STATISTICALLY SELF-AFFINE SETS: HAUSDORFF AND BOX DIMENSIONS

by

Dimitrios Gatzouras*
and
Steven P. Lalley**
Purdue University

Technical Report # 92-35

Department of Statistics
Purdue University

August, 1992
Revised March, 1993

*Supported by Davis Ross grant
**Supported by NSF grant DMS–9146252

△

STATISTICALLY SELF-AFFINE SETS: HAUSDORFF AND BOX DIMENSIONS

by

Dimitrios Gatzouras

and

Steven P. Lalley

Purdue University

ABSTRACT

We introduce a class of random sets in $\mathbf{R}^2$ that include McMullen's "generalized Sierpinski carpets". We give exact expressions for the Hausdorff and Bouligand-Minkowski (box) dimensions of these sets, and find in particular that typically they are not equal. Our expression for the Hausdorff dimension is *not* what one would expect by analogy with McMullen's formula for the Hausdorff dimension of a generalized Sierpinski carpet.

1. Introduction

Let $A_1, A_2, \ldots, A_r$ be affine contractions of $\mathbf{R}^d$. A result of Hutchinson [Hu] implies that there exists a unique, nonempty, compact set $\Lambda \subset \mathbf{R}^d$ such that $\Lambda = \bigcup_{i=1}^{r} A_i(\Lambda)$; such a set Λ is called (strictly) *self-affine*. If the maps A_i are contractive similarity transformations then Λ is called (strictly) *self-similar*. Although dimensional properties of self-similar sets are well understood (see, e.g., [Mo], [Hu], [Ma]), surprisingly little is known about self-affine sets in general. The best general result to date seems to be that of $[\text{Fa}]_1$, which gives a formula for the Hausdorff and Bouligand-Minkowski (box) dimension in an "almost everywhere" sense, but does not indicate for which self-affine sets the formula is valid. Also noteworthy is the main result of [Mc], which gives the Hausdorff dimension explicitly for a rather restricted (and countable) collection of self-affine sets called *generalized Sierpinski carpets*. (See also $[\text{Be}]_{1-2}$, [GL].) This paper indicates that, at least for generalized Sierpinski carpets, the Hausdorff and Bouligand-Minkowski dimensions are equal only in exceptional cases, in contrast to the situation for self-similar sets.

In this paper we introduce and study a class of sets which we will call *statistically self-affine*. These sets are obtained by injecting an element of randomness into the construction of McMullen's generalized Sierpinski carpets. We will give exact expressions for the Hausdorff and Bouligand-Minkowski dimensions of our sets in terms of the statistical parameters of the construction, finding again that the two dimensions are equal only rarely. More noteworthy, our expressions are *not* what one would expect by analogy with [Mc] (in particular, they are not always what one obtains by substituting expectations into the formulas of [Mc]). Thus, the randomness in the construction has a subtle effect on dimensional properties.

Other random constructions have been studied in $[\text{Ma}]_1$, [MW], $[\text{Fa}]_2$, $[\text{Fa}]_4$, [DG], and elsewhere. These constructions lead to what we would call "statistically self-similar sets", as similarity transformations are used in place of affine transformations. Their dimensional properties are quite different from those of our sets – in particular, the values of box and Hausdorff dimensions are generally the same. The machinery of branching processes is an important tool in [MW], $[\text{Fa}]_2$, $[\text{Fa}]_4$, and [DG], as it is here, but the details of its use are somewhat different here.

The paper is organized as follows. Details of the construction are given in section 2, and the main results are stated; relevant features of the theory of branching processes are reviewed in section 3; the box dimension is computed in section 4; and finally, in section 5, the Hausdorff dimension is determined.

2. Statistically Self-Affine Sets

Throughout the paper $(\Omega, \mathcal{F}, P)$ will be the underlying probability space supporting the randomizations used in the construction of our random set K. We assume that this probability space is large enough to accommodate additional random variables independent of those used in the construction of K. To avoid notational clutter we will omit explicit reference to the functional dependence of random sets, random variables, etc., on $\omega \in \Omega$:

for instance we will write K instead of $K(\omega)$.

Let K_0 be the unit square $[0,1]^2$, and let m, n be integers satisfying $1 < m < n$. Divide K_0 into the mn congruent rectangles

$$R_{ij} = [in^{-1}, (i+1)n^{-1}] \times [jm^{-1}, (j+1)m^{-1}],$$

and let $A_\nu, \nu \in \{1, 2, 3, \ldots, mn\}$ be the natural affine transformations of R^2 mapping $[0,1]^2$ onto the rectangles R_{ij}. Let G be a probability distribution on the set of subsets of $\boldsymbol{S} = \{1, 2, 3, \ldots, mn\}$. Build random compact sets

$$K_0 \supset K_1 \supset K_2 \supset \ldots$$

as follows. Choose a random subset S of $\boldsymbol{S}$ according to the distribution G and let K_1 be the union of the $n^{-1} \times m^{-1}$ rectangles $A_\nu K_0$, where $\nu \in S$. Then, for each rectangle A_ν, K_0 chosen in the first stage of the construction, choose another random subset S' of $\boldsymbol{S}$ according to the distribution G (independent of S and of all other random subsets chosen at this stage) and replace $A_\nu K_0$ by the union of the $n^{-2} \times m^{-2}$ rectangles $A_\nu A_{\nu'} K_0$, where ν' ranges over S'. Let K_2 be the aggregate of all $n^{-2} \times m^{-2}$ rectangles so obtained. Continue in this fashion to define $K_0 \supset K_1 \supset K_2 \supset \ldots$: at the $(k+1)$th stage of the construction, replace all of the $n^{-k} \times m^{-k}$ rectangles from the kth stage by the union of randomly chosen $n^{-k-1} \times m^{-k-1}$ subrectangles, with the random choice made according to the distribution G. Define

$$K = \cap_{n=0}^{\infty} K_n.$$

Observe that the construction may terminate after a finite number of steps, if at some stage of the construction all the random subsets of $\boldsymbol{S}$ are empty. In this case, the set K is empty.

It is easily seen that the set K so constructed has the following structure:

$$K = \cup_{i \in S} A_i K(i)$$

where $K(i), i \in S$ are independent random sets, each with the same distribution as K. This could in fact be taken as the defining property of K. Observe that if S is nonrandom, i.e., if the distribution G is concentrated at a single subset of $\boldsymbol{S}$, then K is nonrandom, $K = \cup_{i \in S} A_i K$, and the construction is the same as that of McMullen [Mc].

For each $k \geq 1$, the random set K_k is the union of a random number of $n^{-k} \times m^{-k}$ rectangles contained in the unit square: call this number M_k. Thus,

$$M_k = \# \text{ of } k\text{th generation rectangles.}$$

It is apparent from the construction that the sequence M_k is an (ordinary) Galton-Watson process, and that $K = \phi$ iff $M_k = 0$ for some $k > 0$. We make the following standing

Assumption 1: $EM_1 > 1$.

The rationale for this assumption is transparent: if $EM_1 \leq 1$ then by a fundamental theorem in the theory of branching processes, $M_k = 0$ eventually, with probability one, and therefore $K = \phi$ a.s. – not an interesting case.

Figure 1 here

We will identify points of the y-axis between 0 and 1 with their m-ary expansions. Set

$$\begin{aligned}\mathcal{I} &= \{0, 1, \ldots, m-1\};\\ \mathcal{I}^k &= \mathcal{I} \times \mathcal{I} \times \ldots \times \mathcal{I} \quad (k \text{ times});\\ \mathcal{I}^\infty &= \{\text{sequences with entries in } \mathcal{I}\}.\end{aligned}$$

The mapping $T: \mathcal{I}^\infty \to [0,1]$ given by $T(s) = \Sigma s_i m^{-i}$ sets up a correspondence that is onto and fails to be 1-to-1 only at the points $im^{-j}, j \geq 1$ and $i \leq m^j$. For $y \in [0,1]$ not of the form im^{-j}, we will write $y_1 y_2 y_3 \ldots$ for the unique sequence $T^{-1}(y)$, and we will write y in place of $T^{-1}(y)$ (thus the letter y may represent either a point of [0, 1] or the sequence mapped to it by T). Although this leads to ambiguity for those y of the form $y = im^{-j}$, the ambiguity is of no consequence in the arguments below: when we cover K by balls, we may cover points (x, im^{-j}) twice rather than just once, but this redundancy clearly will have no effect on computations of box or Hausdorff dimensions.

Each finite sequence $s = (s_1 s_2 \ldots s_k) \in \mathcal{I}^k$ corresponds to an interval of length m^{-k} contained in [0, 1], namely

$$\begin{aligned}I_s &= [im^{-k}, (i+1)m^{-k}],\\ i &= \sum_{\nu=1}^{k} s_\nu m^{k-\nu}.\end{aligned}$$

Note that I_s consists of those $y \in [0,1]$ whose m-ary expansion begins with $y_\nu = s_\nu, \nu = 1, 2, \ldots, k$. The significance of the intervals I_s is that the k^{th} generation rectangles in the construction of K are arranged in "rows" $[0,1] \times I_s, s \in \mathcal{I}^k$. For each sequence $s =$

$(s_1, s_2, \ldots)$, finite or infinite, of length $\geq k$, define random variables

$$N_k(s) = \#k^{th} \text{ generation rectangles in } K_k \cap ([0,1] \times I_{s_1 s_2 \ldots s_k}),$$
$$N(s) = N_k(s) \text{ if } k = \text{ length } (s).$$

Observe that for each $k \geq 1, M_k = \Sigma_{s \in \mathcal{I}^k} N(s)$. Even though $EM_1 > 1$ it is not necessarily the case that all, or even any, of the expectations $EN(i), i \in \mathcal{I}$, are ≥ 1; but of course at least one must be positive. Define

$$\mathcal{J} = \{i \in \mathcal{I}: EN(i) > 0\}.$$

Assumption 2: $\#(\mathcal{J}) \geq 2$.

Assumption 3: $\exists i \in \mathcal{J}$ *such that* $P\{N(i) \neq 1\} > 0$.

We use $\#F$ to denote the cardinality of a set F. If Assumption 2 were false, say if $\mathcal{J} = \{i\}$, then K would always be a random Cantor set contained in the segment $[0,1] \times \{i/(m-1)\}$, and would be a special case of the [MW] construction. If Assumption 3 were false, then the Hausdorff and Bouligand-Minkowski dimensions of K could be obtained from those of $\text{proj}_2(K) = \{y: (x,y) \in K\}$, and in this case once again $\text{proj}_2(K)$ would be a special case of the [MW] construction.

Our expressions for the Hausdorff and Bouligand-Minkowski dimensions $\delta_H(K)$ and $\delta_B(K)$ involve a "thermodynamic" function $\psi(\theta)$ which is defined as follows:

$$\psi(\theta) = \log \left\{ \sum_{i \in \mathcal{J}} (EN(i))^\theta \right\}.$$

Observe that unless $EN(i) = EN(j) \;\; \forall \, i,j \in \mathcal{J}$ the function $\psi(\theta)$ has strictly positive second derivative and therefore is strictly convex. If $EN(i) = EN(j) \;\; \forall \, i,j \in \mathcal{J}$ then $\psi(\theta)$ is linear in θ; if $EN(i) = 1 \;\; \forall \, i \in \mathcal{J}$ then $\psi(\theta) \equiv \log\{EM_1\}$. In all but the last case, $\psi(\theta)$ attains its minimum value for $\theta \in [0,1]$ uniquely at some $t \in [0,1]$; if ψ is constant, define $t = 1$. The main results of the paper (Theorems 4.1–5.1) are that P-almost surely on the event $\{K \neq \phi\}$,

$$\delta_B(K) = \frac{\log EM_1}{\log n} + \left(1 - \frac{\log m}{\log n}\right) \psi(t)/\log m \tag{2.1}$$

and

$$\delta_H(K) = \psi(\alpha)/\log m \tag{2.2}$$

where

$$\alpha = \max(t, \log m/\log n).$$

Several remarks are in order. First, if $\alpha = \log_n m$ then (2.2) agrees with the formula of [Mc] in the sense that is obtained by substituting expectations in the [Mc] formula. But examples where $\alpha > \log_n m$ are abundant and easy to construct. Notice that if $\psi'(0) \geq 0$ then $t = 0$ so $\alpha = \log_n m$; this is the case, for example, if $EN(i) \geq 1 \ \forall\, i \in \mathcal{J}$. Thus, the more interesting cases are where $EN(i) < 1$ for some $i \in \mathcal{J}$.

Second, if $t = 1$ then $\alpha = 1$ and so $\psi(\alpha) = \psi(t) = \log(EM_1)$. Comparison of (2.1) and (2.2) shows that in this case $\delta_B(K) = \delta_H(K)$. A sufficient condition for $\alpha = t = 1$ is that $EN(i) < 1 \ \forall\, i \in \mathcal{J}$. A rough explanation for the fact that $\delta_B(K) = \delta_H(K)$ in this case is that the horizontal fibers of K are so sparse that K and $\mathrm{proj}_2 K$ have the same dimensional properties.

Third, if $\psi'(1) > 0, \alpha = t$, and $\psi'' > 0$ then $\psi(\alpha) < \psi(1) = \log EM_1$, hence by (2.1)–(2.2)

$$\delta_B(K) \neq \delta_H(K).$$

Since ψ, EM_1, and t all vary continuously with the parameters of the construction (specifically, the distribution G) it follows that $\delta_B \neq \delta_K$ on an open subset of the parameter space (compare with $[\mathrm{Fa}]_1$).

In section 6 we will show that $\delta_H = \delta_B$ iff $t = 1$ or $EN(i) = EN(j)$ for all $i, j \in \mathcal{J}$.

Finally the conditions as to when $\delta_H = \delta_B$ verify a conjecture of Mandelbrot in $[\mathrm{Ma}]_2$. He conjectures that $\delta_H = \delta_B$ a.s. for two very special types of distributions G: 1) fix an $M \geq 1$ and assign equal probabilities to all the ways of choosing M out of mn rectangles, and 2) fix a $0 < p < 1$ and keep any of the mn rectangles of size $n^{-1} \times m^{-1}$ with probability p, independently from on another. It is easily seen that for both these cases $t < 1$ and $EN(i) = EN(j)$ for all $i, j \in \mathcal{J}$, hence $\delta_H = \delta_B$ a.s.

Notation: We will not refer specifically to the affine transformations A_ν again. However, the notations $\mathcal{I}, \mathcal{J}, M_k, N(s), N_k(s), K, K_k, I_s, \psi, t$, and α will have the same meaning throughout the paper as in this section.

3. Branching Processes in Varying and Random Environments

We have already observed that the sequence $\{M_k\}_{k\geq 0}$ is a supercritical Galton-Watson process. Just as important for our purposes are the temporally inhomogeneous Galton-Watson processes $\{N_k(s)\}_{k\geq 0}$, where $s \in \mathcal{I}^\infty$, which we will call *branching processes in varying environments.* Define

$$G_s = \text{ distribution of } N(s), s \in \bigcup_{k\geq 1} \mathcal{I}_k.$$

For each $s \in \mathcal{I}^\infty$ the sequence $\{N_k(s)\}_{k\geq 0}$ has the following structure: $N_0(s) \equiv 1$, and $N_{k+1}(s)$ is obtained by adding $N_k(s)$ independent random variables each with distribution $G_{s_{k+1}}$. Consequently, for each $s \in \mathcal{J}^\infty$ the sequence $\{N_k(s)/\prod_{\nu=1}^{k} EN(s_\nu)\}_{k\geq 1}$ is a nonnegative martingale relative to the natural filtration, which, by the martingale convergence theorem, implies that

$$N_\infty(s) = \lim_{k\to\infty} \frac{N_k(s)}{\prod_{\nu=1}^{k} EN(s_\nu)} \tag{3.1}$$

exists a.s. It will be important for us to know when $N_\infty(s) > 0$. Certainly if the process $\{N_k(s)\}_{k\geq 1}$ reaches extinction, i.e., $N_k(s) = 0$ eventually, then $N_\infty(s) = 0$; we will see that, at least for certain $s \in \mathcal{J}^\infty, N_\infty(s) > 0$ a.s. on $\{N_k(s) \geq 1 \ \ \forall\, k \geq 1\}$.

When s is itself randomly selected, independently of the random objects used in the construction of K then the sequence $\{N_k(s)\}_{k\geq 0}$ becomes (in the terminology of $[\text{AK}]_{1-2}$) *a branching process in random environments.* Let $\zeta_1, \zeta_2, \ldots$ be a sequence of iid random variables, valued in $\mathcal{I}$, such that (under P) the sequence $\{\zeta_k\}_{k\geq 1}$ and the random sets used in the construction of K are jointly independent. Let $\zeta = \zeta_1\zeta_2\cdots$.

Theorem 3.1 $[\text{AK}]_{1,2}$

$$P\left\{\lim_{k\to\infty} N_k(\zeta) = 0\right\} + P\left\{\lim_{k\to\infty} N_k(\zeta) = \infty\right\} = 1. \tag{3.2}$$

$$P\left\{\lim_{k\to\infty} N_k(\zeta) = \infty\right\} > 0 \ \ iff \ \ E\log E[N(\zeta_1)|\zeta_1] > 0. \tag{3.3}$$

$$P\left\{N_\infty(\zeta) > 0 | \lim_{k\to\infty} N_k(\zeta) = \infty\right\} = 1 \quad \text{if} \quad E\log E[N(\zeta_1)|\zeta_1] > 0. \tag{3.4}$$

Here $E\log[N(\zeta_1)|\zeta_1] = \Sigma_{i\in I}(\log EN(i))P(\zeta_1 = i)$. Thus, it is reasonable to refer to the BPRE as subcritical, critical, or supercritical according as $E\log E[N(\zeta_1)|\zeta_1]$ is negative, zero, or positive.

Later it will be necessary to consider a somewhat more general kind of random environment $\zeta = \zeta_1\zeta_2\ldots$, where successive "blocks" $\zeta_1\zeta_2\ldots\zeta_r, \zeta_{r+1}\zeta_{r+2}\cdots\zeta_{2r}, \ldots$ of length r are iid. The results of Theorem 3.1 are still applicable, because the sequence $\{N_{kr}(\zeta)\}_{k\geq 1}$ is a BPRE to which the results of $[\text{AK}]_{1-2}$ may be applied. The second and third statements of the theorem must be modified as follows:

$$P\left\{\lim_{k\to\infty} N_k(\zeta) = \infty\right\} > 0 \ \ iff \ \ \sum_{\nu=1}^{r} E\log E[N(\zeta_\nu)|\zeta_1\zeta_2\ldots\zeta_r] > 0; \tag{3.5}$$

$$P\{N_\infty(\zeta) > 0 | \lim N_k(\zeta) = \infty\} = 1 \quad \text{if} \quad \sum_{\nu=1}^{r} E\log E[N(\zeta_\nu)|\zeta_1\zeta_2\ldots\zeta_r] > 0. \tag{3.6}$$

Theorem 3.2

Assume that $\zeta = \zeta_1\zeta_2\ldots$ where successive r-blocks are iid and jointly independent of the random sets in the construction of K. If

$$(3.7) \qquad P\left\{\sum_{\nu=1}^{r} \log E[N(\zeta_\nu)|\zeta_1,\ldots,\zeta_r] > 0\right\} = 1$$

then

$$(3.8) \qquad E(N_\infty(\zeta))^2 < \infty.$$

Proof: It suffices to consider the case $r = 1$. Recall that $N_\infty(\zeta)$ is the limit of the martingale (3.1); thus, it suffices to prove that this martingale is L^2-bounded. Let $g(j) = EN(j)$ and $h(j) = \text{var}\,(N(j))$ for $j \in \mathcal{J}$. By (3.7), $P\{g(\zeta_1) > 1\} = 1$; since $g(\zeta_1)$ can assume only finitely many values, there exists $\varepsilon > 0$ such that $g(\zeta_1) \geq 1+\varepsilon$ and $h(\zeta_1) \leq \varepsilon^{-1}$ a.s. Now conditioning on ζ and on $N_k(\zeta)$ one obtains

$$E[N_{k+1}(\zeta)^2|\zeta] = E[N_k(\zeta)^2|\zeta]g_{k+1}(\zeta)^2 + E[N_k(\zeta)|\zeta]h(\zeta_{k+1}), \text{ a.s.}$$

$$\Longrightarrow E\left\{\frac{N_{k+1}(\zeta)}{\prod_{\nu=1}^{k+1} g(\zeta_\nu)}\right\}^2 \leq E\left\{\frac{N_k(\zeta)}{\prod_{\nu=1}^{k} g(\zeta_\nu)}\right\}^2 + E\left\{\frac{N_k(\zeta)}{\prod_{\nu=1}^{k} g(\zeta_\nu)}\right\}\varepsilon^{-1}(1+\varepsilon)^{-k-2}.$$

Since $E\{N_k(\zeta)/\prod_{\nu=1}^{k} g(\zeta_\nu)\} = 1$ it now follows that

$$EN_\infty(\zeta)^2 \leq 1 + \varepsilon^{-2}(1+\varepsilon) < \infty. \qquad \blacksquare$$

Define

$$(3.9) \qquad Z_k = \sum_{s\in\mathcal{I}^k} 1_{\{N(s)\geq 1\}};$$

Z_k is the number of k^{th} generations "rows" still alive, i.e., containing k^{th} generation rectangles of K_k. The following theorem is a consequence of the main result of [De].

Theorem 3.3

$$(3.10) \qquad \lim_{k\to\infty} k^{-1} \log EZ_k = \psi(t).$$

Recall (section 2) that $\psi(\theta)$ attains its minimum value for $\theta \in [0,1]$ at $\theta = t$. Theorem 3.3 will be of crucial importance in the determination of the box dimension $\delta_B(K)$ in section 4.

Proposition 3.4

$$\psi(t) > 0 \tag{3.11}$$

Proof: By Assumption 2 of section 2 there are at least two distinct $i \in \mathcal{J}$, for which $EN(i) > 0$. Suppose there exists $i \in \mathcal{J}$ such that $EN(i) \geq 1$: then $(EN(i))^t \geq 1$; and since there is at least one other $j \in \mathcal{J}$ for which $EN(j) > 0$, it follows that $\Sigma(EN(i))^t > 1$, which implies (3.11). Suppose, then, that $\forall\ i \in \mathcal{J}, EN(i) < 1$. In this case $t = 1$, so $\Sigma(EN(i))^t = \Sigma EN(i) = EM_1 > 1$, by Assumption 1, proving (3.11). ■

Proposition 3.5

On the event $\{K \neq \phi\}$,

$$\lim_{k\to\infty} k^{-1} \log Z_k = \psi(t) \quad a.s. \tag{3.12}$$

Proof: By Theorem 3.3 and Proposition 3.4, for all $\varepsilon > 0$ sufficiently small there exists $k \geq 1$ such that

$$EZ_k \geq \exp\{k\psi(t) - k\varepsilon\} > 1.$$

Construct a Galton-Watson process $Y_i, i \geq 0$, as follows: set $Y_1 = Z_k$; for each $s \in \mathcal{I}^k$ such that $N(s) \geq 1$ *throw away* all but one of the rectangles counted in $N(s)$ before continuing the construction; set $Y_2 =$ the number of $s \in \mathcal{I}^{2k}$ counted in Z_{2k} that are not contained in one of the "thrown away" rectangles from the previous stage; for each $s \in \mathcal{I}^{2k}$ counted in Y_2 throw away all but one of the rectangles counted in $N(s)$ before continuing the construction; continue indefinitely. That $Y_1, Y_2, \ldots$ is in fact a Galton-Watson process is easily verified, and clearly the offspring distribution has mean EZ_k. Consequently, on the event of non-extinction, as $i \to \infty$,

$$Y_i/(EZ_k)^i \longrightarrow W > 0 \quad a.s.,$$

by a standard theorem in branching process theory (see [AN]). Now $Y_i \leq Z_{ik}$, so on the event that $\{Y_i\}_{i\geq 1}$ does not reach extinction

$$\liminf_{i\to\infty} \frac{1}{i} \log Z_{ik} \geq \log EZ_k \geq k(\psi(t) - \varepsilon) \tag{3.13}$$

almost surely.

It is possible that $\{Y_i\}_{i\geq 1}$ may reach extinction on the event $\{K \neq \phi\}$, because we have thrown away "growth opportunities" for K in the above construction. If this happens, go back to one of the "thrown away" rectangles and begin afresh, constructing a new, independent copy of $\{Y_i\}_{i\geq 1}$ in the same manner as before. Once again $Y_{i-i_*} \leq Z_{ik}$ for all i (where i_* is the generation number of the thrown away rectangle). Consequently, on the event that this new copy does not reach extinction, (3.13) holds a.s. But on $\{K \neq \phi\}$ we can keep going back to "thrown away" rectangles until eventually finding one that engenders a copy of $\{Y_i\}_{i\geq 1}$ that does not die out. (Keep in mind that, since $EZ_k > 1, \{Y_i\}_{i\geq 1}$ is supercritical and therefore has probability < 1 of extinction.) It follows that (3.13) holds a.s. on the event $\{K \neq \phi\}$. Since $\varepsilon > 0$ was arbitrary, this proves that on $\{K \neq \phi\}$,

$$\liminf_{k\to\infty} \frac{1}{k} \log Z_k \geq \psi(t) \quad a.s.$$

The opposite inequality

$$\limsup_{k\to\infty} \frac{1}{k} \log Z_k \leq \psi(t) \quad a.s.$$

may be proved by a much more straightforward argument using Theorem 3.3, the Markov inequality, and the Borel-Cantelli lemma. ∎

4. The Bouligand-Minkowski (Box) Dimension

If X is a compact metric space then its Bouligand-Minkowski dimension δ_B (also referred to as "box dimension" and "capacity") is defined by

$$\delta_B = \limsup_{\varepsilon\to 0} \frac{\log C(\varepsilon)}{-\log \varepsilon}$$

where $C(\varepsilon)$ is the cardinality of a minimal covering of X by ε-balls. The object of this section is to prove

Theorem 4.1

Conditional on $K \neq \phi$, the Bouligand-Minkowski dimension $\delta_B(K)$ of K is, with probability one, given by

$$\delta_B(K) = \frac{\log(EM_1)}{\log n} + \left(1 - \frac{\log m}{\log n}\right) \frac{\psi(t)}{\log m}. \tag{4.1}$$

We begin the proof by showing that it suffices to consider coverings of K by certain rectangles which we call (following [Mc]) "approximate squares". For $l = 1, 2, \ldots$ set $k = k_l = [l \log n / \log m]$ (here [] denotes integer part); note that $k_1 < k_2 < \ldots$ is an increasing sequence of positive integers, and that for each $l = 1, 2, \ldots$

$$m^k \leq n^l < m^{k+1}.$$

Now for each $l \geq 1$ define the l^{th} generation approximate squares $R_l(p, q)$, where $p \in \{0, 1, \ldots, n^l - 1\}$ and $q \in \{0, 1, \ldots, m^k - 1\}$, by

$$R_l(p, q) = [pn^{-l}, (p+1)n^{-l}] \times [qm^{-k}, (q+1)m^{-k}].$$

Note that for distinct pairs $(p,q),(p',q')$ the rectangles $R_l(p,q)$ and $R_l(p',q')$ overlap either in a line segment, a point, or not at all. Moreover, the ratio of height to width of $R_l(p,q)$ is always between 1 and m; hence the term "approximate square".

Lemma 4.2

Let C_l be the cardinality of a minimal covering of K by l^{th} generation approximate squares. Then

$$\delta_B(K) = \limsup_{l\to\infty} \frac{\log C_l}{l \log n}.$$

Proof: It suffices to show that no covering of K by ε-balls is much more efficient than the best covering by approximate squares of roughly the same size. Let $\varepsilon > 0$ be such that $n^{-l-1} \leq \varepsilon < n^{-l}$ for some $l \geq 1$. Then any ε-ball's intersection with K is contained in the union of (at most) nine l^{th} generation approximate squares. Consequently, for any covering of K by ε-balls there is a covering by l^{th} generation approximate squares with nine times as many members. The lemma follows easily from this. ■

Define

$$d = \frac{\log(EM_1)}{\log n} + \left(1 - \frac{\log m}{\log n}\right) \frac{\psi(t)}{\log m}. \tag{4.2}$$

To show that $\delta_B(K) \leq d$ it suffices to exhibit an efficient covering of K by l^{th} generation approximate squares for each $l = 1, 2, \ldots$. Consider the collection $\mathcal{V}_l$ consisting of those l^{th} generation approximate squares $R_l(p,q)$ satisfying

$$\text{Interior } (R_l(p,q)) \cap K_k \neq \phi.$$

Clearly, $\mathcal{V}_l$ is a covering of K_k, hence also of K.

Lemma 4.3

$$\limsup_{l\to\infty} \left\{ \frac{\log(\#\mathcal{V}_l)}{l} \right\} \leq d \log n \quad a.s. \tag{4.3}$$

Proof: In order that an approximate square $R_l(p,q)$ be included in the collection $\mathcal{V}_l$ it is necessary and sufficient that it be entirely contained in one of the M_l rectangles that make up K_l. Thus,

$$\#\mathcal{V}_l = \sum_{j=1}^{M_l} Y_j^{(k-l)} \tag{4.4}$$

where $Y_j^{(k-l)}$ is the number of distinct l^{th} generation approximate squares contained in the j^{th} of the M_l rectangles that make up K_l. Observe that, conditional on $M_l = r \geq 1$, the random variables $Y_1^{(k-l)}, Y_2^{(k-l)}, \ldots, Y_r^{(k-l)}$ are iid. Moreover, $Y_1^{(k-l)}$ has the same distribution as Z_{k-l}, defined by (3.9), so

$$EY_1^{(k-l)} = EZ_{k-l}$$

is as described in Theorem 3.3. To see this, consider the process by which the random set K_k is constructed: First, K_l is constructed, consisting of a random number M_l of nonoverlapping n^{-l} by m^{-l} rectangles. Then in each of these M_l rectangles the process is continued another $k-l$ steps. For one of the M_l rectangles in K_l, say the j^{th}, those "rows" of width n^{-l} and height m^{-k} that "survive" another $k-l$ generations, i.e., intersect K_k, are *precisely* the approximate squares counted in $Y_j^{(k-l)}$.

It follows from the foregoing representation that

$$E(\#\mathcal{V}_l) = (EM_l)\mu_{k-l} = (EM_1)^l \mu_{k-l},$$

where $\mu_{k-l} = EZ_{k-l}$. Consequently, by the Markov inequality and the Borel-Cantelli lemma, for any $\varepsilon > 0$

$$P\{\#(\mathcal{V}_l) \geq (EM_1)^l \mu_{k-l} e^{\varepsilon l}\} \leq e^{-\varepsilon l}$$
$$\Longrightarrow P\{\#(\mathcal{V}_l) \geq (EM_1)^l \mu_{k-l} e^{\varepsilon l} \;\; i.o.\} = 0.$$

Since $k = k_l = [l \log n / \log m]$, Theorem 3.3 now shows that with probability one the inequality (4.3) in the statement of the lemma must hold. ■

Each of the approximate squares in the collection $\mathcal{V}_l$ has width n^{-l} and height $m^{-k} \in [n^{-l}, mn^{-l}]$. Therefore, Lemma 4.3 implies the upper bound

$$\delta_B(K) \leq d \;\; a.s. \tag{4.5}$$

It remains to establish the reverse inequality. This we will accomplish by showing that the coverings cannot be improved in an essential way. Before doing this, however, we will obtain an asymptotic *lower* bound for $\#\mathcal{V}_l$ to complement Lemma 4.3.

Lemma 4.4

$$\liminf_{l \to \infty} \left\{ \frac{\log(\#\mathcal{V}_l)}{l} \right\} \geq d \log n \;\; a.s. \; on \;\; \{K \neq \phi\}. \tag{4.6}$$

Proof: By (4.4), $\#\mathcal{V}_l$ is the sum of M_l iid (conditional on the value of M_l) random variables $Y_j^{(k-l)}$ each having the same distribution as the random variable Z_{k-l} defined by (3.9). Recall that $\{M_l\}_{l \geq 0}$ is a supercritical Galton-Watson process and that $\{K \neq \phi\}$ is precisely the event that $\{M_l\}_{l \geq 0}$ does not reach extinction. Consequently, there is a random variable W such that $W > 0$ on $\{K \neq \phi\}$ and as $l \to \infty$,

$$M_l/(EM_1)^l \xrightarrow{a.s.} W.$$

Recall from Proposition 3.5 that on $\{K \neq \phi\}$, as $k \to \infty$,

$$\frac{1}{k} \log Z_k \xrightarrow{a.s.} \psi(t)$$

Hence, since $k - l \to \infty$ as $l \to \infty$, for each $\varepsilon > 0$

$$P\{Z_{k-l} \geq \exp\{(k-l)(\psi(t) - \varepsilon)\}\} \geq \frac{1}{2}$$

for all sufficiently large l. Since conditional on M_l the random variables $Y_j^{(k-l)}$ are iid copies of Z_{k-l} it follows that for all sufficiently large l,

$$\#\mathcal{V}_l = \sum_{j=1}^{M_l} Y_j^{(k-l)} \geq \left(\sum_{j=1}^{M_l} \xi_j^{(k-l)} \right) \exp\{(k-l)(\psi(t) - \varepsilon)\}$$

where conditional on $M_l, \xi_1^{(k-l)}, \ldots, \xi_{M_l}^{(k-l)}$ are independent 0-1 random variables each with $P(\xi_i^{(k-l)} = 1, M_l \geq i) \geq \frac{1}{2} P(M_l \geq i)$ (just take $\xi_j^{(k-l)} = 1$ if $Y_j^{(k-l)} \geq \exp\{(k-l)(\psi(t) - \varepsilon)\}$ and $\xi_j^{(k-l)} = 0$ otherwise). Now for any $0 < \delta < 1$ we have by Markov's inequality

$$\begin{aligned}
&P\left(\sum_{j=1}^{M_l} \xi_j^{(k-l)} < \frac{1}{4} M_l \text{ and } M_l \geq (EM_1)^{(1-\delta)l} \right) \\
&\leq P\left(\sum_{j=1}^{M_l} P(\xi_j^{(k-l)} = 1 | M_\ell) - \sum_{j=1}^{M_l} \xi_j^{(k-l)} > \frac{1}{4} M_l \text{ and } M_l \geq (EM_1)^{(1-\delta)l} \right) \\
&\leq 4(EM_1)^{-(1-\delta)l}
\end{aligned}$$

and since $M_l/(EM_1)^l \to W > 0$ a.s. on $\{K \neq \phi\}$, the Borel-Cantelli lemma implies that

$$\liminf_{l \to \infty} \frac{1}{l} \left[\log \left\{ \sum_{j=1}^{M_l} \xi_j^{(k-l)} \right\} - \log M_l \right] \geq 0$$

a.s. on $\{K \neq \phi\}$. Using this inequality together with the one for $\#\mathcal{V}_l$ above and the fact that $M_l/(EM_1)^l \to W > 0$ again, one easily concludes that a.s. on $\{K \neq \phi\}$,

$$\liminf_{l \to \infty} \left\{ \frac{\log(\#\mathcal{V}_l)}{l} \right\} \geq \log(EM_1) + (\psi(t) - \varepsilon) \left(\frac{\log n}{\log m} - 1 \right).$$

Since $\varepsilon > 0$ was arbitrary, (4.6) now follows. ∎

To complete the proof of Theorem 4.1 we must show that on the event $\{K \neq \phi\}$,

$$\delta_B(K) \geq d \quad a.s.$$

By Lemma 4.2 it suffices to show that on $\{K \neq \phi\}$,

$$\limsup_{l \to \infty} \frac{1}{l} \log C_l \geq d \log n, \tag{4.7}$$

where C_l is the cardinality of a minimal covering of K by l^{th} generation approximate squares. Clearly, any covering of K by l^{th} generation approximate squares must include all members of the collection

$$\mathcal{W}_l = \{R_l(p,q) \colon \text{Interior } (R_l(p,q)) \cap K \neq \phi\}.$$

Note that $\mathcal{W}_l \subset \mathcal{V}_l$. We will argue that on $\{K \neq \phi\}$,

$$\liminf_{l \to \infty} \left(\frac{\#\mathcal{W}_l}{\#\mathcal{V}_l}\right) \geq p \quad a.s., \tag{4.8}$$

where $p > 0$ is the survival probability of the Galton-Watson process $\{M_j\}_{j \geq 1}$, equivalently, $p = P\{K \neq \phi\}$. In view of (4.6) this will prove (4.7) and therefore complete the proof of Theorem 4.1.

Recall that $\mathcal{V}_l$ consists of those l^{th} generation approximate squares $R_l(p,q)$ such that Int $(R_l(p,q)) \cap K_k \neq \phi$. Thus, $R_l(p,q) \in \mathcal{V}_l$ implies that $R_l(p,q)$ contains at least one of the n^{-k} by m^{-k} rectangles that make up K_k. Each such n^{-k} by m^{-k} rectangle has probability p of containing a point of K in its interior, and the events of "survival" for the various n^{-k} by m^{-k} rectangles in K_k are independent. Therefore,

$$\#\mathcal{W}_l \geq \sum_{j=1}^{\#\mathcal{V}_l} \xi_j^{(l)}$$

where $\xi_1^{(l)}, \ldots, \xi_{\#\mathcal{V}_l}^{(l)}$ are independent 0-1 random variables (conditional on $\#\mathcal{V}_l$) each with $P(\xi_i^{(l)} | \#\mathcal{V}_l) \geq p$. Since $\#\mathcal{V}_l \to \infty$ at an exponential rate, the result (4.8) follows by a routine argument. ■

5. The Hausdorff Dimension

For any subset X of $\mathbf{R}^k$ the δ-dimensional Hausdorff outer measure of X is defined as follows:

$$H_\delta(X) = \liminf_{\varepsilon \to 0} \left\{ \sum_{i=1}^{\infty} |U_i|^\delta \colon X \subset \bigcup_{i=1}^{\infty} U_i \text{ and } |U_i| \leq \varepsilon \right\}.$$

Here $|U|$ indicates the diameter of U and the infimum is over all coverings of X by sets of diameter $\leq \varepsilon$. When H_δ is restricted to the Borel subsets of $\mathbf{R}^k$ it becomes a measure – see [Fa]$_3$. The Hausdorff dimension $\delta_H(X)$ of X is defined by

$$\delta_H(X) = \inf\{\delta \geq 0 \colon H_\delta(X) < \infty\}.$$

It is easily checked that

$$\delta_H(X) \le \delta_B(X)$$

where $\delta_B(X)$ is the Bouligand-Minkowski (box) dimension of X.

The purpose of this section is to prove

Theorem 5.1

Set $\alpha = \max(t, \log_n m)$. Then on the event $\{K \neq \phi\}$ the Hausdorff dimension $\delta_H(K)$ of K is P-a.s. given by

$$\delta_H(K) = \psi(\alpha)/\log m. \tag{5.1}$$

We note at the outset that $\delta_H(K)$ is almost surely constant on the event $\{K \neq \phi\}$. The argument is as follows. Conditioning on the first k generations of the construction shows that K is the union of M_k sets $A_s K^{(s)}$, where each A_s is an affine map and $K^{(s)}$ are iid copies of K. Consequently,

$$\delta_H(K) = \max_{1 \le s \le M_k} \delta_H(K^{(s)}),$$

where, conditional on M_k, the random variables $\delta_H(K^{(s)})$ are iid, each with the same distribution as $\delta_H(K)$. Since $\{K \neq \phi\} = \{\lim_{k\to\infty} M_k = \infty\}$, it follows that

$$\delta_H(K) = \delta_* 1\{K \neq \phi\} \quad a.s.,$$

where $\delta_* = \sup\{x \in \mathbf{R}^+ : P\{\delta_H(K) > x\} > 0\}$.

In proving Theorem 5.1 we will consider the cases $\psi'(1) \le 0$ and $\psi'(1) > 0$ separately. The first case is simpler: in this case K has the same Hausdorff and box dimensions as its projection on the y-axis. So we will begin by computing the dimension(s) of this projection. For any set $F \subset \mathbf{R}^2$, define

$$\text{proj}_2 F = \{y \in \mathbf{R} : \mathbf{R} \times \{y\} \cap F \neq \phi\}.$$

Proposition 5.2

On the event $\{K \neq \phi\}$,

$$\delta_H(\text{proj}_2 K) = \delta_B(\text{proj}_2 K) = \psi(t)/\log m \quad a.s.$$

Proof: Let $\mathcal{G}_l$ be the covering of $\text{proj}_2 K$ consisting of those intervals $[jm^{-l}, (j+1)m^{-l}]$ whose interiors intersect $\text{proj}_2 K_l$. These intervals are indexed by those $s \in \mathcal{J}^l$ that are counted in Z_l (see (3.9)); hence $\#\mathcal{G}_l = Z_l$. Therefore, by Proposition 3.5, on $\{K \neq \phi\}$

$$\delta_B(\text{proj}_2 K) \le \psi(t)/\log m \quad a.s.$$

To complete the proof it suffices to show that on the event $\{K \neq \phi\}, \delta_H(\text{proj}_2 K) \geq \psi(t)/\log m$. We will accomplish this by showing that $\forall \varepsilon > 0$ there is a set $L^* \subset K$ such that $\delta_H(\text{proj}_2 L^*) \geq (\psi(t) - \varepsilon)/\log m$. The set $\text{proj}_2 L^*$ will be a special case of the [MW] construction.

By Theorem 3.3 there exists for any $\varepsilon > 0$ an integer $k = k_\varepsilon$ sufficiently large that $EZ_k \geq \exp\{k(\psi(t) - \varepsilon)\}$. Fix this k, and consider the sets $K_k \supset K_{2k} \supset \ldots$ in the construction of K; recall that each K_{jk} is the union of $(jk)^{th}$ generation rectangles which are arranged in the "rows" $[0,1] \times I_s, s \in \mathcal{J}^{jk}$. Define sets $L_j \subset K_{jk}$ as follows. Throw away all but the leftmost of the k^{th} generation rectangles in each row, and let L_1 be the union of those remaining. Note that $\text{proj}_2 L_1 = \text{proj}_2 K_k$. To obtain $L_{j+1} \subset K_{(j+1)k}$, throw away all but the leftmost of the $((j+1)k)^{th}$ generation rectangles in each row of $K_{(j+1)k} \cap L_j$, and let L_{j+1} be the union of those remaining. Clearly, $L_1 \supset L_2 \supset \ldots$, so we may define $L = \bigcap_{j \geq 1} L_j$. By construction, $L \subset K$ and hence $\text{proj}_2 L \subset \text{proj}_2 K$. But $\text{proj}_2 L$ is a random Cantor set of the type considered in [MW] and $[\text{Fa}]_2$; the main results of either of these papers implies that on the event $\{L \neq \phi\}$,

$$\delta_H(\text{proj}_2 L) = \frac{\log EZ_k}{k \log m} \geq \frac{(\psi(t) - \varepsilon)}{\log m} \quad a.s.$$

It is of course possible that $L = \phi$ even though $K \neq \phi$. This difficulty may be handled by the same device used in the proof of Proposition 3.5. If $L = \phi$, return to the first rectangle that was thrown away during the construction of L, and begin the entire procedure again in this rectangle. The result will be another random set L', independent of L, and such that for a suitable affine map A, AL' has the same law as L. Thus on $\{L' \neq \phi\}$ the dimension $\delta_H(L')$ satisfies $\delta_H(L') \geq (\psi(t) - \varepsilon)/\log m$ a.s. If $L' = \phi$, go back to yet another thrown away rectangle, and continue until eventually obtaining a nonempty copy of L. ■

The proof of Theorem 5.1 in the case $\psi'(1) \leq 0$ may now be given. If $\psi'(1) \leq 0$ then $t = 1$, hence $\alpha = t$, and so $\psi(\alpha) = \psi(t) = \psi(1) = \log(EM_1)$. But by Theorem 4.1,

$$\delta_B(K) = \psi(\alpha)/\log m \quad a.s.$$

on $\{K \neq \phi\}$. By Proposition 5.2,

$$\delta_H(K) \geq \delta_H(\text{proj}_2 K) = \psi(\alpha)/\log m \ a.s.$$

on $\{K \neq \phi\}$. Since $\delta_H \leq \delta_B$ it follows that in this case both are equal, and so a.s. on $\{K \neq \phi\}$

$$\delta_H(K) = \delta_B(K) = \psi(\alpha)/\log m.$$

■

The Case $\psi'(1) > 0$

ASSUME for the remainder of section 5 that $\psi'(1) > 0$. Thus, $0 \leq t \leq \alpha < 1$, and $\psi'(\theta) > 0$ for every $\theta > t$.

The proof of (5.1) will be given in two stages: first, it will be shown that $\psi(\alpha)/\log m$ is a lower bound for $\delta_H(K)$; then, that it is an upper bound.

A. The Lower Bound

The strategy here is based on a theorem of Marstrand (see [Fa]$_3$, Theorem 5.8 and Exercise 5.2). For $F \subset \mathbf{R}^2$ and $y \in \mathbf{R}$ define

$$F(y) = \{x \in \mathbf{R}: (x,y) \in F\},$$
$$\text{proj}_2 F = \{y \in \mathbf{R}: F(y) \neq \phi\}.$$

Marstrand's Theorem

If for each $y \in$ $proj_2 F$ the fiber $F(y)$ has Hausdorff dimension $\delta_H(F(y)) \geq \delta_1$ and if $\delta_H(proj_2 F) \geq \delta_2$ then

$$\delta_H(F) \geq \delta_1 + \delta_2.$$

To use Marstrand's Theorem we will need to estimate the Hausdorff dimensions of fibers $F(y)$ and projections $\text{proj}_2 F$ of various F. The key tool for doing so is a lemma of Frostman. Let ν be a Borel probability measure on a Euclidean space $\mathbf{R}^d$; define its Hausdorff dimension $\delta_H(\nu)$ to be the infimum of $\delta_H(A)$ for sets A satisfying $\nu(A) = 1$. Clearly, if F is any Borel set and ν is any Borel probability measure supported by F (i.e., if $\nu(F) = 1$) then $\delta_H(\nu)$ is a lower bound for $\delta_H(F)$.

Frostman's Lemma

If

$$\delta_1 \leq \liminf_{r \to 0} \frac{\log \nu(B(x,r))}{\log r} \leq \limsup_{r \to 0} \frac{\log \nu(B(x,r))}{\log r} \leq \delta_2$$

for ν-a.e. x, then

$$\delta_1 \leq \delta_H(\nu) \leq \delta_2.$$

Here $B(x,r)$ denotes the ball of radius r centered at x. For a proof of Frostman's lemma (in this form) see [Yo].

Define probability measures $\mu_\theta, \theta \in \mathbf{R}$, on the unit interval [0, 1] as follows. Let $\zeta_1, \zeta_2, \ldots$ be iid $\mathcal{J}$-valued random variables with distribution

(5.2) $$P\{\zeta_1 = j\} = e^{-\psi(\theta)} g(j)^\theta, \quad j \in \mathcal{J}$$

where

$$g(j) = EN(j).$$

Define μ_θ to be the distribution of $\Sigma_{j=1}^\infty \zeta_j m^{-j}$. Note that

(5.3) $$E \log g(\zeta_1) = \psi'(\theta),$$
$$\text{var } (\log g(\zeta_1)) = \psi''(\theta).$$

Lemma 5.3

Fix $\theta > t$, and let $G \subset [0,1]$ be any Borel set such that $\mu_\theta(G) = 1$. Then P-a.s. on the event $\{K \neq \phi\}$,

$$\delta_H(G \cap \operatorname{proj}_2 K) \geq \frac{\psi(\theta) - \theta\psi'(\theta)}{\log m}. \tag{5.4}$$

Lemma 5.4

Assume that for every $j \in \mathcal{J}, EN(j) > 1$. Then for each $\theta > t$, on the event $\{K \neq \phi\}$

$$\mu_\theta\{y \in [0,1]\colon \delta_H(K(y)) \geq \psi'(\theta)/\log n\} = \mu_\theta(\operatorname{proj}_2 K) \tag{5.5}$$

almost surely (P).

We have not been able to determine whether Lemma 5.4 remains true without the hypothesis that $EN(j) > 1 \ \forall\, j \in \mathcal{J}$. Consequently, we will have to take a slightly more roundabout route to proving the lower bound in the general case.

Before proving Lemmas 5.3–5.4, we will show how they imply the lower bound

$$\delta_H(K) \geq \psi(\alpha)/\log m \tag{5.6}$$

a.s. (P) on $\{K \neq \phi\}$ in the special case where $EN(j) > 1$ for all $j \in \mathcal{J}$. Fix $\theta > t$. By Lemma 5.4 there exists, a.s. (P) on $\{K \neq \phi\}$, a Borel set $G \subset [0,1]$ such that $\mu_\theta(G) = 1$ and such that for every $y \in G \cap \operatorname{proj}_2 K$,

$$\delta_H(K(y)) \geq \psi'(\theta)/\log n.$$

By Lemma 5.3, $\delta_H(G \cap \operatorname{proj}_2 K) \geq (\psi(\theta) - \theta\psi'(\theta))/\log m$. Consequently, Marstrand's theorem applied to $G \cap \operatorname{proj}_2 K$ implies that P-a.s. on $\{K \neq \phi\}$,

$$\delta_H(K) \geq \frac{\psi'(\theta)}{\log n} + \frac{\psi(\theta) - \theta\psi'(\theta)}{\log m}. \tag{5.7}$$

It is a simple exercise in calculus to verify that the supremum over $\theta > t$ of the expression on the right is attained at $\theta = \alpha$, proving (5.6). ■

Proof *of Lemma 5.3*: First we show that $\forall\, \theta > t$,

$$P(\mu_\theta(\operatorname{proj}_2 K) > 0 | K \neq \phi) = 1. \tag{5.8}$$

Conditioning on the first step of the construction one sees that $P\{\mu_\theta(\operatorname{proj}_2 K) = 0\}$ is a solution of the equation

$$z = \sum_{j=1}^{\infty} z^j P\{M_1 = j\}.$$

There are two solutions to this equation, $z = 1$ and $z = P\{K = \phi\}$. To see that $P\{\mu_\theta(\text{proj}_2 K) = 0\}$ must be the smaller of these, observe that

$$\begin{aligned} E(\mu_\theta(\text{proj}_2 K)) &= \int_{[0,1]} P\{y \in \text{proj}_2 K\} d\mu_\theta(y) \\ &= P\{N_k(\zeta) \geq 1 \ \ \forall\, k \geq 1\} \\ &> 0 \end{aligned}$$

by Theorem 3.1 (specifically, (3.3)), since $E \log E[N(\zeta_1)|\zeta_1] = \psi'(\theta) > 0$ for $\theta > t$. (Note: $\zeta = \zeta_1\zeta_2\dots$ where $\zeta_1, \zeta_2, \dots$ are iid with distribution (5.2).) Thus,

$$P\{\mu_\theta(\text{proj}_2 K) > 0\} = P\{K \neq \phi\},$$

and (5.8) clearly follows from this.

In view of (5.8),

$$\nu_\theta(A) = \frac{\mu_\theta(A \cap \text{ proj}_2 K)}{\mu_\theta(\text{proj}_2 K)}$$

defines a probability measure ν_θ on the Borel sets of [0, 1] (P-a.s. on $\{K \neq \phi\}$). We will use Frostman's lemma to obtain a lower bound for $\delta_H(\nu_\theta)$. Let $y \in [0,1]$ and $0 < r < 1$; then

$$\begin{aligned} \frac{\log \nu_\theta(B(y,r))}{\log r} &= \frac{\log \mu_\theta(B(y,r) \cap \text{ proj}_2 K)}{\log r} - \frac{\log \mu_\theta(\text{proj}_2 K)}{\log r} \\ &\geq \frac{\log \mu_\theta(B(y,r))}{\log r} - \frac{\log \mu_\theta(\text{proj}_2 K)}{\log r}, \end{aligned}$$

which implies that

$$\liminf_{r \to 0} \frac{\log \nu_\theta(B(y,r))}{\log r} \geq \liminf_{r \to 0} \frac{\log \mu_\theta(B(y,r))}{\log r}.$$

But it is a routine consequence of the SLLN that

$$\lim_{r \to 0} \frac{\log \mu_\theta(B(y,r))}{\log r} = \frac{\psi(\theta) - \theta\psi'(\theta)}{\log m}$$

a.e. μ_θ, and hence also a.e. ν_θ. Since $\nu_\theta(G) = 1$ whenever $\mu_\theta(G) = 1$, it follows that for any G such that $\mu_\theta(G) = 1$ it must be the case that $\nu_\theta(G \cap \text{ proj}_2 K) = 1$. But Frostman's lemma and the preceding argument implies that $\delta_H(\nu_\theta) \geq (\psi(\theta) - \theta\psi'(\theta))/\log m$, so it follows that ($P$-a.s. on $\{K \neq \phi\}$), $\delta_H(G \cap \text{ proj}_2 K) \geq (\psi(\theta) - \theta\psi'(\theta))/\log m$. ■

Proof *of Lemma 5.4*: Recall that $\zeta = \zeta_1\zeta_2\dots$ where $\zeta_1, \zeta_2, \dots$ are iid $\mathcal{J}$-valued random variables each with distribution (5.2), and that μ_θ is the distribution of $\Sigma\zeta_i m^{-i}$. Consequently, to prove (5.5) it suffices to prove that

$$\delta_H(K(\zeta)) \geq \psi'(\theta)/\log n$$

a.s. on $\{K(\zeta) \neq \phi\}$. By Frostman's Lemma it suffices to exhibit a probability measure λ supported by $K(\zeta)$ such that

$$\liminf_{r \to 0} \frac{\log \lambda(B(x,r))}{\log r} \geq \psi'(\theta)/\log n \tag{5.9}$$

for λ-a.e. x. Define λ_k to be *normalized* Lebesgue measure on $K_k(\zeta), k \geq 1$; by Helly's selection principle every subsequence of $\{\lambda_k\}_{k\geq 1}$ has a subsequence that converges weak-$*$ to a limiting probability measure λ. Observe that any such λ is supported by $K(\zeta)$, because λ_k is supported by $K_k(\zeta), K_1(\zeta) \supset K_2(\zeta) \supset \ldots$ are all compact sets, and $K(\zeta) = \bigcap_{k\geq 1} K_k(\zeta)$. We will show that for some such λ, (5.9) is valid. (Note: With more care it can be shown that in fact there is only one possible limit λ.)

Fix $k \geq 1$ and let $J_k^{(i)}, i = 1, 2, \ldots, N_k(\zeta)$, be the nonoverlapping k^{th} generation intervals whose union is $K_k(\zeta)$. Each $J_k^{(i)}$ is the intersection of $[0,1] \times \{y\}$ with one of the k^{th} generation rectangles whose union is K_k; hence each $J_k^{(i)}$ has length n^{-k}. Moreover, each $J_k^{(i)}$ begets its own random construction, and, conditional on ζ and the details of the construction through generation k, the different constructions $i = 1, 2, \ldots, N_k(\zeta)$ are iid replicas (scaled by the factor n^{-k}) of the original construction for the sequence $\sigma^k\zeta$. Here $\sigma: \mathcal{I} \to \mathcal{I}$ denotes the shift, so

$$\sigma^k \zeta = \zeta_{k+1}\zeta_{k+2}\cdots .$$

Consequently,

$$\lambda_{k+r}(J_k^{(i)}) = \frac{N_r^{(i)}(\sigma^k\zeta)}{N_{k+r}(\zeta)}$$

where for each $i = 1, 2, \ldots, N_k(\zeta)$ the process $\{N_r^{(i)}(\sigma^k\zeta)\}_{r\geq 0}$ is, conditional on ζ, a replica of $\{N_r(\sigma^k\zeta)\}_{r\geq 0}$, and the different processes $i = 1, 2, \ldots, N_k(\zeta)$ are, conditional on ζ and $N_k(\zeta)$, independent. Note that

$$N_{k+r}(\zeta) = \sum_{i=1}^{N_k(\zeta)} N_r^{(i)}(\sigma^k\zeta).$$

It now follows from (3.1) that there are random variables $N_\infty^{(i)}(\sigma^k\zeta)$ and $N_\infty(\zeta)$ such that for any weak-$*$ limit λ of a subsequence of $\{\lambda_r\}_{r\geq 1}$,

$$\lambda(J_k^{(i)}) = \frac{N_\infty^{(i)}(\sigma^k\zeta)}{\sum_{j=1}^{N_k(\zeta)} N_\infty^{(j)}(\sigma^k\zeta)} = \frac{N_\infty^{(i)}(\sigma^k\zeta)}{N_\infty(\zeta)\prod_{\nu=1}^{k} E[N(\zeta_\nu)|\zeta]} \tag{5.10}$$

Using this representation, the hypothesis that $EN(i) > 1 \ \ \forall\, i \in \mathcal{J}$, and Theorem 3.2 we will prove (5.9). (Note: On the event $\{K(\zeta) \neq \phi\} = \{N_k(\zeta) \geq 1 \ \ \forall\, k\}$, the random variable $N_\infty(\zeta)$ is a.s. positive, by Theorem 3.1.)

The random variables $N_\infty(\zeta)$ and $N_\infty^{(i)}(\sigma^k\zeta)$ are not independent (even conditionally on ζ) but they *are* identically distributed. If they were in fact *bounded* random variables then it would follow from (5.10) and the fact that $N_\infty(\zeta) > 0$ on the event $\{K(\zeta) \neq \phi\}$ that for large k,

$$\log \lambda(J_k^{(i)}) \sim \sum_{\nu=1}^{k} \log E[N(\zeta_\nu)|\zeta] \sim k\psi'(\theta).$$

From this (5.9) could easily be deduced, since each interval $J_k^{(i)}$ has length n^{-k}.

Unfortunately, the random variables $N_\infty^{(i)}(\sigma^k\zeta)$ are *not* bounded. This is where our hypothesis that $EN(i) > 1 \;\; \forall\, i \in \mathcal{J}$ enters the argument: in conjunction with Theorem 3.2 (the case $r = 1$) it implies that $EN_\infty(\zeta)^2 < \infty$. Together with (3.1) this implies that $EN_\infty(\zeta) = 1$. Hence, by the Markov inequality, for every $\varepsilon > 0$ and $k = 1, 2, \ldots, P\{N_\infty(\zeta) \geq e^{\varepsilon k}\} \leq e^{-\varepsilon k}$, and so the Cauchy-Schwartz inequality implies

$$\sum_{k=1}^{\infty} E\{N_\infty(\zeta) 1\{N_\infty(\zeta) \geq e^{\varepsilon k}\}\} \leq \sum_{k=1}^{\infty} \{EN_\infty(\zeta)^2\}^{1/2} e^{-\varepsilon k/2} < \infty.$$

Fix $\varepsilon > 0$, and define $\mathcal{B}_k = \mathcal{B}_k^\varepsilon$ to be the set of indices i among $1, 2, \ldots, N_k(\zeta)$ such that $N_\infty^{(i)}(\sigma^k\zeta) \geq e^{\varepsilon k}$. By the result of the preceding paragraph and the fact that $N_\infty(\zeta)$ and $N_\infty^{(i)}(\sigma^k\zeta)$ have the same distribution,

$$E\left\{\sum_{k=1}^{\infty} \left(N_k(\zeta)^{-1} \sum_{i\in\mathcal{B}_k} N_\infty^{(i)}(\sigma^k\zeta)\right) 1\{N_k(\zeta) \geq 1\}\right\} < \infty.$$

This implies, in particular, that the random variable inside the braces {} is almost surely finite. Since $N_k(\zeta) / \prod_{\nu=1}^{k} E[N(\zeta_\nu)|\zeta] \to N_\infty(\zeta)$ a.s., by (3.1), it now follows from (5.10) that

$$\sum_{k=1}^{\infty} \sum_{i\in\mathcal{B}_k} \lambda(J_k^{(i)}) < \infty \tag{5.11}$$

almost surely on the event $\{K(\zeta) \neq \phi\}$.

The result (5.9) is now easily deduced from (5.10) and (5.11). Consider a point x in the support of λ: x is an element of $K(\zeta)$ so for each $k = 1, 2, \ldots$ the point x is an element of one (or two) of the intervals $J_k^{(i)}$. Call this interval $J_k(x)$. Then $J_k(x)$ abuts on zero, one, or two other $J_k^{(i)}$; call these $J_k'(x)$ and $J_k''(x)$ (if they exist). Inequality (5.11) and the Borel-Cantelli lemma, together with (5.10), imply that

$$\lambda\{x: \lambda(J_k(x)) + \lambda(J_k'(x)) + \lambda(J_k''(x)) \geq \frac{3e^{\varepsilon k}}{N_\infty(\zeta) \prod_{\nu=1}^{k} E[N(\zeta_\nu)|\zeta]},\; i.o.\} = 0$$

P-a.s. on the event $\{K(\zeta) \neq \phi\}$. By the SLLN, $\{\prod_{\nu=1}^{k} E[N(\zeta_\nu)|\zeta]\}^{1/k} \to e^{\psi'(\theta)}$ a.s. (see (5.3)), so it follows that

$$\lambda\{x: \lambda(J_k(x)) + \lambda(J'_k(x)) + \lambda(J''_k(x)) \geq \frac{3e^{2\varepsilon k}}{N_\infty(\zeta)} e^{-k\psi'(\theta)} \ i.o.\} = 0.$$

Finally, take x in the support of λ, and consider $\lambda(B(x,r))$ for $r > 0$ small. In proving (5.9) it is enough to consider $r = n^{-k}, k = 1, 2, \ldots$. If $r = n^{-k}$ then $B(x,r) \subset J_k(x) \cup J'_k(x) \cup J''_k(x)$. Consequently, by the result of the previous paragraph,

$$\liminf_{r \to 0} \frac{\log \lambda(B(x,r))}{\log r} \geq (\psi'(\theta) - \varepsilon)/\log n$$

for λ-a.e. x, P-a.s. on the event $\{K(\zeta) \neq \phi\}$. Since $\varepsilon > 0$ was arbitrary, (5.9) follows. ■

We now turn to the general case, where the hypothesis $EN(i) > 1 \ \forall\, i \in \mathcal{J}$ may fail. The proof of Lemma 5.4 breaks down in this case. To circumvent this difficulty we will replace the original family of measures μ_θ by families $\mu_\theta^{(r)}, r = 2, 3, \ldots$, for which the proof of Lemma 5.4 remains valid. For $r \geq 2$ define

$$\mathcal{J}_+^r = \{y_1 y_2 \ldots y_r \in \mathcal{J}^r : \prod_{\nu=1}^{r} g(y_\nu) > 1\}$$

and

$$\psi_r(\theta) = \log\left\{\sum_{\mathcal{J}_+^r} \prod_{\nu=1}^{r} g(y_\nu)^\theta\right\}$$

where $g(i) = EN(i)$ as earlier. Let $\xi_j^{(r)} = \xi_j, j \geq 1$, be iid $\mathcal{J}_+^r$-valued random variables with distributions

$$P\{\xi_j^{(r)} = y_1 y_2 \ldots y_r\} = e^{-\psi_r(\theta)} \prod_{\nu=1}^{r} g(y_\nu)^\theta,$$

and let $\zeta = \zeta_1\zeta_2\zeta_3\ldots$ be the sequence obtained by concatenating the finite sequences $\xi_1, \xi_2, \ldots$; thus $\xi_k = \zeta_{r(k-1)+1}\zeta_{r(k-1)+2}\ldots\zeta_{rk}$. Define $\mu_\theta^{(r)}$ to be the distribution of $\Sigma\zeta_j m^{-j}$.

Lemma 5.3*

Fix $\theta > t$ and let $G = G_r \subset [0,1]$ be any Borel set such that $\mu_\theta^{(r)}(G) = 1$. Then for all $r = 2, 3, \ldots$,

$$P\left(\delta_H(G \cap \operatorname{proj}_2 K) \geq \frac{\psi_r(\theta) - \theta\psi'_r(\theta)}{r \log m}\middle| K \neq \phi\right) > 0.$$

Proof: This is very similar to the proof of Lemma 5.3; however, in place of (5.8) it suffices here to prove that

$$P(\mu_\theta^{(r)}(\operatorname{proj}_2 K) > 0 | K \neq \phi) > 0.$$

This follows because by (3.5), $P\{\lim_{k\to\infty} N_k(\zeta) = \infty\} > 0$, which implies that $P\{K(\zeta) \neq \phi\} > 0$.

The rest of the proof is virtually identical to the second half of the proof of Lemma 5.3: on the event $\{\mu_\theta^{(r)}(\text{proj}_2 K) > 0\}$, define

$$\nu_\theta^{(r)}(A) = \frac{\mu_\theta^{(r)}(A \cap \text{ proj}_2 K)}{\mu_\theta^{(r)}(\text{proj}_2 K)}$$

and proceed as before. ■

Lemma 5.4*

Fix $\theta > t$. On the event $\{\mu_\theta^{(r)}(\text{proj}_2 K) > 0\}$,

$$\mu_\theta^{(r)}\{y \in [0,1] \colon \delta_H(K(y)) \geq \psi_r'(\theta)/r \log n\} = \mu_\theta^{(r)}(\text{proj}_2 K)$$

almost surely (P).

Proof: This is the same as that of Lemma 5.4. ■

The same arguments as used earlier now lead to a lower bound for $\delta_H(K)$ on the event $\{K \neq \phi\}$. By Lemma 5.3*–5.4* there exists, with positive P-probability, a Borel set G such that

$$\delta_H(G \cap \text{ proj}_2 K) \geq \frac{\psi_r(\theta) - \theta\psi_r'(\theta)}{r \log m}$$

and such that for every $y \in G \cap$ proj$_2 K$,

$$\delta_H(K(y)) \geq \psi_r'(\theta)/r \log n.$$

Consequently, Marstrand's theorem implies that with positive P-probability, for every $\theta > t$

(5.12) $$\delta_H(K) \geq \frac{\psi_r'(\theta)}{r \log n} + \frac{\psi_r(\theta) - \theta\psi_r'(\theta)}{r \log m}.$$

Lemma 5.5

For each $\theta > t$,

(5.13) $$\lim_{r\to\infty} r^{-1}\psi_r(\theta) = \psi(\theta)$$

and

(5.14) $$\lim_{r\to\infty} r^{-1}\psi_r'(\theta) = \psi'(\theta).$$

Proof: Recall that for $\theta > t, \psi'(\theta) = E \log g(\zeta_1) > 0$ (see (5.3)). Now

$$e^{r\psi(\theta)} = \sum\nolimits_{\mathcal{J}^r} \prod_{\nu=1}^{r} g(y_\nu)^\theta$$

(note that the sum is over all of $\mathcal{J}^r$, not just $\mathcal{J}_+^r$), so

$$\frac{\exp\{\psi_r(\theta)\}}{\exp\{r\psi(\theta)\}} = \frac{\sum_{\mathcal{J}_+^r} \prod_{\nu=1}^{r} g(y_\nu)^\theta}{\sum_{\mathcal{J}^r} \prod_{\nu=1}^{r} g(y_\nu)^\theta} = P\left\{\sum_{\nu=1}^{r} \log g(\zeta_\nu) > 0\right\}$$

where $\zeta_1, \zeta_2, \ldots$ are iid with distribution (5.2). But the SLLN (or WLLN) implies that this probability converges to 1 as $r \to \infty$, since $\psi'(\theta) = E \log g(\zeta_1) > 0$. This clearly implies (5.13). Since $\psi(\theta)$ and $\psi_r(\theta)$ are analytic functions of θ, (5.14) follows directly. ■

In view of Lemma 5.5, (5.12) implies that for every $\theta > t$, (5.7) holds with positive P-probability. But $\delta_H(K)$ is a.s. constant on the event $\{K \neq \phi\}$, so in fact (5.7) holds a.s. on $\{K \neq \phi\}$. Taking the supremum over $\theta > t$ now yields

$$\delta_H(K) \geq \psi(\alpha)/\log m$$

a.s. on $\{K \neq \phi\}$, as before. ■

B. The Upper Bound

The basic strategy here is the same as in $[\mathrm{Be}]_2$, but the details of the coverings are different, thanks to the randomness in the construction. We will partition K into subsets whose projections on the y-axis consist of points whose m-ary expansions are approximately "generic" for certain probability measures; and using again the results of section 3 we will construct efficient covers for each of these subsets.

For $y \in [0,1]$ with m-ary expansion $y_1 y_2 y_3 \ldots \in \mathcal{J}^\infty$ define the frequency distributions $f_k(y), k = 1, 2, \ldots$, as follows:

$$f_k(y) = (f_k^{(i)}(y))_{i \in \mathcal{J}},$$

$$f_k^{(i)}(y) = k^{-1} \sum_{\nu=1}^{k} 1\{y_\nu = i\}, i \in \mathcal{J}.$$

We shall only consider points y whose m-ary expansions are contained in $\mathcal{J}^\infty$ because only such y occur in $\mathrm{proj}_2 K$ (recall that $EN(i) = 0 \;\; \forall\, i \in \mathcal{I} \backslash \mathcal{J}$). For points y with multiple m-ary expansions there are frequency distributions for each expansion; we will treat these as essentially different points of [0, 1].

Define $\mathcal{P}$ to be the set of probability vectors on the index set $\mathcal{J}$: i.e., $\mathcal{P} = \{\boldsymbol{p} \in [0,1]^{\mathcal{J}} : \Sigma_{i \in \mathcal{J}} p_i = 1\}$. For each $\boldsymbol{p} \in \mathcal{P}$, define

$$H(\boldsymbol{p}) = -\sum_{i \in \mathcal{J}} p_i \log p_i$$

(with $0 \log 0 = 0$), and for $\delta > 0$ let $B_\delta(\boldsymbol{p})$ be the L^∞-ball of radius δ centered at $\boldsymbol{p}$, i.e., $B_\delta(\boldsymbol{p}) = \{\boldsymbol{q} \in \mathcal{P}: |p_i - q_i| < \delta \ \ \forall\, i \in \mathcal{J}\}$. For $\delta > 0$ and $\boldsymbol{p} \in \mathcal{P}$ define

$$\begin{aligned}
\mu(\boldsymbol{p}) &= \textstyle\sum_{i \in \mathcal{J}} p_i \log EN(i); \\
d_1(\boldsymbol{p}) &= \frac{H(\boldsymbol{p})}{\log m} + \frac{\mu(\boldsymbol{p})}{\log n}; \\
d_2(\boldsymbol{p}) &= \frac{H(p) + \mu(\boldsymbol{p})}{\log m}; \\
A(\boldsymbol{p}, \delta) &= \{y \in [0,1]: f_k(y) \in B_\delta(\boldsymbol{p}) \ \ i.o.\}; \\
A^*(\boldsymbol{p}, \delta) &= \left\{ y \in A(\boldsymbol{p}, \delta): \limsup_{k \to \infty} \mu(f_k(y)) \le \sup_{\boldsymbol{q} \in B_\delta(\boldsymbol{p})} \mu(\boldsymbol{q}) \right\}; \\
F^*(\boldsymbol{p}, \delta) &= \{(x, y) \in K: y \in A^*(\boldsymbol{p}, \delta)\}.
\end{aligned}$$

Observe that $d_2(\boldsymbol{p})$ is greater than, less than, or equal to $d_1(\boldsymbol{p})$ according as $\mu(\boldsymbol{p})$ is greater than, less than, or equal to 0. Observe also that for any fixed $\delta > 0$, the unit interval [0, 1] may be covered by finitely many of the sets $A^*(\boldsymbol{p}, \delta), \boldsymbol{p} \in \mathcal{P}$, and consequently K may be covered by finitely many $F^*(\boldsymbol{p}, \delta)$. Thus, the problem of obtaining an upper bound for $\delta_H(K)$ essentially reduces to that of obtaining upper bounds for the sets $F^*(\boldsymbol{p}, \delta)$.

Lemma 5.6

For each $\varepsilon > 0$ there exists $\delta > 0$ such that for every $\boldsymbol{p} \in \mathcal{P}$, with probability one,

$$\delta_H(F^*(\boldsymbol{p}, \delta)) \le \max(0, \min(d_1(\boldsymbol{p}), d_2(\boldsymbol{p}))) + \varepsilon. \tag{5.15}$$

Proof: We will show that $\delta_H(F^*(\boldsymbol{p}, \delta)) \le \max\{0, d_1(\boldsymbol{p})\} + \varepsilon$ and $\delta_H(F^*(\boldsymbol{p}, \delta)) \le \max\{0, d_2(\boldsymbol{p})\} + \varepsilon$. We will begin with the second inequality which is simpler.

The condition defining $A^*(\boldsymbol{p}, \delta)$ is not needed for this inequality and so we will show something slightly stronger than $\delta_H(F^*(\boldsymbol{p}, \delta)) \le \max\{0, d_2(\boldsymbol{p})\} + \varepsilon$. We will show that $\delta_H(F(\boldsymbol{p}, \delta)) \le \max\{0, d_2(\boldsymbol{p})\} + \varepsilon$ where

$$F(\boldsymbol{p}, \delta) = \{(x, y) \in K: y \in A(p, \delta)\}.$$

Since clearly $F^*(\boldsymbol{p}, \delta) \subset F(\boldsymbol{p}, \delta)$ this will be sufficient.

For any finite sequence $s \in \mathcal{J}^k$ define the frequency distributions $f_r(s) = (f_r^{(i)}(s))_{i \in \mathcal{J}}$, for $1 \le r \le k$, in the same manner as for infinite sequences, specifically,

$$f_r^{(i)}(s) = \frac{1}{r} \sum_{\nu=1}^{r} 1\{s_\nu = r\}, \quad i \in \mathcal{J}.$$

For $k = 1, 2, \ldots$ define

$$\begin{aligned}
A_k(\boldsymbol{p}, \delta) &= \{s \in \mathcal{J}^k: f_k(s) \in B_\delta(\boldsymbol{p})\}, \\
F_k(\boldsymbol{p}, \delta) &= \{(x, y) \in K_k: y_1 y_2 \ldots y_k \in A_k(\boldsymbol{p}, \delta)\}.
\end{aligned}$$

Clearly

$$F(\boldsymbol{p},\delta) \subset \bigcup_{k\geq K} F_k(\boldsymbol{p},\delta)$$

for all $k \in \mathbb{N}$.

Consider the covering of $F_k(\boldsymbol{p},\delta)$ consisting of all k^{th} generation rectangles lying in "rows" indexed by sequences $s \in A_k(\boldsymbol{p},\delta)$; call this covering $\mathcal{U}_k$. Thus

$$\mathcal{U}_k = \{[jm^{-k},(j+1)m^{-k}] \times I_s \subseteq K_k : s \in A_k(\boldsymbol{p},\delta)\}.$$

Each element of $\mathcal{U}_k$ is a rectangle with height m^{-k} and width n^{-k} and therefore has diameter comparable to m^{-k} (recall $m < n$). The cardinality of $\mathcal{U}_k$ is

$$\#\mathcal{U}_k = \sum_{s\in A_k(\boldsymbol{p},\delta)} N_k(s).$$

Now notice that if $s \in A_k(\boldsymbol{p},\delta)$ then $\mu(f_k(s)) \leq \mu(\boldsymbol{p}) + \gamma\delta$, where $\gamma = \Sigma_{i\in\mathcal{J}}|\log EN(i)$ is a constant independent of k and $\boldsymbol{p}$. Thus for each $s \in A_k(\boldsymbol{p},\delta)$

$$EN_k(s) = \exp\{k\mu(f_k(s))\} \leq \exp\{k\mu(\boldsymbol{p}) + k\gamma\delta\}.$$

Furthermore, by standard estimates, the cardinality of $A_k(\boldsymbol{p},\delta)$ does not exceed $\exp\{kH(\boldsymbol{p}) + d\gamma'(\delta)\}$, where $\gamma'(\delta) > 0$ may be chosen so that $\gamma'(\delta) \to 0$, as $\delta \to 0$, uniformly in $\boldsymbol{p}$. It now follows that

$$E[\#\mathcal{U}_k] \leq \exp\{k(\mu(\boldsymbol{p}) + H(\boldsymbol{p})) + k\beta(\delta)\},$$

where $\beta(\delta) = \gamma\delta + \gamma'(\delta)$, and hence

$$\lim_{\delta\to 0} \beta(\delta) = 0 \text{ uniformly in } \boldsymbol{p}.$$

Consequently, by the Markov inequality and the Borel-Cantelli lemma,

$$P(\#\mathcal{U}_k \geq \exp\{k(\mu(\boldsymbol{p}) + H(\boldsymbol{p})) + k\beta(\delta)\}, \quad i.o.) = 0,$$

and hence, for any $d > d_2(\boldsymbol{p}) + 2\beta(\delta)/\log m$ and $d > 0$,

$$P\left(\sum_{k=1}^{\infty} m^{-kd}(\#\mathcal{U}_k) < \infty\right) = 1 \tag{5.16}$$

Fix $\rho > 0$, and let $k(\rho)$ be such that $m^{-k(\rho)} < \rho \leq m^{-k(\rho)-1}$. Since each $\mathcal{U}_k$ is a covering of $F_k(\boldsymbol{p},\delta)$ and since $(x,y) \in F(\boldsymbol{p},\delta)$ implies that $y \in A(\boldsymbol{p},\delta)$ (so $f_k(y) \in B_\delta(\boldsymbol{p})$ infinitely often), $\cup_{k\geq k(\rho)}\mathcal{U}_k$ is a covering of $F(\boldsymbol{p},\delta)$ for each $\rho > 0$. Furthermore, each $\mathcal{U}_k$ consists of finitely many rectangles, each of maximum sidelength m^{-k} and hence of diameter $< \sqrt{2}m^{-k}$. Thus each element of $\cup_{k\geq k(\rho)}\mathcal{U}_k$ has diameter $< \sqrt{2}\rho$ and by (5.16) we have that, with probability one,

$$H_d(F(\boldsymbol{p},\delta)) \leq \lim_{\rho\to 0} \sum_{k=k(\rho)}^{\infty} m^{-kd}(\#\mathcal{U}_k) = 0$$

for every $d > d_2(\boldsymbol{p}) + 2\beta(\delta)/\log m$ and $d > 0$. Since $\beta(\delta) \to 0$, as $\delta \to 0$, uniformly in $\boldsymbol{p}$, this proves that for sufficiently small $\delta > 0$,

$$\delta_H(F^*(\boldsymbol{p},\delta)) \leq \delta_H(F(\boldsymbol{p},\delta)) \leq \max\{0, d_2(\boldsymbol{p})\} + \varepsilon \quad a.s.$$

For the other inequality implicit in (5.15) we need to restrict ourselves to the y's in $A^*(\boldsymbol{p},\delta)$. The proof is similar, but requires a less obvious covering, this by the approximate squares that were used in secton 4. As in section 4, for $l = 1, 2, \ldots$, let $k = k_l = [l \log_m n]$, and define

$$A_l^*(\boldsymbol{p},\delta) = \left\{ s \in A_k(\boldsymbol{p},\delta)\colon\ \mu(f_l(s)) \leq \sup_{\boldsymbol{q}\in B_\delta(\boldsymbol{p})} \mu(\boldsymbol{q}) + \delta \right\},$$

$$F_l^*(\boldsymbol{p},\delta) = \{(x,y) \in K_k\colon\ y_1y_2\ldots y_k \in A_l^*(\boldsymbol{p},\delta)\}$$

Once again, for every $L = 1, 2, \ldots$,

$$F^*(\boldsymbol{p},\delta) \subset \bigcup_{l\geq L} F_l^*(\boldsymbol{p},\delta).$$

Let $\mathcal{V}_l$ be the set of all l^{th} generation approximate squares whose interiors intersect $F_l^*(\boldsymbol{p},\delta)$; then clearly $\mathcal{V}_l$ is a covering of $F_l^*(p,\delta)$ by rectangles of height m^{-k} and width $n^{-l} \approx m^{-k}$. An l^{th} generation approximate square is included in $\mathcal{V}_l$ iff (1) it contained a rectangle of K_k, and (2) it is contains in a "row" [0, 1] $\times I_s$, for some $s \in A_l^*(\boldsymbol{p},\delta)$. Now an l^{th} generation approximate square contained in a "row" $[0,1] \times I_s$, for some $s \in \mathcal{J}^k$, can only contain a rectangle of K_k if it is contained in one of the rectangles of K_l and so the number of "nonempty" l^{th} generation approximate squares contained in the "row" $[0,1] \times I_s$ is at most $N_l(s)$. Consequently,

$$\#\mathcal{V}_l = \sum_{s\in A_l^*(\boldsymbol{p},\delta)} N_l(s).$$

As earlier,

$$EN_l(s) = \exp\{l\mu(f_l(s))\} \leq \exp\left\{ l \sup_{\boldsymbol{q}\in B_\delta(\boldsymbol{p})} \mu(\boldsymbol{q}) + l\delta \right\} \leq \exp\{l\mu(\boldsymbol{p}) + l(\gamma+1)\delta\}$$

and since $\#A_k^*(\boldsymbol{p},\delta) \leq \#A_k(\boldsymbol{p},\delta)$ we obtain the estimate

$$E[\#\mathcal{V}_l] \leq \exp\{kH(\boldsymbol{p}) + l\mu(\boldsymbol{p}) + kc(\delta)\}$$

where again $c(\delta) \to 0$, as $\delta \to 0$, uniformly in $\boldsymbol{p}$.

Now consider the coverings $\cup_{l \geq L} \mathcal{V}_l, L = 1, 2, \ldots$ of $F^*(\boldsymbol{p}, \delta)$ to conclude as in the preceding case, that for $\delta > 0$ sufficiently small

$$\delta_H(F^*(\boldsymbol{p}, \delta)) \leq \max\{d_1(\boldsymbol{p}), 0\} + \varepsilon \quad a.s. \qquad \blacksquare$$

An upper bound for $\delta_H(K)$ is now easily obtained. Choose $\delta > 0$ sufficiently small that the estimate (5.15) of Lemma 5.6 is valid. Recall that K is the union of finitely many $F^*(\boldsymbol{p}, \delta), \boldsymbol{p} \in \mathcal{P}$, from which it follows that $\delta_H(K)$ is bounded above by $\max_{\boldsymbol{p}} \delta_H(F^*(\boldsymbol{p}, \delta))$. Therefore, by (5.15),

$$\delta_H(K) \leq \sup_{\boldsymbol{p} \in \mathcal{P}} \max(0, \min(d_1(\boldsymbol{p}), d_2(\boldsymbol{p}))) + \varepsilon,$$

and since $\varepsilon > 0$ is arbitrary, we may in fact set $\varepsilon = 0$. It remains only to establish

Lemma 5.7

$$\sup_{\boldsymbol{p} \in \mathcal{P}} \min(d_1(\boldsymbol{p}), d_2(\boldsymbol{p})) = \psi(\alpha)/\log m > 0.$$

Proof: Both $d_1(\boldsymbol{p})$ and $d_2(\boldsymbol{p})$ are linear combinations of $H(\boldsymbol{p})$ and $\mu(\boldsymbol{p})$ with positive coefficients. Consider first $\mu(\boldsymbol{p}) = \Sigma_{i \in \mathcal{J}} p_i \log EN(i)$: as $\boldsymbol{p}$ varies over the simplex $\mathcal{P}, \mu(\boldsymbol{p})$ varies continuously over the interval $[a, b]$, where $a = \min_{i \in \mathcal{J}} \log EN(i)$ and $b = \max_{i \in \mathcal{J}} \log EN(i)$. Consider $\boldsymbol{p}^\theta$, defined by

$$p_i^\theta = e^{-\psi(\theta)}(EN(i))^\theta, i \in \mathcal{J}.$$

As θ varies over $\mathbf{R}, \mu(\boldsymbol{p}^\theta)$ varies over (a, b) (unless $a = b$, in which case $\mu(\boldsymbol{p}) = a \;\; \forall\, \boldsymbol{p}$); in particular, for each $c \in (a, b)$ there exists $\theta \in \mathbf{R}$ such that $\mu(\boldsymbol{p}^\theta) = c$. Now consider the maximization of $H(\boldsymbol{p})$ over all $\boldsymbol{p}$ satisfying $\mu(\boldsymbol{p}) = c$. Lagrange multipliers shows that for each $c \in (a, b)$ the maximum must occur at some $\boldsymbol{p}^\theta, \theta \in \mathbf{R}$. (Note: If $a = b$ then for $c = a$ the max occurs at the uniform distribution on $\mathcal{J}$, which coincides with $\boldsymbol{p}^0$.) This proves that

$$\sup_{\boldsymbol{p} \in \mathcal{P}} \min(d_1(\boldsymbol{p}), d_2(\boldsymbol{p})) = \sup_{\theta \in \mathbf{R}} \min(d_1(\boldsymbol{p}^\theta), d_2(\boldsymbol{p}^\theta)).$$

Next, consider the functions $D_1(\theta) = d_1(\boldsymbol{p}^\theta)$ and $D_2(\theta) = d_2(\boldsymbol{p}^\theta)$. Evidently,

$$D_1(\theta) = \frac{\psi(\theta) - \theta\psi'(\theta)}{\log m} + \frac{\psi'(\theta)}{\log n}$$

and

$$D_2(\theta) = \frac{\psi(\theta) - \theta\psi'(\theta) + \psi'(\theta)}{\log m},$$

so the minimum $D(\theta)$ of $D_1(\theta), D_2(\theta)$ is given by

$$D(\theta) = \begin{cases} \dfrac{\psi(\theta) + (1-\theta)\psi'(\theta)}{\log m}, & \theta \le t; \\ \dfrac{\psi(\theta) - \theta\psi'(\theta)}{\log m} + \dfrac{\psi'(\theta)}{\log n}, & \theta \ge t. \end{cases}$$

It is now a routine exercise in calculus to verify that the maximum of $D(\theta)$ occurs at $\theta = \alpha$, and that $D(\alpha) = \psi(\alpha)/\log m$. (Note: This uses the assumption $\psi'(1) > 0$, which implies that $0 \le t < 1$.)

Finally, observe that if $\alpha = t$ then $D(\alpha) = \psi(t)/\log m$ which, by Proposition 3.4, is positive; and if $\alpha = \log_n m > t$ then $D(\alpha) \ge D(t) = \psi(t)/\log m > 0$. ∎

6. When is $\delta_H = \delta_B$?

Proposition 6.1

$\delta_H = \delta_B$ a.s. iff $t = 1$ or $EN(i) = EN(j)$ for all $i, j \in \mathcal{J}$. (Recall $\mathcal{J}$ consists of those indices i for which $EN(i) > 0$.)

Proof: If $t = 1$ then $\delta_H = \delta_B$ follows by directly comparing (2.1) and (2.2) and observing that $t = 1$ also implies $\alpha = 1$.

If $t < 1$ and $EN(i) = \beta$ for all $i \in \mathcal{J}$ and some $\beta \in (0, +\infty)$ then ψ is linear with slope $\log \beta$. By linearity, $t < 1$ implies that $t = 0$ must be the case. Thus $\psi(t) = \psi(0) = \log(\#\mathcal{J})$ and $\alpha = \log_n m$. Hence, a.s. on $\{K \ne \emptyset\}$.

$$\delta_H(K) = \log_m \left\{ \sum_{i \in \mathcal{J}} \beta^{\log_n m} \right\} = \log_m(\#\mathcal{J}) + \log_n \beta = \delta_B(K).$$

It follows that $\delta_B = \delta_H$ a.s. (since on $\{K = \emptyset\}, \delta_B(K) = \delta_H(K) = 0$).

Conversely, suppose that $\delta_B(K) = \delta_H(K)$ a.s. and that $t < 1$. First we show that we must have $t < \log_n m$. Indeed, a.s. on $\{K \ne \phi\}$,

$$\delta_B(K) = \delta_H(K) + \frac{\log EM_1 - \psi(t)}{\log n} \tag{6.1}$$

if $t \ge \log_n m$. Now ψ is either strictly convex or linear with all $EN(i)$ having a common value > 1 for $i \in \mathcal{J}$ (recall we are assuming $t < 1$). In the latter case ψ has positive slope, so t must be 0. If ψ is strictly convex then it is strictly increasing on $[t, 1]$, so $\psi(1) - \psi(t) = \log EM_1 - \psi(t) > 0$ and $\delta_B(K) > \delta_H(K)$, a.s. on $\{K \ne \emptyset\}$, by (6.1), which contradicts the fact $\delta_B(K) = \delta_H(K)$ a.s. Thus, $t < \log_n m$.

Now $\delta_B(K) = \delta_H(K)$ a.s. implies that

$$\log_n EM_1 + \frac{\psi(t)}{\log m} - \frac{\psi(t)}{\log n} = \frac{\psi(\log_n m)}{\log m} \tag{6.2}$$

Assume that the $EN(i)$, for $i \in \mathcal{J}$, are not all the same. Then ψ is strictly convex and so

$$\psi(\log_n m) < \frac{1-\log_n m}{1-t}\psi(t) + \frac{\log_n m - t}{1-t}\psi(1)$$

or equivalently

$$t\psi(1) + (1-t)\psi(\log_n m) < \left[\frac{\psi(t)}{\log m}(1-\log_n m) + \frac{\psi(1)}{\log n}\right]\log m. \tag{6.3}$$

But ψ is strictly convex and $t < \log_n m < 1$. This implies that $\psi(1) \geq \psi(\log_n m)$ or

$$t\psi(1) + (1-t)\psi(\log_n m) \geq \psi(\log_n m) \tag{6.4}$$

As (6.3) and (6.4) together contradict (6.2) (recall $EM_1 = \psi(1)$) we must have that all the $EN(i), i \in \mathcal{J}$, are equal. ■

REFERENCES

[AK]$_1$ K.B. Athreya, and S. Karlin: On branching processes in random enviroments: I. Extinction probability, *Ann. Math. Stat.*, **42** (1971), 1499–1520.

[AK]$_2$ K.B. Athreya, and S. Karlin: On branching processes in random environments: II. Limit theorems, *Ann. Math. Stat.*, **42** (1971), 1843–1858.

[AN] K.B. Athreya, and P.E. Ney: Branching processes, *Die Grundlehren der mathematischen Wissenschaften in Einzeldarstelung*, **196**, Springer-Verlag, New York, Heidelberg Berglin (1972).

[Be]$_1$ T. Bedford: Crinkly curves, Markov partitions,and dimension. Ph. D. dissertation, University of Warwick (1984).

[Be]$_2$ T. Bedford: On Weierstrass-like functions and random recurrent sets, *Math. Proc. Camb. Phil. Soc.*, **106** (1989), 325–342.

[De] F.M. Dekking: On the survival probability of a branching process in a finite state i.i.d. random environment, *Stoch. Proc. and Appl.*, **27** (1988), 151–157.

[DG] F.M. Dekking and G.R. Grimmett: Superbranching processes and projections of random Cantor sets, *Probability Theory and Related Fields*, **78** (1988), 335–355.

[Fa]$_1$ K.J. Falconer: The Hausdorff dimension of self-affine fractals, *Math. Proc. Camb. Phil. Soc.*, **103** (1988), 339–350.

[Fa]$_2$ K.J. Falconer: Random Fractals, *Math. Proc. Camb. Phil. Soc.*, **100** (1986), 559–582.

[Fa]$_3$ K.J. Falconer:The Geometry of Fractal Sets, Cambridge University Press, Cambridge (1985).

[Fa]$_4$ K.J. Falconer: Projections of Random Cantor Sets, *Journal of Theoretical Probability*, **2** (1989), 65–70.

[GL] S.P. Lalley, and D. Gatzouras: Hausdorff and box dimensions of certain self-affine fractals, *Indiana Univ. Math. J.*, to appear.

[Hu] J. Hutchinson: Fractals and Self-Similarity, *Indiana Univ. Math. J.*, **30** (1981), 713–747.

[Ma]$_1$ B. Mandelbrot: The Fractal Geometry of Nature, Freeman, San Francisco (1983).

[Ma]$_2$ B. Mandelbrot: Self-Affine Fractal Sets III: Hausdorff dimension anomalies and their implications, in *Fractals in Physics*, L. Pietronero, E. Tosatti (editors), Edlsevier Science Publishers, 21–28.

[Mc] C. McMullen: The Hausdorff dimension of general Sierpinski Carpets, *Nagoya Math J.*, **96** (1984), 1–9.

[MW] R.D. Mauldin, and S.C. Williams: Random recursive constructions: Asymptotic geometric and topological properties, *Trans. of the AMS*, **295** (1986), 325–346.

[Yo] Lai-Sang Young: Dimension, entropy, and Lyapunov exponents, *Ergodic Theory and Dynamical Systems,* **2** (1982), 102–124.

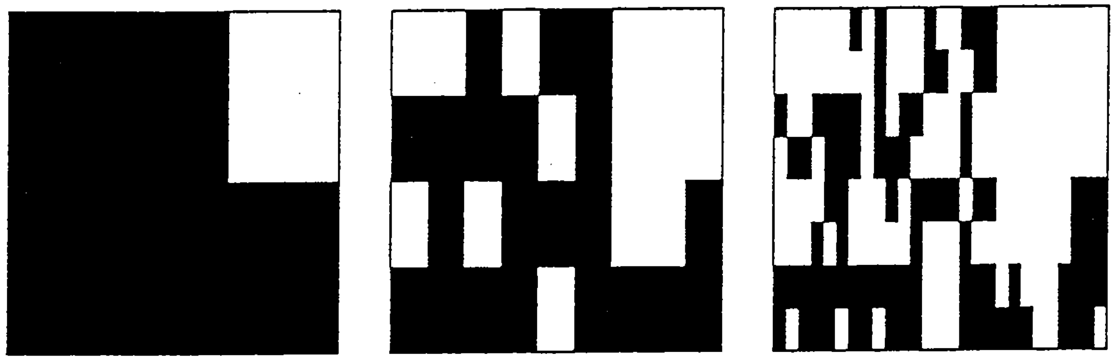

Figure 1: This shows 3 generations in the construction of a statistically self-affine set. In each generation, every surviving rectangle is divided into 6 congruent rectangles, arranged in 2 rows and 3 columns. Each of these new rectangles is then discarded with probability .3 .