia, I. (1986) Scaling structure of the surface layer of diffusion-limited aggregates, *Physical Review Letters*, Vol. 56, No. 8, 854-857

Hawkes, J. (1974) Hausdorff measure, entropy, and the independence of small sets, *Proc. London Math. Soc.* (3) **28**, 700-724

Hentschel H., and Procaccia, I. (1982) Intermittency exponent in fractally homogeneous turbulence, *Physical Review Letters*, Vol. 49, No. 16, 1158-1161

Hurewicz, W. (1948) *Dimension Theory*, Princeton: Princeton Univ. Press

Johnson, N. and Kotz, S. (1977) *Urn Models and Their Applications*, New York: Wiley

Johnson, L. and Reiss, R. (1982) *Numerical Analysis, 2nd ed.*, New York: Addison-Wesley

Kapitalnik, A. and Deutscher, G. (1982) Percolation characteristics in discontinuous thin films of Pb, *Physical Review Letters*, Vol. 49, No. 19, 1444-1448

Kaplan, J. and Yorke, J. (1978) Chaotic behavior of multidimension difference equations, *Functional Differential Equations and the Approximation of Fixed Points, Proceedings,* Bonn, Peitgen and Walther, eds.

Kolmogorov, A. and Tihomorov, V. (1959-1961) Epsilon-entropy and epsilon-capacity of sets in functional spaces, *Translations Amer. Math. Soc. (Ser. 2)* **17**, 277-364

Krummel, J. Gardner, R., Sugihara, G. O'Heill, R. and Coleman, P. (1987) Landscape patterns in a disturbed environment, *Oikos*, **48**, 321-324

Lewis, M. and Rees, D. (1985) Fractal surfaces of proteins, *Science*, **230**, 1163-1165

Lewontin, R. (1974) *The Genetic Basis of Evolutionary Change*, New York: Columbia University Press

Mandelbrot, B. B. (1963) The variation of certain speculative prices, *J. of Business* XXXVI, 394-419

Mandelbrot, B. B. (1969) Long-run linearity, locally Gaussian processes, H spectra and infinite variance, *Int. Econ. Review*, **10**, 82-111

Mandelbrot, B. B. (1971) A fast fractional gaussian noise generator, *Water Resources Research*, **7**, 543-553

Mandelbrot, B. B. (1977) *Fractals: Form, Chance, and Dimension*, San Francisco: Freeman

Mandelbrot, B. B. (1982) *Geometry of Nature*, San Francisco: Freeman

Mandelbrot, B. B. and Van Ness, J. (1968) Fractional Brownian motions, fractional noises, and applications, *SIAM Review*, **10**, 422-437

Mandelbrot, B. B. and McCamy, J (1970) On the secular pole motion and the Chandler wobble, *Geophys. J. Roy. Astron. Soc.* **21**, 217-232

Mandelbrot, B. B. and Wallis, J. (1969) Some long run properties of geophysical records, *Water Resources Research*, **5(2)**, 321-340

Marcus, M. B. (1976) Capacity of level sets of certain stochastic processes *Zeitschrift fur Wahrscheinlichkeitstheorie und Verwandte Gebiete*, **34**, 279-284

Mardia, K., Kent, J., and Bibby J. (1979) *Multivariate Analysis*, New York: Academic Press

Mark, D. M. and Aronson, P. B. (1984) Scale-dependent fractal dimensions of topographic surfaces: an empirical investigation, with applications in geomorphology and computer mapping, *J. of Math. Geology*, Vol. 15, No. 7, 671-683

Morse, D., Lawton, J., Dodson, M. and Williamson, M. (1985) Fractal dimension of vegetation and the distribution of arthropod body lengths, *Nature*, **314**, 731-733

O'Brien, T. (1986) A discrete time option model dependent on expected return: A note, *J. of Finance*, Vol. XLI, No. 2, 515-520

Orey, S. (1970) Gaussian sample functions and the Hausdorff dimension of level crossings, *Zeitschrift fur Wahrscheinlichkeitstheorie und Verwandte Gebiete*, **15**, 249-256

Ott, E., Sauer, D., and Yorke, J. (1989) Lyapunov partition functions for the dimensions of chaotic sets, *Physical Review Letters*, Vol. 39, No. 8, 4212-4222

Passamante, A., Hediger, T., and Gollub, M. (1989) Fractal dimension and local intrinsinc dimension, *Physical Review Letters: A. General Physics*, Series 3, Vol. 39, No. 7, 3640-3645

Peitgen, H. O. and Richter, P. (1986) *The Beauty of Fractals*, Berlin: Springer

Peitgen, H. O. and Saupe, D. (1988) *The Science of Fractal Images*, New York: Springer-Verlag

Ramsey, P. (1989) Personal communication

Rogers, C. A. (1970) *Hausdorff Measures*, Cambridge: Cambridge Univ. Press

Rothrock D., and Thorndike, A. (1980) Geometric properties of underside sea ice, *J. of Geophysical Research*, Vol. 85, No. C7, 3955-3963

Ruelle, D. (1982) Repellers for real analytic maps, *Ergod. Theory and Dynamical Systems*, **2**, 99-108

Ryszard, E. (1978) *Dimension Theory*, New York: North-Holland Pub. Co.

Saupe, D. (1987) Efficient computation of Julia sets and their fractal dimension *Physica 28D*, 358-370

Saupe, D. (1988) Algorithms for random fractals, in *The Science of Fractal Images*, Peitgen, H.O., and Saupe, D., eds., New York: Springer-Verlag

Schaffer, W. (1984) Stretching and folding in lynx fur returns: evidence for a strange attractor in nature? *The American Naturalist* Vol. 124, No. 6, 796 - 820

Schaffer, W. and Kot, M. (1985a) Nearly one dimensional dynamics of an epidemic, *Journal of Theoretical Biology* **112,** 403-427

Schaffer, W. and Kot, M. (1985b) Do strange attractors govern ecological systems? *Bioscience*, **35**, 342-350

Smith, C. (1976) Option Pricing, *J. of Financial Economics*, **3**, 616-664

Terrell, G. and Scott, D. (1985) Oversmoothed nonparametric density estimates *J. of the American Statistical Association*, Vol. 80, No. 389, 209-214

Viczek, T. (1989) *Fractal Growth Phenomena*, New Jersey: World Scientific

Voss, R. (1988) Fractals in nature: From characterization to simulation, in *The Science of Fractal Images*, Peitgen, H.O., and Saupe, D., eds., New York: Springer-Verlag

Voss, R., Laibowitz, R., and Allessundrini, E. (1982) Fractal (scaling) clusters in thin gold films near the percolation threshold, *Physical Review Letters,* Vol. 49, No. 19, 1441-1444

Appendix A

In this appendix, I will evaluate several competing and related methods of estimating the Hausdorff dimension of a set given a random sample of points on the set. All sets discussed in this appendix are subsets of the interval [0, 1]. The data to be analyzed by the dimension estimating program (DEP) are generated by a pseudo-random number generator for three different dimensions, 0.4307, 0.6309, and 1, corresponding to the sets $C_{2,5}$, $C_{2,3}$, and [0, 1]. The data are then generated (and analyzed) as in a repeated measures design. However, the data are not independent, but highly dependent, and so multivariate methods must be employed to analyze the data.

Some of the questions to be answered in this appendix are the following:

(i) Are the estimators unbiased?

(ii) Do the estimators have the same bias?

(iii) What do the dispersion parameters look like?

We want to evaluate the expression

$$\lim_{n\to\infty} n^{-1} \limsup_{k\to\infty} \left[\log \eta_k(A) - \log \eta_k(u_n(x) \cap A)\right].$$

In order to evaluate the formula, we must extrapolate to the limit. Several methods are considered here. The methods are outlined in Table A.1 below. L_1 stands for regression minimizing the absolute deviations, L_2 stands for regression minimizing the squared deviations, and Richardson extrapolation is really polynomial interpolation, using the intercept as the limit. For the outside limit, regression on 1/n will be called "1/n" regression, and regression on n will be called "n" regression.

The original data generated do not follow any ellipsoidal distribution--the support is a Cantor set. However, after classifying the data into intervals, and performing several calculations, including many additions and subtractions, I believe that the data may in fact follow a multinormal distribution closely enough that the normal-theory inference procedures should work. To test this out, I did a test for multinormality, as given in Mardia, Kent, and Bibby (1979). The skewness and kurtosis statistics were computed, because they give sensitivity of the tests on means and variances, respectively, to non-normality. Under the null hypothesis of normality, the skewness statistic is approximately χ^2 with 56 degrees of freedom, and the kurtosis statistic is approximately $N(48, 4.38^2)$. See Table A.2 for results.

The kurtosis is fairly close to expected, so inference on dispersion parameters is valid for all three groups of data. However, the $C_{2,5}$ set has a significant skewness, so that inferences on the means regarding the $C_{2,5}$ set must be regarded with some suspicion.

To check out bias of the different estimation methods, some Hotelling T^2 tests were performed. Because the data turned out to be heteroscedastic, each group was analyzed separately. The results are reported in Table A.3.

From Table A.3, we see that all the estimators are biased, but the bias is less than the classical entropy method. Also, the different methods produce different biases. The means of the different methods are shown in Table A.4 in terms of the excess over the dimension they are estimating. Notice that the "n" regression method is better than the "1/n" method. Is the extrapolation method important? Doing some Hotelling T^2 to test for equality of extrapolation methods, we get the results in Table A.5.

This shows that the extrapolation method is important. So, looking at the means in Table A.4 above, it appears that Richardson extrapolation is best.

To summarize:

- Richardson extrapolation for the inside limit
- "n" regression for the outside limit

Looking at the histogram of the estimates of the dimension generated from this procedure, it does appear that the data are close enough to normal for normal theory procedures to work. The histogram can be found in Fig. A.1.

Table A.1. Methods of evaluting limits. The combinations in Table A.1 give the six methods that will be investigated.

Inside limit	Outside limit	
Richardson	L_2	regress limits on 1/n
L_1		weights $\sqrt{n}$
L_2	L_2	regress limits on n

Table A.2. Results of test for multinormality.

Group	Skewness	p-value	Kurtosis
$C_{2,5}$	101.2	0.000206	49.9
$C_{2,3}$	61.9	0.273736	43.0
[0, 1]	40.1	0.946238	37.0

Table A.3. Some tests by group.

Group	Estimates unbiased? T^2	p-value	Same bias? T^2	p-value
$C_{2,5}$	388	$\ll 0.01$	89	< 0.01
$C_{2,3}$	41.9	< 0.01	27.1	≈ 0.02
[0, 1]	52.8	< 0.01	36.0	< 0.01

Table A.4. Sample means of the different methods (adjusted for true dimension).

Group	Richardson		L_1		L_2	
	1/n	n	1/n	n	1/n	n
$C_{2,5}$	0.0065	0.0267	0.0357	0.462	0.0353	0.0476
$C_{2,3}$	0.0817	0.572	0.0829	0.0513	0.0852	0.0541
[0, 1]	-0.0419	-0.036	-0.467	-0.0557	-0.0443	-0.512

Table A.5. Comparison of extrapolation methods.

Group	$C_{2,5}$	$C_{2,3}$	[0, 1]
T^2	73.4**	13.03	23.6*

df=19, p=4

*Significant at the 0.01 level

**Significant at almost any level

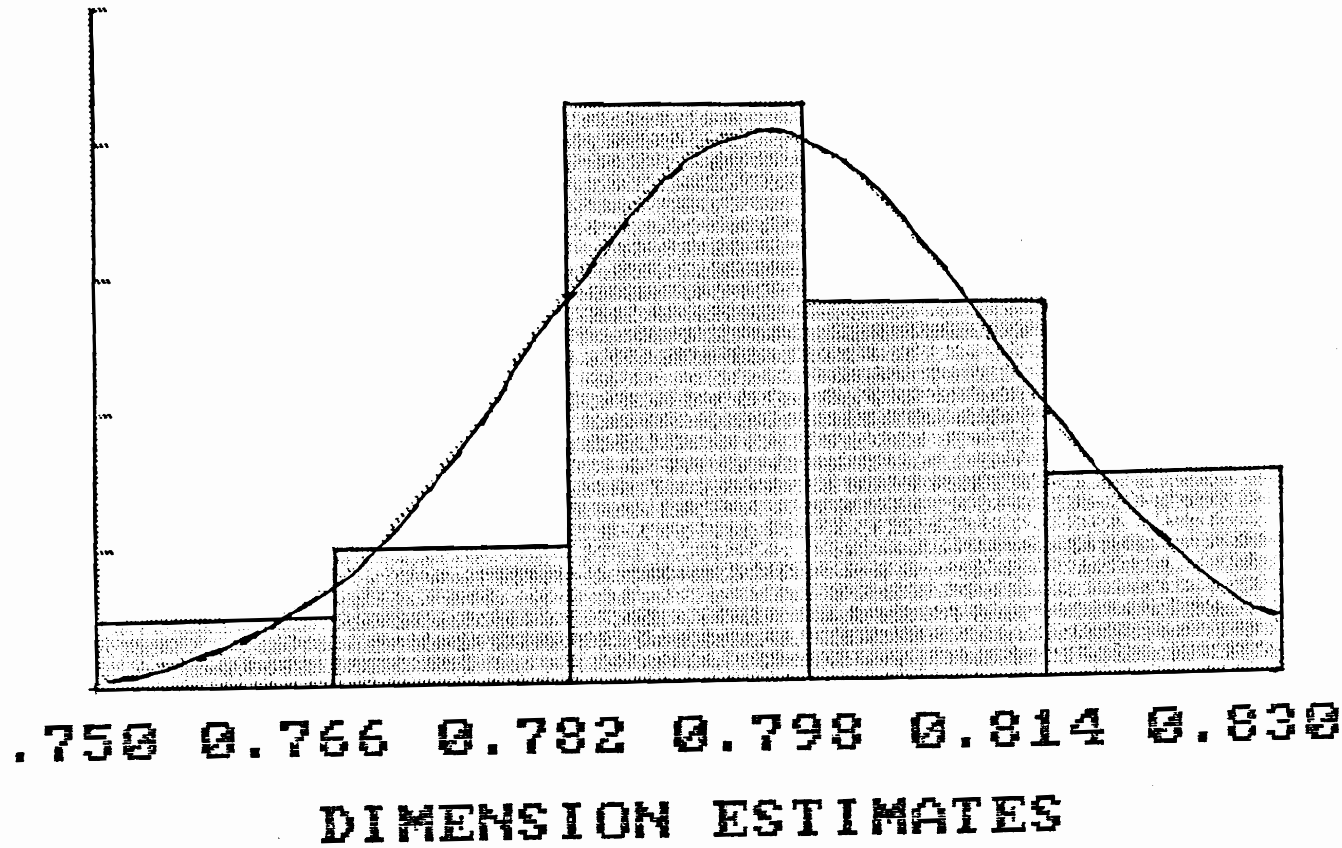

Figure A.1. Histogram of the estimate of the Hausdorff dimension, calculated by using Richardson extrapolation on the inside limit, and L_2-regression on n for the outside limit. An optimal number of intervals for constructing an histogram was used (Terrell and Scott 1985). A χ^2 test for Normality was performed with 2 degrees of freedom, and a p-value of 0.32465 was computed.

Appendix B

The program in this appendix estimates the Haudorff dimension using the nonparametric approach of Chapter 2. A description of how the program works follows, and then a listing of the program appears at the end of this appendix.

The program is written in Turbo Pascal^{TM} because Pascal handles truly dynamic variables very nicely. Other languages (such as QuickBASIC^{TM}) have what should truly be called quasi-dynamic variables.

Suppose the data are e-dimensional. Then the data are stored in an 2^e-leafed tree; i.e., there are 2^e branches coming out of each node. Each node represents a net set: an order n node represents an order n net set, and in each node is stored the number of order k net sets that are occupied ($\eta_k(A_n)$ for some appropriate A_n.)

The tree is deepened one level at a time until the number of leaves exceeds 90% of the sample size (find k so that $\eta_k(A) \geq 0.90$ t.)

The relative entropy estimate with averaging is computed by moving up and down the tree, adding the logarithms of the counts across each level. The sum is then divided by the number of nodes comprising the sum. Of course, the logarithms are computed to the base 2.

Standard errors of the relative entropy estimator are then computed using equation (2.21). The program ends by performing a weighted least squares regression, and displaying the slope as the estimate.

After this is done, the user is then allowed the choice of dropping one data point from the regression (one of the two ends).

A listing of typical output follows on the next page along with an explanation of the various values outputted. On the page after is a listing of the source code, written in Pascal and implemented on a Tandy 1000 personal computer.

Typical output. Superscripts denote items to be explained below.

Data has been read in. Read in 1000[1] data points.

Data has been scaled.

The tree has been built, and it contains 316 nodes.

It is 6[2] levels deep.

Checked 10 (out of 316) nodes.[3]

Checked 20 (out of 316) nodes.

⋮

Checked 310 (out of 316) nodes.

<table>
<tr><th>n</th><th>rel. entropy</th><th></th><th>std. error</th><th>residual</th><th>z-value</th></tr>
<tr><td>2</td><td>2.512|[4]</td><td>1.256[5]</td><td>0.14837[6]</td><td>-0.21674[7]</td><td>-1.46080[8]</td></tr>
<tr><td>3</td><td>3.645|</td><td>1.215</td><td>0.14795</td><td>-0.04795</td><td>-0.31447</td></tr>
<tr><td>4</td><td>4.952|</td><td>1.238</td><td>0.14154</td><td>0.29422</td><td>2.07871</td></tr>
<tr><td>5</td><td>5.835|</td><td>1.167</td><td>0.12331</td><td>0.27186</td><td>2.20460</td></tr>
<tr><td>6</td><td>1.043|</td><td>1.043</td><td>0.11524</td><td>-0.31349</td><td>-2.72034</td></tr>
</table>

The slope is 0.9442[9] +/- 0.09816[10]

Is the fit OK (y/n)? <u>n</u>

Drop bottom or top (b/t)? <u>b</u>

Underlined type is user's response. The regression would be rerun, but omitting observation n=6.

1. The sample size, t. 2. The depth of the tree: k so that $\eta_k(A) \geq 0.9t$. 3. Progress of computer. 4. For each n, $\log \eta_k(A) - \overline{\log V_2}$. 5. For each n, $\hat{\alpha}_n = n^{-1}[\log \eta_k(A) - \overline{\log V_2}]$. 6. The standard error for $n\hat{\alpha}_n$. 7. $r_n = n\hat{\alpha}_n -$regression line. 8. $r_n \div$ standard error. 9. Slope of regression line. 10. Mean squared error of residuals of regression line divided by $[\sum (n - \bar{n})^2]^{1/2}$.

HAUSDORF.PAS

```
program hausdorf;

const    e = 2;            { Dimension of space object is embedded in }
        num_dir = 4;      { Number of directions to go: 2^e }
        max_data = 1000;  { Maximum number of data points allowed }
        max_levels = 50;  { Maximun levels of the tree }
        ln2 = 0.6931471;  { Natural log of 2 }

type     tree = ^net;
        net = record
                count : integer;
                root : tree;                      { Points backwards }
                direction : array[1..num_dir] of tree; { Points forwards  }
            end;

        raw_data = array[1..max_data, 1..e] of real;

        twig_data = array[1..max_data] of tree;
        entropies = array[1..max_levels] of real;

function log2(x:integer):real;
begin
   log2 := ln(x) / ln2
end;

function log2real(x:real):real;
begin
   log2real := ln(x) / ln2
end;

function power(base : real; exponent : integer) : real;
```

```
begin
   power := exp ( exponent * ln(base))
end;

function two_to_the(n:integer):real;
begin
   two_to_the := exp(n * ln2)
end;

procedure read_data( var  x : raw_data;
                     var  n : integer);
{ x contains the data.  The sample size is n. }
var       i : integer;
     f_name : string[65];
      input : text;

begin
   n := 0;
   write('In which file is the data?   ');
   readln(f_name);
   assign(input,f_name);
   reset(input);
   while not(eof(input)) and (n < max_data) do
   begin
      n := n + 1;
      for i:=1 to e-1 do read(input, x[n,i]);
      readln(input, x[n,e])
   end;
end; { read_data }

procedure scale_data( var  x : raw_data;
                           n : integer);
{ x contains the data.  The sample size is n. }
```

```
var     i, j : integer;
        l, u : real;

begin
   for j:=1 to e do begin
      l := x[1,j];
      u := x[1,j];
      for i:=1 to n do
         if x[i,j] < l then
            l := x[i,j]
         else
            if x[i,j] > u then
               u := x[i,j];
      for i:=1 to n do
         if l<u then
            x[i,j] := (x[i,j]-l)/(u-l)
         else
            x[i,j] := 0.0
   end; { for }
end; { scale_data }

procedure new_tree(sample_size: integer;
           var  head_node  : tree;
           var  twig       : twig_data);

var i:integer;

begin
   new(head_node);
   head_node^.count := 1;
   head_node^.root := nil;
   for i:=1 to num_dir do begin
      head_node^.direction[i] := nil;
```

```
   end;
   for i:=1 to sample_size do twig[i] := head_node;
end;

procedure add_level(sample_size : integer;
                   head_node   : tree;
               var x           : raw_data;
               var twig        : twig_data;
               var num_node    : integer);

{ This procedure adds a level to an already existing
  tree. It deposits the proper counts in the appropriate
  place in the tree. }

var     curr_node, new_root : tree;
      i, j, dir_to_go : integer;

    function which_dir(i : integer) : integer;  { Returns the direction }
    var j, k, temp : integer;                    { in which to proceed   }
    begin                                        { next for building the }
       temp := 1;                                { tree. SIDE EFFECT:    }
       for j:=1 to e do begin                    { X array has 2*X mod 1 }
          x[i,j] := 2.0 * x[i,j];                { after execution.      }
          if int(x[i,j])>=1 then begin
             temp := temp + round(two_to_the(j-1));
             x[i,j] := x[i,j] - 1.0
          end
       end;
       which_dir := temp;
    end;  { function which_dir }

begin  { add_level }
   for i:=1 to sample_size do begin        { Resets all the counts }
```

```
      if twig[i]^.count > 0 then begin    { back to zero.          }
         curr_node := twig[i];
         while curr_node^.count > 0 do begin
            curr_node^.count := 0;
            curr_node := curr_node^.root
         end
      end
   end;                                   { End of reset.          }

   for i:=1 to sample_size do begin  { Begin to add next level of nodes }
      dir_to_go := which_dir(i);     { For this data point, which dir? }
      if twig[i]^.direction[dir_to_go] = nil then begin  { If new node }
         num_node := num_node + 1;
         new_root := twig[i];
         new(twig[i]);                          { Get new node }
         new_root^.direction[dir_to_go] := twig[i]; { Record node  }
         twig[i]^.count := 1;                   { One box occupied }
         twig[i]^.root := new_root;                { Set root properly }
         for j:=1 to num_dir do twig[i]^.direction[j] := nil;
                                          { Initialize directions. }
         curr_node := twig[i];                  { Add 1 to each count in }
         while curr_node^.root <> nil do begin  { tree below that twig.  }
            curr_node := curr_node^.root;
            curr_node^.count := curr_node^.count + 1
            end { while }
         end { then }
      else                                      { Already made twig. }
         twig[i] := twig[i]^.direction[dir_to_go]; { Record new twig.   }
   end; { for }
end; { add_level }

procedure get_alphas(head_node    : tree;
                     depth        : integer;
```

```
            sample_size : integer;
        var alpha      : entropies;
        var stderr     : entropies;
            num_nodes  : integer);

{ This procedure outputs the alpha hats and their standard errors,
  using the data in the tree pointed to by HEAD_NODE.
  The tree is DEPTH levels.  Alpha hat constructed with averaging
  (over space, not data.) }

var  stack_ptr : array [1..max_levels] of tree;
     stack_dir : array [1..max_levels] of integer;
     stack_lvl, n, dir_to_go, node_count : integer;
     log_tot : real;
     curr_node : tree;
     b : array [1..max_levels] of integer;

   procedure push( ptr : tree;  dir : integer );
   begin
      stack_lvl := stack_lvl + 1;
      stack_ptr[stack_lvl] := ptr;
      stack_dir[stack_lvl] := dir
   end;

   procedure pop(var ptr : tree; var  dir : integer);
   begin
      ptr := stack_ptr[stack_lvl];
      dir := stack_dir[stack_lvl];
      stack_lvl := stack_lvl - 1
   end;

   function next_dir(curr_node : tree; dir_to_go : integer) : integer;
   var i : integer;
```

```
   all_done : boolean;
begin
   i := dir_to_go + 1;
   if i<=num_dir then
      all_done := false
   else
      all_done := true;
   while (curr_node^.direction[i] = nil) and not(all_done) do begin
      i := i + 1;                        { Keep incrementing i     }
      if i>num_dir then all_done := true  { until i is too big, or }
   end; { while }                        { valid direction found. }
   if all_done then
      next_dir := 0
   else
      next_dir := i
end; { function next_dir }

function variance(m2 : integer) : real;
{ This function outputs the variance of an estimator }
var     m, m1 : integer;
        e1, s1, e2, s2, e, c : real;

begin
   m  := head_node^.count;
   m1 := m - m2;
   e1 := m * ( 1 - power(1-1/m,sample_size));
   s1 := m1 * (m1-1) * power(1-2/m, sample_size)
      + m1 * power(1-1/m, sample_size)
      - sqr(m1) * power(1-1/m, 2*sample_size);
   e2 := m2 * ( 1 - power((1 - 1/m),sample_size) );
   s2 := m2 * (m2-1) * power(1-2/m, sample_size)
      + m2 * power(1-1/m, sample_size)
      - sqr(m2) * power(1-1/m, 2*sample_size);
```

```
        e  := e1 + e2;
         c  := m2 * m1
            * ( power(1-2/m,sample_size) - power(1-1/m,2*sample_size) );
        if m2 > 0 then
            variance := ( s1 / sqr(e1)  +  s2 / sqr(e2)  -
                          2 * ( c + s2) / e / e2 ) / sqr(ln2)
        else
            variance := s1 / sqr(e1) / sqr(ln2);
    end; { function variance }

begin  { procedure get_alphas }
    stack_lvl := 0;
    node_count := 0;
    push(head_node,0);
    for n:=2 to depth do begin
        b[n] := 0;
        alpha[n] := 0.0;
        stderr[n] := 0.0
    end; { for }
    dir_to_go := 0;
    curr_node := head_node;
    while stack_lvl > 0 do begin
        dir_to_go := next_dir(curr_node, dir_to_go);
        if dir_to_go = 0 then
            pop(curr_node,dir_to_go)
        else begin
            push(curr_node,dir_to_go);
            curr_node := curr_node^.direction[dir_to_go];
            dir_to_go := 0;
            alpha[stack_lvl] := alpha[stack_lvl] + log2(curr_node^.count);
            if stack_lvl < depth then
                stderr[stack_lvl] := stderr[stack_lvl] +
                                     variance(curr_node^.count)
```

```
        else
          stderr[stack_lvl] := stderr[stack_lvl] + variance(0);
        b[stack_lvl] := b[stack_lvl] + 1;
        node_count := node_count + 1;
        if (node_count div 10 * 10) = node_count then   { Every 10 nodes, }
                                                        { report progress }
            writeln('Checked ',node_count,' (out of ',num_nodes,
                              ') nodes.');
      end; { if }
   end; { while }
   log_tot := log2(head_node^.count);
   for n:=2 to depth do begin
      alpha[n] := log_tot - alpha[n]/b[n];
      stderr[n] := sqrt(stderr[n]) / b[n];
   end; { for }
end;

procedure make_tree(  sample_size : integer;
            var   x           : raw_data;
            var   head_node   : tree;
            var   levels      : integer;
            var   num_node    : integer);

var    twig  : twig_data;
       limit : integer;

begin
   new_tree(sample_size,head_node,twig);   { Create a new tree }

   num_node := 1;                          { One node created }

   limit := sample_size * 9 div 10;   { If number of nodes gets within
                            90% of sample size, that's good enough }
```

```
  levels := 1;
  while limit >= head_node^.count do begin        { Keep adding levels    }
    add_level(sample_size,
              head_node, x, twig, num_node);   { until number of nodes }
    levels := levels + 1                       { exceeds LIMIT.        }
  end;
end; { make_tree }

procedure l2_regression(  y  : entropies;
                          x  : entropies;
        { weight -> }     w  : entropies;  { Actually, sqrt of weight }
                          depth       : integer;
                      var slope       : real;
                      var stderr_slope : real);

{ Data points are in positions 2, 3, ..., depth. The weights in w[.] are
  reciprocals of the standard deviations of the y's.                    }

var     y_star, x_star : entropies;
      y_bar, x_bar, sum_w, syy, ssx, sxy,  sigma_2 : real;
      i, bot, top : integer;
      good_fit : boolean;
      q : char;

  function entrp( y : real; i : integer): real;
  begin
    entrp := y / i
  end;

begin  { l2_regression }
 good_fit := false;
 top := depth;
 bot := 2;
```

```
repeat

  ssx := 0;       sxy   := 0;       x_bar := 0;    { Initialization }
  syy := 0;       y_bar := 0;        sum_w := 0;

  for i:=bot to top do begin
     sum_w := sum_w + w[i];
     y_bar := y_bar + y[i];
     x_bar := x_bar + x[i]
  end;
  y_bar := y_bar / (top - bot + 1);
  x_bar := x_bar / (top - bot + 1);
  for i:=bot to top do begin
     y_star[i] := (y[i] - y_bar) * (top - bot + 1) * w[i] / sum_w;
     x_star[i] := (x[i] - x_bar) * (top - bot + 1) * w[i] / sum_w;
  end;

  for i:=bot to top do begin
     syy := syy + sqr(y_star[i]);
     ssx := ssx + sqr(x_star[i]);
     sxy := sxy + x_star[i] * y_star[i]
  end;
  slope := sxy / ssx;
  sigma_2 := (syy - sqr(sxy)/ssx) / (top-bot-1);
  stderr_slope := sqrt(sigma_2/ssx);

 { Write out the results that have been accomplished so far }
  writeln(' ');
  writeln('In summary of what has been done (rel. entropy = c + dim n):');
  writeln(' ');
  writeln('n':4,' ','rel. entropy':15,' ','std. error':15,
          ' ','residual':15,' ','z-value':15);
  for i:=bot to top do
```

```
        writeln(i:4,'  ',y[i]:7:3,'|',entrp(y[i],i):7:3,
                '  ',1/w[i]:15:5,
                (y_star[i]-slope*x_star[i]):15:5,'  ',
                ((y_star[i]-slope*x_star[i])*w[i]):15:5);

    writeln(' ');
    writeln('The slope is ',slope:7:4,' +/- ',stderr_slope:8:5);
    writeln(' ');
    write('Is the fit OK (y/n)? ');
    readln(q);                             { This section asks and      }
    if (q='y') or (q='Y') then             { decides whether to drop    }
      good_fit := true                     { an observation. It drops   }
    else begin                             { on the ends of the data.   }
      write('Drop bottom or top (b/t)? '); { The top is the largest     }
      readln(q);                           { observation. However, on   }
      if (q='t') or (q='T') then           { screen it reverses, so I   }
        bot := bot + 1                     { must reverse the question  }
      else                                 { so it makes sense.         }
        top := top - 1;
      good_fit := false  { just to make sure it repeats }
      end;
  until good_fit or (top-bot <= 1);

end; { l2_regression }

procedure do_it;
var   num_nodes, sample_size, i, levels : integer;
      x : raw_data;
      head_node : tree;
      alpha, stderr, weight, regressor : entropies;
      estimate, error : real;

begin { do_it }
```

```
    read_data(x, sample_size);
    writeln('Data has been read in.  Read in ',sample_size,' data points.');
    scale_data(x, sample_size);
    writeln('Data has been scaled.');
    make_tree(sample_size, x, head_node, levels, num_nodes);
    writeln(' ');
    write('Tree for storing data has been created, and it contains ');
    writeln(num_nodes,' nodes.');
    writeln('It is ',levels,' levels deep.'); writeln(' ');
    get_alphas(head_node, levels, sample_size, alpha, stderr, num_nodes);
    for i:=2 to levels do begin
        weight[i] := 1/stderr[i];
        regressor[i] := i
    end;
    l2_regression( alpha, regressor, weight, levels, estimate, error);
end; { do_it }

begin {main}
    do_it
end.
```

Appendix C

The purpose of this appendix is to establish the asymptotic normality of the vector (V_1, V_2). I will make the same assumptions relating sample size t to number of urns in the two groups as was made in Chapter 2. The assumptions are that the conditions (i) and (ii) below hold.

(i) $t/m_i \to \infty$, i=1, 2, and $t/m \to \infty$

(ii) $m_i \exp(-t/m_i) \to \infty$, i=1, 2, and $m \exp(-t/m) \to \infty$

The following proposition (Billingsley 1968, pp. 29) will enable us to infer the asymptotic joint distribution from the asymptotic marginal distributions.

Notation. Convergence in distribution is denoted by $\to_{\mathcal{D}}$. Convergence in probability is denoted by $\to_{\mathcal{P}}$.

Proposition C.1. Suppose that X and Y are independent. If $X_t \to_{\mathcal{D}} X$ and $Y_t \to_{\mathcal{P}} Y'$, where Y' has the same distribution as Y, and if

$$P\{X_t \in A, Y_t \in B\} \to P\{X \in A\}\, P\{Y' \in B\} \tag{C.1}$$

for all Borel sets A and B, then $(X_t, Y_t) \to_{\mathcal{D}} (X, Y)$.

Now, suppose there are m urns, divided into two groups. The first group has m_1 urns, the second has m_2 urns. Hence, $m = m_1 + m_2$. A total of t marbles are thrown into the urns, each urn having equal probability 1/m of receiving a marble. Suppose t_1 marbles land in the first group of urns, while t_2 land in the second. So, we must have $t = t_1 + t_2$. Notice that t_i is distributed as a binomial variable with parameters m_i/m and t. The standardized number of occupied urns will be called $Z_i = (V_i - E(V_i)) / (Var(V_i))^{1/2}$. Johnson and Kotz (1977) give numerous conditions for which $V_1 + V_2$ is asymptotically normal. I will use their results to show that (V_1, V_2) is asymptotically bivariate normal. From Johnson and Kotz (1977) we get Proposition C.2.

Proposition C.2. If assumptions (i) and (ii) above are satisfied, then $V_1 + V_2$ is asymptotically normal.

Before the main theorem of this appendix is proved, a lemma is needed.

Lemma C.3. Assume that conditions (i) and (ii) hold. Then

$$\lambda_i \equiv P[\, t_i/m_i \le q_{i1},\quad m_i \exp(-t_i/m_i) \le q_{i2}\,] \to 0, \quad i=1,\, 2.$$

Proof. We have the inequality

$$\lambda_i \le P(\, t_i/m_i \le q_{i1}\,).$$

Consider $\lambda_{i1} \equiv P(\, t_i/m_i \le q_{i1}\,)$. Since t_i is distributed Binomial(m_i/m, t), $E(t_i/m_i)= t/m$, and $Var(t_i/m_i)=(t/m)\cdot m_{3-i}/(m_i\, m)$. Hence, using Chebyshev's inequality,

$$P(\, t_i/m_i \le q_{i1}\,) \le P\left(\left| \frac{t_i/m_i - E(t_i/m_i)}{\sqrt{Var(t_i/m_i)}} \right| \ge \left| \frac{q_{i1} - t/m}{\sqrt{(t/m)\cdot m_{3-i}/(m_i\, m)}} \right| \right)$$

$$\le \frac{\frac{t}{m}\cdot\frac{m_{3-i}}{m_i\, m}}{(q_{i1} - t/m)^2} = \frac{\frac{m_{3-i}}{m_i\, m}}{\frac{t}{m}\left(\frac{q_{i1}}{t/m} - 1\right)^2} \to 0,$$

since $q_{i1} \div (t/m) \to 0$ and $m_{3-i}/m_i\, m \le 1$. Therefore, $\lambda_i \to 0$. □

We are now ready to state the theorem.

Theorem C.4. Suppose there exists a random variable Y so that $Z_2 \to_{\mathfrak{D}} Y$, where Y has a standard normal distribution. Let N_1 and N_2 be independent, standard normal variables. Then $(Z_1,\, Z_2) \to_{\mathfrak{D}} (N_1,\, N_2)$, where Z_i is V_i standardized to have mean 0 and variance 1.

Proof. From the remarks at the end of section 2.4, the limiting variables of the vector $(Z_1,\, Z_2)$ are uncorrelated. So, if they are jointly normal, then they are independent. If (C.1) is established, the theorem will be proved, by Proposition C.1. The remainder of this proof establishes the condition (C.1).

Let $\mathfrak{B}$ be the Borel sets on the real line. Given $0 < \epsilon < 1$. Let $\epsilon' = \epsilon/5$. Now, let $b(\cdot)$ be the binomial probabilities

$$b(x) = \binom{t}{x}\left(\frac{m_1}{m}\right)^x \left(\frac{m_2}{m}\right)^{t-x}.$$

Conditioning on t_1, we obtain

$$P(Z_1 \in A_1, Z_2 \in A_2) = \sum_{t_1=0}^{t} P(Z_1 \in A_1 \mid t_1) \, P(Z_2 \in A_2 \mid t_2) \, b(t_1),$$

(recalling that $t_2 = t - t_1$,) since V_1 and V_2 are conditionally independent (Johnson and Kotz 1977). Define $\mathcal{A}_i = \mathcal{A}_i(m_i^*, t_i^*)$ to be

$$\mathcal{A}_i = \{ (m_i, t_i) : t_i/m_i \geq t_i^*/m_i^*, \; m_i \exp(-t_i/m_i) \geq m_i^* \exp(-t_i^*/m_i^*) \}.$$

By Proposition C.2, with V_i in place of $V_1 + V_2$, there exists (m_i^*, t_i^*) so that, for all $(m_i, t_i) \in \mathcal{A}_i$,

$$| P(Z_i \in A_i \mid t_i) - P(N_i \in A_i) | \leq \epsilon' \text{ for all } A_i \in \mathfrak{B}, \; i=1, 2. \qquad \text{(C.2)}$$

By Lemma C.3, there exists (m^*, t^*) so that

$$t/m \geq t^*/m^*, \quad m \exp(-t/m) \geq m^* \exp(-t^*/m^*)$$

implies that $P(\mathcal{A}_i^c) \leq \epsilon'$. It is easy to verify, upon noting that $P(N_i \in A_i) \leq 1$ and $\epsilon'^2 \leq \epsilon$, that equation (C.2) above implies

$$| P(Z_1 \in A_1 \mid t_1) \, P(Z_2 \in A_2 \mid t_2) - P(N_1 \in A_1) \, P(N_2 \in A_2) | \leq 3\epsilon'.$$

Then,

$$| P(Z_1 \in A_1, Z_2 \in A_2) - P(N_1 \in A_1) \, P(N_2 \in A_2) |$$

$$\leq \sum_{t_1 \in \mathcal{A}_1^c \cup \mathcal{A}_2^c} | P(Z_1 \in A_1 \mid t_1) \, P(Z_2 \in A_2 \mid t_2) - P(N_1 \in A_1) \, P(N_2 \in A_2) | \, b(t_1)$$

$$+ \sum_{t_1 \in \mathcal{A}_1 \cap \mathcal{A}_2} | P(Z_1 \in A_1 \mid t_1) P(Z_2 \in A_2 \mid t_2) - P(N_1 \in A_1) P(N_2 \in A_2) | b(t_1)$$

$$\leq \sum_{t_1 \in \mathcal{A}_1^c \cup \mathcal{A}_2^c} 1 \cdot b(t_1) + \sum_{t_1 \in \mathcal{A}_1 \cap \mathcal{A}_2} (3\epsilon') b(t_1) \leq 2\epsilon' + 3\epsilon' = \epsilon.$$

Since ϵ was arbitrary, we have established (C.1), and the proof is complete. □

VITA

Russell Lynn Reeve, son of Mr. Dale T. Reeve and Mrs. LaVon M. Reeve, was born in Bountiful, Utah, on July 6, 1963.

He graduated from Viewmont High School in Bountiful, Utah, in 1981. He received a B.S. degree with a double major in mathematics and statistics in 1985 from Utah State University in Logan, Utah.

In May, 1985, the author accepted a graduate teaching assistantship in the Dept. of Operations Research and Statistics at Rensselaer Polytechnic Institute (RPI) in Troy, New York. He received his M.S. in statistics from RPI in May, 1987.

In August, 1987, the author enrolled in the Statistics Department at Virginia Polytechnic Institute and State University. As a graduate teaching assistant at Virginia Tech, he taught an introductory statistics course and worked extensively with clients through the Statistical Consulting Laboratory. From March, 1988, to October, 1989, he was president of the statistics honor society Mu Sigma Rho. In the spring of 1990, he was elected a member of the national research society Sigma Chi. He is also a member of the professional societies: American Statistical Association, and Operations Research Society of America.

On May 19, 1990, the author married Vanita D'Silva of Bombay, India.

Paper published: Reeve, R., Smith, E., Wallace, B. (1990) Components of fitness become effectively neutral in equilibrium populations, *Proceedings of the National Academy of Science USA*, Vol. 87, pp. 2018-2020.