Random subsets of Cantor sets generated by trees of coin flips

Pieter C. Allaart

University of North Texas
(allaart@unt.edu)

Geometry and Fractals under the Midnight Sun
Oulu, June 27, 2024

(joint with Taylor Jones)

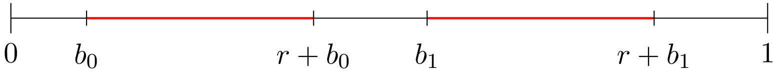
Preprint: arXiv:2308.04569

The Cantor set

Fix $r \in (0, 1/2]$. Consider the IFS $\{f_0, f_1\}$ given by

$$f_0(x) = rx + b_0, \qquad f_1(x) = rx + b_1,$$

where $b_0 < b_1$ and $\{f_0, f_1\}$ satisfies the Open Set Condition with $(0, 1)$ as the open set.



Let C be the attractor of the IFS. Then C is:

- a Cantor set if $r < 1/2$;
- the interval $[0, 1]$ if $r = 1/2$.

Either way,

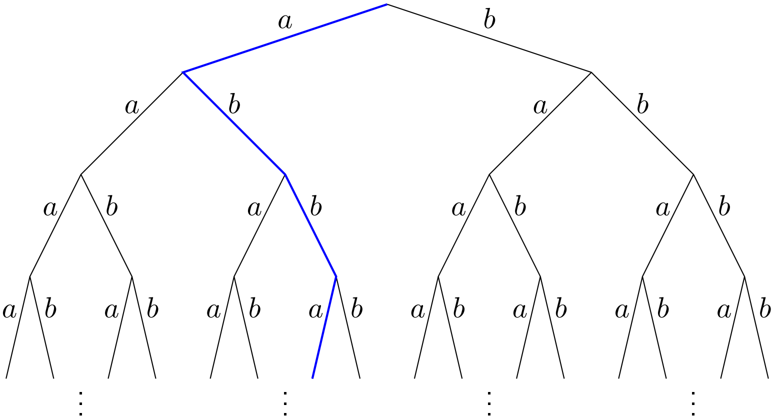
$$\dim_H C = \dim_B C = -\frac{\log 2}{\log r}.$$

The grand plan

We are going to construct a random subset of C by flipping coins... but not in the usual way!

The binary tree

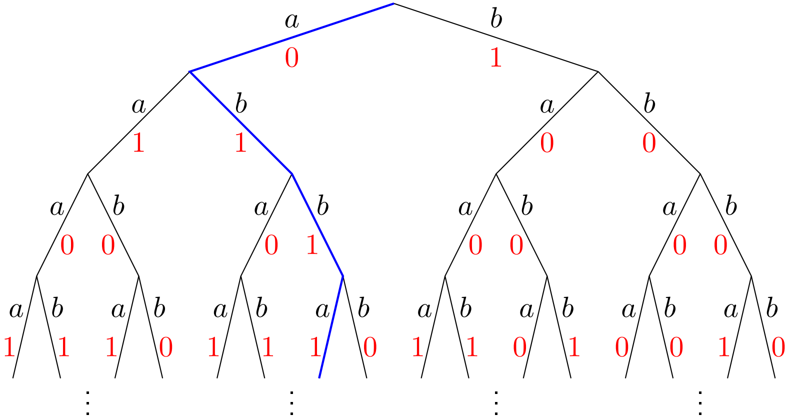
Consider a full infinite binary tree:



The blue path has coding $(a, b, b, a, \ldots)$.

The binary tree with random labeling

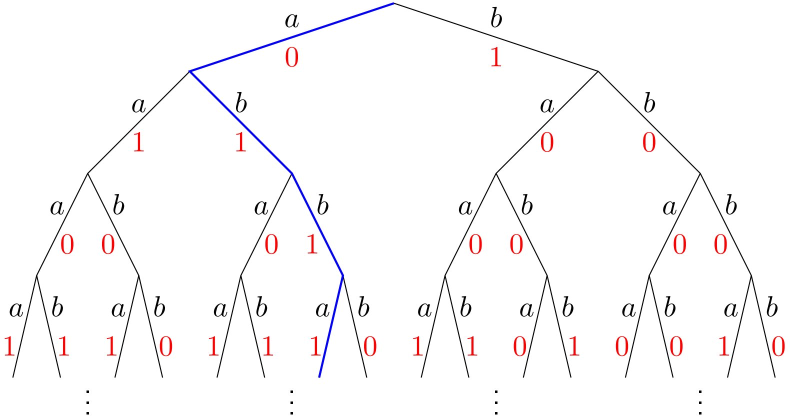
We label the edges independently: 0 with probability p, or 1 with probability $1-p$.



The blue path has labeling $(0,1,1,1,\dots)$.

If a path from the root has labeling $X_1, X_2, \ldots$, we associate with it a point

$$x := \lim_{n\to\infty} f_{X_1} \circ f_{X_2} \circ \cdots \circ f_{X_n}([0,1]) \in C.$$



The blue path leads to a point in $f_0 \circ f_1^3([0,1])$.

The random subset F

- Let $\mathcal{E}$ denote the set of edges of the binary tree.
- Let $\Omega := \{0,1\}^{\mathcal{E}}$ be the set of all labelings of the tree.
- We equip Ω with the product topology and μ_p, the Bernoulli$(p, 1-p)$ measure on its Borel σ-algebra.
- For a labeling $\omega \in \Omega$, each infinite path from the root down the tree gives a point in C as explained on the previous slide.
- We denote the set of all these points by $F(\omega)$.
- This makes F a random subset of C. It is by construction non-empty!

Example

For the sample labeling ω on the previous slide, $F(\omega)$ does not intersect $f_0^2((0,1))$ or $f_1^2((0,1))$. It does intersect $f_0 \circ f_1 \circ f_0 \circ f_1([0,1])$.

Branching random walk

- Helpful analogy: Branching random walk.
- Start with a single ancestor at generation 0 occupying the interval $[0,1]$.
- In each generation, each individual produces two children, which each receive a random label, 0 or 1.
- A child with label 0 moves to the left subinterval; a child with label 1 to the right.
- In the nth generation, there are 2^n individuals. Their positions form some (random) distribution over the 2^n basic intervals of C at level n.
- Thanks to *Sascha Troscheit* for pointing out this connection to us!

Remark

Our construction is not quite new: It was considered in a paper by I. Benjamini, O. Gurel-Gurevich & B. Solomyak in 2009, but with a different focus.

The dimension of F

Theorem (A. & Jones, 2023)

We have, μ_p-almost surely,

$$\frac{\log(p^2+(1-p)^2)}{\log r} \le \dim_H F \le \overline{\dim}_B F \le h(\xi(p)),$$

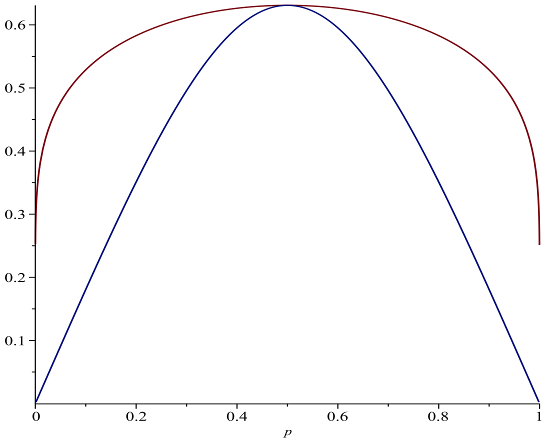
where

$$h(x) := \frac{x\log x + (1-x)\log(1-x)}{\log r}$$

and

$$\xi(p) := \frac{\log(2p)}{\log p - \log(1-p)}.$$

Graphs of the bounds



The upper and lower bounds for $\dim_H F$, when $r = 1/3$.

Corollary

If $p = 1/2$*, then* $\dim_H F = \dim_B F = -\log 2/\log r$*,* μ_p*-a.s.*

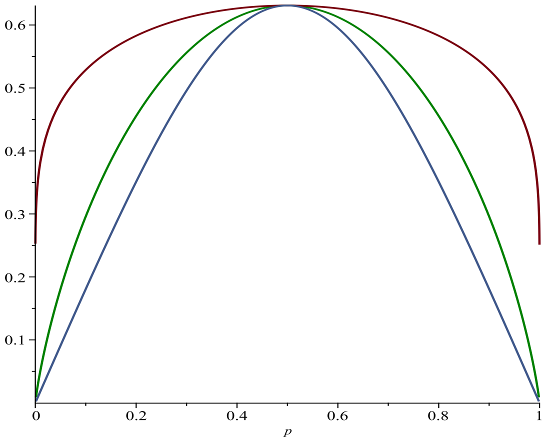
A reasonable guess

We might guess the true dimension of F to be

$$h(p) = \frac{p \log p + (1-p)\log(1-p)}{\log r}.$$

- This is the Hausdorff dimension of the set of points in C whose codings contain the digit 0 with frequency p, and 1 with frequency $1-p$.
- Indeed, $h(p)$ lies between the upper and lower bound!

The bounds and our guess



The upper and lower bounds, with our guess in green.

Key ideas of the proof – notation

Definition

For a label sequence $\mathbf{i} \in \{0,1\}^n$ and path $\mathbf{j} \in \{a,b\}^n$, let

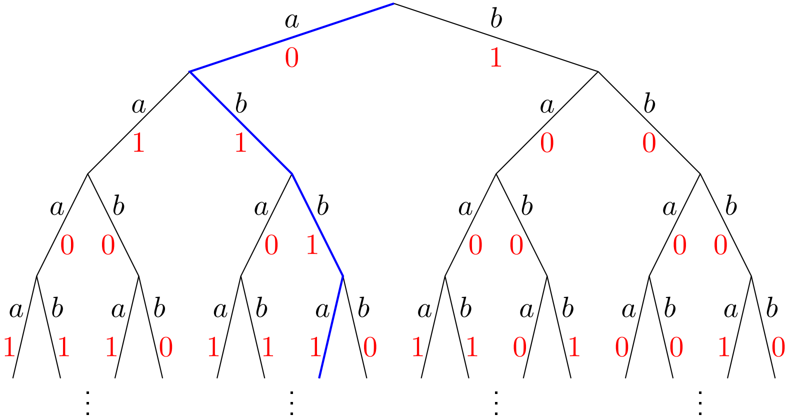
$$A(\mathbf{i},\mathbf{j}) := \{\omega \in \Omega : \text{ the path } \mathbf{j} \text{ receives label sequence } \mathbf{i} \text{ from } \omega\}.$$

Example

For instance, in our sample labeling, the blue path shows that $\omega \in A(0111, abba)$.

The binary tree with random labeling

We label the edges independently: 0 with probability p, or 1 with probability $1-p$.



The blue path shows that $\omega \in A(0111, abba)$.

Key ideas of the proof – lower bound

- Write $I_{\mathbf{i}} := f_{i_1} \circ \cdots \circ f_{i_n}([0,1])$. Define a (random) measure m_ω on $F(\omega)$ by
$$m_\omega(I_{\mathbf{i}}) := \frac{1}{2^n} \sum_{\mathbf{j} \in \{a,b\}^n} 1_{A(\mathbf{i},\mathbf{j})}(\omega).$$
In terms of the BRW, this is the proportion of individuals in generation n that occupy the interval $I_{\mathbf{i}}$.
- We estimate the energy
$$\Phi_t(m_\omega) := \iint_{[0,1]\times[0,1]} |x-y|^{-t} dm_\omega(x) dm_\omega(y),$$
and show that, for
$$t < \frac{\log(p^2 + (1-p)^2)}{\log r},$$
we have $\mathbb{E}_p\left[\Phi_t(m_\omega)\right] < \infty$, so that $\Phi_t(m_\omega) < \infty$, μ_p-a.s.

Key ideas of the proof – upper bound

- For $\mathbf{i} \in \{0,1\}^n$, let

$$\begin{aligned} a_{\mathbf{i}} &:= \mu_p(\text{at least one path of length } n \text{ has label sequence } \mathbf{i}) \\ &= \mu_p\left(\bigcup_{\mathbf{j}\in\{a,b\}^n} A(\mathbf{i},\mathbf{j})\right). \end{aligned}$$

- Let Z_n be the number of basic intervals at level n needed to cover F. Then
$$\mathbb{E}_p(Z_n) = \sum_{\mathbf{i}\in\{0,1\}^n} a_{\mathbf{i}}.$$

- We simply estimate

$$a_{\mathbf{i}} \leq \sum_{\mathbf{j}\in\{a,b\}^n} \mu_p(A(\mathbf{i},\mathbf{j})) = 2^n \mu_p(A(\mathbf{i},\cdot)).$$

(How much do we "throw away" by ignoring the overlaps?)

Key ideas of the proof – upper bound

- If $\mathbf{i}$ has k 1s and $n-k$ 0s, then

$$\mu_p(A(\mathbf{i},\cdot)) = p^k(1-p)^{n-k},$$

so

$$a_{\mathbf{i}} \le 2^n \mu_p(A(\mathbf{i},\cdot)) = 2^n p^k (1-p)^{n-k}.$$

- Thus,

$$\mathbb{E}_p(Z_n) \le \sum_{k=0}^{n} \binom{n}{k} \min\left\{2^n p^k (1-p)^{n-k}, 1\right\}.$$

- It can be shown that this bound grows exponentially at rate

$$\left[\xi^{-\xi}(1-\xi)^{-(1-\xi)}\right]^n,$$

where

$$\xi = \xi(p) = \frac{\log(2p)}{\log p - \log(1-p)}.$$

Hausdorff measure in the symmetric case

Recall: When $p = 1/2$, we have

$$\dim_H F = \dim_H C = s := -\frac{\log 2}{\log r}, \qquad \mu_p - a.s.$$

Proposition (A. & Jones, 2023)

When $p = 1/2$,

$$\mathcal{H}^s(F) = 0, \qquad \mu_p - a.s.$$

A generalization

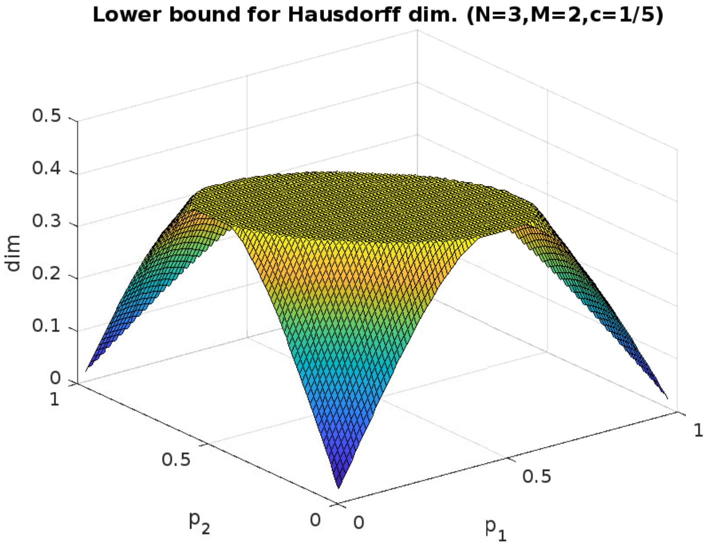
- We can generalize our construction, by considering N similarities $f_1, \ldots, f_N$ with ratio r satisfying the OSC, and an M-ary tree.
- Let $\mathbf{p} = (p_1, \ldots, p_N)$ be a probability vector, and equip the space of labelings $\Omega = \{1, 2, \ldots, N\}^{\mathcal{E}}$ with the Bernoulli$(\mathbf{p})$ measure.
- This generates a random subset $F_{N,M}$ of C.

Theorem (A. & Jones, 2024)

We have the lower bound

$$\dim_H F_{N,M} \geq \min\left\{-\frac{\log M}{\log r}, \frac{\log(p_1^2 + \cdots + p_N^2)}{\log r}\right\} \qquad \mu_{\mathbf{p}}\text{-a.s.}$$

Graph of the lower bound



Generalization: Upper bound

We trivially have:

$$\overline{\dim}_B F_{N,M} \leq \min\left\{-\frac{\log N}{\log r}, -\frac{\log M}{\log r}\right\}.$$

Theorem (A. & Jones, 2024)

If

$$\prod_{i=1}^{N} p_i^{-p_i} \leq M \leq \left(\prod_{i=1}^{N} p_i\right)^{-1/N},$$

then there is a unique $\lambda \in [0,1]$ *satisfying*

$$\sum_{i=1}^{N} p_i^{\lambda} \log(M p_i) = 0,$$

and

$$\overline{\dim}_B F_{N,M} \leq -\frac{\lambda \log M + \log\left(\sum_{i=1}^{N} p_i^{\lambda}\right)}{\log r}, \qquad \mu_{\mathbf{p}}\text{-a.s.}$$

Some remarks

1. The inequalities

$$\prod_{i=1}^{N} p_i^{-p_i} \leq M \leq \left(\prod_{i=1}^{N} p_i \right)^{-1/N}$$

are always satisfied when $M = N$, but may hold for a wider range of M-values.

2. If

$$M \leq \min \left\{ \prod_{i=1}^{N} p_i^{-p_i}, \frac{1}{p_1^2 + \cdots + p_N^2} \right\},$$

then

$$\dim_H F = \dim_B F = -\frac{\log M}{\log r}, \qquad \mu_{\mathbf{p}} - a.s.$$

3. If $M \geq N$ and $p_i = 1/N$ for each i, we have

$$\dim_H F = \dim_B F = \dim_H C = -\frac{\log N}{\log r}, \qquad \mu_{\mathbf{p}} - a.s.$$

Thank you!

Kiitos!