

HAUSDORFF DIMENSION OF CANTOR SERIES

GODOFREDO IOMMI AND BARTLOMIEJ SKORULSKI

ABSTRACT. In 1996 Y. Kifer obtained a variational formula for the Hausdorff dimension of the set of points for which the frequencies of the digits in the Cantor series expansion is given. In this note we present a slightly different approach to this problem that allow us to solve the variational problem of Kifer's formula.

1. INTRODUCTION

Let $A = \{b_n\}_{n \in \mathbb{N}}$ be a sequence of positive integers such that $b_n \in \mathbb{N} \setminus \{1\}$. Every real number $x \in [0, 1]$ can be written as

$$x = \sum_{n=1}^{\infty} \frac{\epsilon_n(x)}{b_1 b_2 \cdots b_n},$$

where $\epsilon_n(x) \in \{0, 1, \dots, b_n - 1\}$. We write

$$x = [\epsilon_1(x)\epsilon_2(x)\dots\epsilon_n(x)\dots]_A$$

and call it the A -Cantor Series of x with respect to the base $A = \{b_n\}_{n \in \mathbb{N}}$. For every base A the Cantor series is unique except for a countable number of points. Note that if A is the sequence such that for every $n \in \mathbb{N}$ we have $b_n = b$, then the A -Cantor series corresponds to the base b expansion of $x \in [0, 1]$.

In 1996 Kifer (see [Kif1]) studied the problem of computing the size of the level sets determined by the frequency of digits in the A -Cantor series expansion. More precisely, for each $n \in \mathbb{N}$, $k \in \mathbb{N}$ and $x \in (0, 1)$ set

$$\tau_j(x, n) := \text{card}\{1 \leq k \leq n : \epsilon_k(x) = j\}.$$

Whenever there exists the limit

$$(1) \quad \tau_j(x) = \lim_{n \rightarrow \infty} \frac{\tau_j(x, n)}{n},$$

it is called the *frequency* of the number j in the A -Cantor series expansion of x . A sequence $\alpha = \{\alpha_n\}_{n \in \mathbb{N}}$ is called a *stochastic vector*, if $\sum_{j=0}^{\infty} \alpha_j = 1$ and for every $j \in \mathbb{N}$ we have $\alpha_j \geq 0$. We consider the set of points for which the frequency of the digit j is equal to α_j , that is

$$(2) \quad J_A(\alpha) = \{x \in [0, 1] : \tau_j(x) = \alpha_j \text{ for every } j \in \{0, 1, \dots\}\}.$$

Date: November 19, 2009.

G. I. would like to acknowledge the support of Proyecto Fondecyt 11070050. B. S. would like to acknowledge the support of Proyecto Fondecyt 11060538. The authors would also like to thank Mike Todd for useful conversations on this subject.

The question we are interested in is: *What is the size of these sets?*

In the case that the sequence A is constant, with $b_n = b$, this problem was studied in 1949 by Eggleston (see [Eg]). In this setting the only digits that appear (hence, that can have positive frequency) are $\{0, 1, \dots, b-1\}$. Note that, by Borel Normal Number Theorem, the Lebesgue measure of $J(\alpha)$ is positive if, and only if, $\alpha_j = 1/b$ for all $j = 0, \dots, b-1$. Therefore, in order to quantify the size of $J(\alpha)$ Eggleston considered its Hausdorff dimension, that we denote by $\dim_H(\cdot)$. In [Eg] he proved that

$$\dim_H(J(\alpha)) = \frac{\sum_{j=0}^{b-1} \alpha_j \log \alpha_j}{\log b}.$$

The general problem, when the sequence A is non-constant, is more subtle. Indeed, the frequency of the digit k depends also on the frequency of the base b_n in the sequence A . For example, let $A = (2, 2, 3, 2, 2, 3, 2, 2, 3, \dots)$. That is, the frequency of the base 3 in the sequence A is equal to $1/3$ and of the base 2 is equal to $2/3$. Since the digit $k = 2$ can only appear when $b_n = 3$ we have that for every $x \in [0, 1]$ the following bound holds $\tau_2(x) \leq 1/3$.

The first studies of Hausdorff dimension of sets of numbers defined in terms of their frequencies of digits in a Cantor expansion were done by Peyrière [Pey]. He considered sets for which the conditional frequencies of the digit i subject to the frequency of the base b_n are fixed and computed its Hausdorff dimension. Let us be more precise. Define the frequency of the base symbol k in the sequence A by

$$(3) \quad d_k = \lim_{n \rightarrow \infty} (1/n) D_k(n),$$

where

$$(4) \quad D_k(n) = \text{card}\{i \in \{1, 2, \dots, n\} : b_i = k\}.$$

Let

$$(5) \quad \pi = \left\{ P = (p_{n,j})_{\substack{j=0, \dots, n-1 \\ n=2, \dots}} : \sum_{j=0}^{b_n-1} p_{n,j} = 1, p_{n,j} \geq 0 \right\}.$$

Denote by

$$(6) \quad \tau_{k,j}(x, n) := \text{card}\{i \in \{1, 2, \dots, n\} : b_i = k \text{ and } \epsilon_i(x) = j\},$$

and define the frequency by

$$(7) \quad \tau_{k,m}(x) := \lim_{n \rightarrow \infty} \tau_{k,m}(n, x)$$

whenever the limit exists. For $P \in \pi$ consider the set

$$(8) \quad J_A(P) = \{x \in [0, 1] : \tau_{k,b_n}(x) = d_{b_n} p_{k,b_n}, n \in \mathbb{N}\}.$$

In 1977 Peyrière, see [Pey], proved that:

Theorem 1.1 (Peyrière). *Let $A = \{b_n\}_{n \in \mathbb{N}}$ be a sequence of positive integers $b_n \in \mathbb{N} \setminus \{1\}$ such that $\sum_{n=2}^{\infty} d_n = 1$. Let $P \in \pi$, then*

$$(9) \quad \dim_H(J_A(P)) = \frac{\sum_{n=2}^{\infty} d_n \sum_{j=0}^{n-1} p_{n,j} \log p_{n,j}}{\sum_{n=1}^{\infty} d_n \log n}.$$

Let us stress that the set $J_A(\alpha)$ defined by (2), where α is a stochastic vector, has a much more complicated structure than the sets $J_A(P)$ considered in (8). Not only the set $J_A(\alpha)$ contains a non-denumerable family of sets of the form $J_A(P)$. Indeed, if $x \in J_A(P)$ then

$$(10) \quad \tau_j(x) = \sum_{n=1}^{\infty} \tau_{n,j}(x) = \sum_{n=1}^{\infty} d_n p_{n,j}.$$

But, it is possible for $\tau_j(x)$ to exist whereas $\tau_{n,j}(x)$ does not exist for any $n \in \mathbb{N}$.

Kifer obtained a formula relating the Hausdorff dimension of $J_A(\alpha)$ with the Hausdorff dimension of certain sets $J_A(P)$. Let

$$\pi(\alpha) := \{P \in \pi : \alpha_j = \sum_{n=1}^{\infty} d_n p_{n,j}\}.$$

In [Kif1], Kifer proved that

Theorem 1.2 (Kifer). *Let $A = \{b_n\}_{n \in \mathbb{N}}$ be a sequence of positive integers $b_n \in \mathbb{N} \setminus \{1\}$ such that $\sum_{n=2}^{\infty} d_n = 1$. Let $\{\alpha_n\}_{n \in \mathbb{N}}$ be a stochastic vector, then*

$$(11) \quad \dim_H(J(\alpha)) = \sup_{P \in \pi(\alpha)} \dim_H(J_A(P)) = \sup_{P \in \pi(\alpha)} \frac{\sum_{n=1}^{\infty} d_n \sum_{j=0}^{n-1} p_{n,j} \log p_{n,j}}{\sum_{n=1}^{\infty} d_n \log n}.$$

This formula has however a disadvantage. Even in very simple cases it is difficult to calculate the Hausdorff dimension of $J(\alpha)$. We propose a slightly different approach that will allow us to overcome this difficulty. This approach is based in the following sequences:

Put $j_0 := \min\{j : \alpha_j \neq 0\}$ and $d_1 := 0$. Let

$$A_n := \sum_{j=j_0}^{n-1} \alpha_j - \sum_{k=j_0+1}^{n-1} d_k$$

and

$$r_n := \frac{1}{A_n} \prod_{k=j_0+1}^{n-1} \left(1 - \frac{d_k}{A_k}\right) \text{ and } t_j := \frac{\alpha_j}{\prod_{k=j_0+1}^j \left(1 - \frac{d_k}{A_k}\right)}.$$

We prove the following,

Theorem A. *Let $A = \{b_n\}_{n \in \mathbb{N}}$ be a sequence of positive integers $b_n \in \mathbb{N} \setminus \{1\}$ such that $\sum_{n=2}^{\infty} d_n = 1$ and $\alpha = \{\alpha_i\}_{i \in \mathbb{N}}$ be a stochastic vector then*

$$\dim_H(J(\alpha)) = \frac{\sum_{j=j_0}^{\infty} \alpha_j \log t_j + \sum_{i=j_0+1}^{\infty} d_i \log r_i}{\sum_{n=2}^{\infty} d_n \log n}.$$

It is direct consequence of the above Theorem that the supremum in Kifer's result (11) is attained at the level $J_A(P^\alpha)$, where

$$(12) \quad P^\alpha = (p_{n,j}^\alpha) \text{ and } p_{n,j}^\alpha = r_n t_j.$$

This is yet another example of the phenomenon described by Cajar in [Ca] where the Hausdorff dimension of a non-denumerable union of sets corresponds to the supremum of the Hausdorff dimension of each set. This property is, of course, false

in general, but in the case of sets defined in terms of the frequencies of digits holds for a large class of systems, see [Ca].

Remark 1.1. *The problem considered in this note can be addressed using techniques from random dynamical systems, but with those techniques we were not able to obtain better results than the ones presented here. The main formulas (9), (11) and (A) can be thought of as the quotient of a random entropy over a random Lyapunov exponent. Compare with the work of Kifer [Kif1, Kif2].*

2. PROOF OF THEOREM A

The following basic result from dimension theory will be used several times in the rest of note, for a detailed exposition on the subject see [Fal, Pes]. Let μ be a Borel finite measure, the *pointwise dimension* of μ at the point x is defined, whenever the limit exists, by

$$(13) \quad d_\mu(x) = \lim_{r \rightarrow 0} \frac{\log \mu(B(x, r))}{\log r},$$

where $B(x, r)$ is the ball of center x and radius r . The following result can be found in [Pes, p.42],

Proposition 2.1. *Let μ be a finite Borel measure. If $d_\mu(x) \leq d$, for every $x \in F$, then $\dim_H(F) \leq d$.*

Construction of the optimal sequence. We start with a general lemma. Let $L \in \mathbb{N} \cup \{\infty\}$ and let $d = (d_n)_{n=1}^L$ and $\alpha = (\alpha_j)_{j=0}^{L-1}$ be stochastic vectors. Put $j_0 := \min\{j : \alpha_j \neq 0\}$.

Lemma 2.1. *There exist $(r_n)_{n=j_0+1}^L$ and $(t_j)_{j=0}^{L-1}$ where $r_n, s_j \geq 0$ such that for*

$$(14) \quad p_{n,j} = r_n t_j$$

we have that

- (1) $\sum_{j=j_0}^{n-1} p_{n,j} = 1$ for every $n \geq j_0 + 1$ and
- (2) $\sum_{k=j+1}^L d_k p_{k,j} = \alpha_j$ for every $j \geq j_0$.

Proof. Let $n \in \mathbb{N}$ be such that $j_0 < n \leq L$, we define inductively the numbers A_n and $\alpha_j^{(n)}$. For every $j \geq j_0$ let

$$\alpha_j^{(j_0)} := \alpha_j \text{ and } A_{j_0} := \alpha_{j_0}.$$

If $j \geq n - 1$ then we define $\alpha_j^{(n)} := \alpha_j^{(n-1)} = \alpha_j$. If $j_0 \leq j < n - 1$ then we define

$$(15) \quad \alpha_j^{(n)} := \alpha_j^{(n-1)} \left(1 - \frac{d_{n-1}}{A_{n-1}}\right) = \alpha_j \prod_{k=j+1}^{n-1} \left(1 - \frac{d_k}{A_k}\right) = \alpha_j \frac{\prod_{k=j_0+1}^{n-1} \left(1 - \frac{d_k}{A_k}\right)}{\prod_{k=j_0+1}^j \left(1 - \frac{d_k}{A_k}\right)}$$

and

$$(16) \quad A_n := \sum_{j=j_0}^{n-1} \alpha_j^{(n)}.$$

Put

$$r_n := \frac{1}{A_n} \prod_{k=j_0+1}^{n-1} \left(1 - \frac{d_k}{A_k}\right) \text{ and } t_j := \frac{\alpha_j}{\prod_{k=j_0+1}^j \left(1 - \frac{d_k}{A_k}\right)}.$$

It follows from equation (14) and (15) that

$$(17) \quad p_{n,j} := \frac{\alpha_j^{(n)}}{A_n}.$$

Note that from equation (16) we have

$$\sum_{j=j_0}^{n-1} p_{n,j} = \sum_{j=j_0}^{n-1} \frac{\alpha_j^{(n)}}{A_n} = \frac{1}{A_n} \sum_{j=j_0}^{n-1} \alpha_j^{(n)} = \frac{A_n}{A_n} = 1.$$

This proves item (1).

Note that applying equation (17) and (15) we obtain,

$$(18) \quad \alpha_j^{(n)} = \alpha_j^{(n-1)} - d_{n-1} p_{n-1,j} = \alpha_j - \sum_{k=j+1}^{n-1} d_k p_{k,j}.$$

Also note that if $j \leq n-1$, from equation (18) and from item (1) of the Lemma (that we already proved), we have that

$$\begin{aligned} A_n &= \sum_{j=j_0}^{n-1} \alpha_j^{(n)} = \sum_{j=j_0}^{n-1} \alpha_j - \sum_{j=j_0}^{n-1} \sum_{k=j+1}^{n-1} d_k p_{k,j} = \sum_{j=0}^{n-1} \alpha_j - \sum_{k=j_0+1}^{n-1} d_k \sum_{j=j_0}^{k-1} p_{k,j} \\ &= \sum_{j=j_0}^{n-1} \alpha_j - \sum_{k=j_0+1}^{n-1} d_k = \left(1 - \sum_{k=j_0+1}^{n-1} d_k\right) - \left(1 - \sum_{j=j_0}^{n-1} \alpha_j\right) = \sum_{k=n}^L d_k - \sum_{j=n}^{L-1} \alpha_j. \end{aligned}$$

Therefore, if $L = \infty$, then

$$\lim_{n \rightarrow \infty} A_n = 0.$$

We have proved that if $L = \infty$ then the series $\lim_{n \rightarrow \infty} A_n = \sum_{j=j_0}^{\infty} \alpha_j^{(n)} = 0$. In particular, $\lim_{n \rightarrow \infty} \alpha_j^{(n)} = 0$. Then, from equation (18) we obtain that

$$(19) \quad \sum_{k=j+1}^{\infty} d_k p_{k,j} = \alpha_j.$$

On the other hand, if $L < \infty$, we obtain the following equality

$$A_L = d_L.$$

This finishes the proof. \square

Definition 2.1. Let $(p_{n,j})$ be as in Lemma 2.1 and let $P^\alpha = (p_{n,j}^\alpha)$ be defined by

$$p_{n,j}^\alpha = \begin{cases} p_{n,j} & \text{if } n \geq j_0 + 1, j \geq j_0 \\ 0 & \text{if } n \geq j_0 + 1, j < j_0 \\ 1/n & \text{if } n \leq j_0, j < j_0 \end{cases}$$

It is a direct consequence of Lemma 2.1 that $P^\alpha \in \pi(\alpha)$. That is, $J_A(P^\alpha) \subset J_A(\alpha)$.

Construction of the measure. Recall that, except for a countable number of points, $x \in [0, 1]$ can be written in a unique way in base A as

$$x = [\epsilon_1(x)\epsilon_2(x) \dots \epsilon_n(x) \dots]_A.$$

Let $n \in \mathbb{N}$ and consider the cylinder set defined by

$$C(\epsilon_1, \dots, \epsilon_n) = \{x \in [0, 1] : \epsilon_1(x) = \epsilon_1, \epsilon_2(x) = \epsilon_2, \dots, \epsilon_n(x) = \epsilon_n\}.$$

The collection of all cylinders form a semi-algebra that generates the Borel σ -algebra in $[0, 1]$. Consider the probability measure μ defined on cylinders by

$$(20) \quad \mu(C(\epsilon_1, \dots, \epsilon_n)) = \prod_{k=1}^n p_{b_k, \epsilon_k}^\alpha = \prod_{k=1}^n r_{b_k} t_{\epsilon_k}.$$

Note that,

$$\mu(C(\epsilon_1(x), \dots, \epsilon_n(x))) = \left(\prod_{j=0}^{\infty} t_j^{\tau_j(x, n)} \right) \left(\prod_{i=2}^{\infty} r_i^{D_i(n)} \right).$$

Therefore, for every $x \in J_A(\alpha)$ we have that

$$(21) \quad \begin{aligned} d_\mu(x) &= \lim_{r \rightarrow 0} \frac{\log \mu(B(x, r))}{\log r} = \lim_{n \rightarrow \infty} \frac{\log \mu(C(\epsilon_1(x), \dots, \epsilon_n(x)))}{\log \prod_{i=1}^m b_i} \\ &= \frac{\lim_{n \rightarrow \infty} \frac{1}{n} \log \mu(C(\epsilon_1(x), \dots, \epsilon_n(x)))}{\sum_{m=2}^{\infty} d_m \log m} \\ &= \frac{\lim_{n \rightarrow \infty} \frac{1}{n} \log \left(\prod_{j=0}^{n-1} t_j^{\tau_j(n)} \prod_{i=2}^n r_i^{D_i(n)} \right)}{\sum_{m=2}^{\infty} d_m \log m} \\ &= \frac{\lim_{n \rightarrow \infty} \left(\sum_{j=0}^{n-1} \frac{\tau_j(n)}{n} \log t_j + \sum_{i=2}^n \frac{D_i(n)}{n} \log r_i \right)}{\sum_{m=2}^{\infty} d_m \log m} \\ &= \frac{\sum_{j=j_0}^{\infty} \alpha_j \log t_j + \sum_{i=j_0+1}^{\infty} d_i \log r_i}{\sum_{m=2}^{\infty} d_m \log m}. \end{aligned}$$

In virtue of Proposition 2.1 we obtain that

$$(22) \quad \dim_H(J_A(\alpha)) \leq \frac{\sum_{j=j_0}^{\infty} \alpha_j \log t_j + \sum_{i=j_0+1}^{\infty} d_i \log r_i}{\sum_{m=2}^{\infty} d_m \log m}.$$

In order to obtain the lower bound note that since $J_A(P^\alpha) \subset J_A(\alpha)$ we have that $\dim_H(J_A(P^\alpha)) \leq \dim_H(J_A(\alpha))$. Proceeding as in (21) and making use of (20) we obtain that for every $x \in J_A(P^\alpha)$, the pointwise dimension is given by

$$d_\mu(x) = \frac{\sum_{n=1}^{\infty} d_n \sum_{i=0}^{n-1} p_{i,n}^\alpha \log p_{i,n}^\alpha}{\sum_{n=2}^{\infty} d_n \log n}.$$

Since $J_A(P^\alpha) \subset J_A(\alpha)$, we have that

$$\frac{\sum_{n=1}^{\infty} d_n \sum_{i=0}^{n-1} p_{i,n}^\alpha \log p_{i,n}^\alpha}{\sum_{n=2}^{\infty} d_n \log n} = \frac{\sum_{j=j_0}^{\infty} \alpha_j \log t_j + \sum_{i=j_0+1}^{\infty} d_i \log r_i}{\sum_{m=2}^{\infty} d_m \log m}$$

and then, by (9), we obtain

$$\dim_H(J_A(\alpha)) \geq \frac{\sum_{j=j_0}^{\infty} \alpha_j \log t_j + \sum_{i=j_0+1}^{\infty} d_i \log r_i}{\sum_{m=2}^{\infty} d_m \log m}.$$

This finishes the proof of the theorem.

REFERENCES

- [Ca] Cajar, Helmut *Billingsley dimension in probability spaces*. Lecture Notes in Mathematics, 892. Springer-Verlag, Berlin-New York, 1981. i+106 pp.
- [Eg] Eggleston, H.: *The fractional dimension of a set defined by decimal properties* Quat. J. Math. Oxford Ser. 20 31-36 (1949).
- [Fal] Falconer, Kenneth *Fractal geometry. Mathematical foundations and applications*. Second edition. John Wiley and Sons, Inc., Hoboken, NJ, (2003.) xxviii+337 pp
- [Kif1] Y. Kifer *Fractal dimensions and random transformations* Trans. Amer. Math. Soc. 348 (1996), no. 5, 2003-2038.
- [Kif2] Y. Kifer *Fractals via random iterated function systems and random geometric constructions*. Fractal geometry and stochastics (Finsterbergen, 1994), 145–164, Progr. Probab., 37, Birkhuser, Basel, 1995.
- [Pey] J. Peyrière *Calculs de dimensions de Hausdorff* Duke Math. J. 44 (1977), no. 3, 591–601.
- [Pes] Y. Pesin *Dimension Theory in Dynamical Systems* CUP (1997).

FACULTAD DE MATEMÁTICAS, PONTIFICIA UNIVERSIDAD CATÓLICA DE CHILE (PUC), AVENIDA VICUÑA MACKENNA 4860, SANTIAGO, CHILE

E-mail address: `giommi@mat.puc.cl`

URL: `http://www.mat.puc.cl/~giommi/`

UNIVERSIDAD CATÓLICA DEL NORTE, FACULTAD DE CIENCIAS, DEPARTAMENTO DE MATEMÁTICAS, AVENIDA ANGAMOS 0610, ANTOFAGASTA, CHILE

E-mail address: `bskorulski@ucn.cl`