RANDOM SERIES IN POWERS OF ALGEBRAIC INTEGERS: HAUSDORFF DIMENSION OF THE LIMIT DISTRIBUTION

by

Steven P. Lalley
Purdue University

Technical Report #95-15

Department of Statistics
Purdue University

April 1995

RANDOM SERIES IN POWERS OF ALGEBRAIC INTEGERS: HAUSDORFF DIMENSION OF THE LIMIT DISTRIBUTION

STEVEN P. LALLEY
PURDUE UNIVERSITY

ABSTRACT. We study the distributions $F_{\theta,p}$ of the random sums $\sum_1^\infty \varepsilon_n\theta^n$, where $\varepsilon_1, \varepsilon_2, \ldots$ are i.i.d. Bernoulli-p and θ is the inverse of a Pisot number (an algebraic integer β whose conjugates all have moduli less than 1) between 1 and 2. It is known that, when $p = .5$, $F_{\theta,p}$ is a singular measure with exact Hausdorff dimension less than 1. We show that in all cases the Hausdorff dimension can be expressed as the top Lyapunov exponent of a sequence of random matrices, and provide an algorithm for the construction of these matrices. We show that for certain β of small degree, simulation gives the Hausdorff dimension to several decimal places.

1. INTRODUCTION

Let $\epsilon_1, \epsilon_2, \ldots$ be independent, identically distributed Bernoulli-p random variables, and consider the random variable X defined by the series

$$X = \sum_{n=1}^{\infty} \epsilon_n \theta^n$$

where $\theta \in (0,1)$. By the "Law of Pure Types" the distribution $F = F_{\theta,p}$ of X is either absolutely continuous or purely singular (but of course nonatomic). Erdös proved (a) that there exist values of θ larger than $\frac{1}{2}$, e.g., the inverse of the "golden ratio", such that F is singular [5]; but (b) that there exists $\gamma < 1$ such that for *almost every* $\theta \in (\gamma, 1)$, $F_{\theta,\frac{1}{2}}$ is absolutely continuous [6]. [1] Erdös' argument shows that in fact F is singular whenever θ is the inverse of a "Pisot" number. Recall that a *Pisot number* is an algebraic integer greater than 1 whose algebraic conjugates are all smaller than 1 in modulus (see [18]) and an algebraic integer is a root of an irreducible monic, integer polynomial. There are in fact infinitely many Pisot numbers in the interval $(1,2)$: for example, for each $n = 2, 3, \ldots$ the largest root β_n of the polynomial

$$p_n(x) = x^n - x^{n-1} - x^{n-2} - \cdots - x - 1 \tag{1}$$

is a Pisot number: see [19]. We will call these the *simple* Pisot numbers.

One may ask for those values of θ such that $F = F_{\theta,p}$ is singular: is it necessarily the case that F is concentrated on a set of Hausdorff dimension strictly less than 1, and if so, what

Date. April 24, 1995.
Key words and phrases. Hausdorff dimension, entropy, Lyapunov exponent, Pisot number.
Supported by NSF grant DMS-9307855.

[1] Solomyak [20] has recently shown that statement (b) is true for $\gamma = .5$.

is the minimal such dimension? (we shall call this minimum the *Hausdorff dimension* of F). Przytycki and Urbanski [16], enlarging on an argument of Garsia [11], showed that if θ is the inverse of a Pisot number between 1 and 2 then in fact $F_{\theta,.5}$ has Hausdorff dimension less than 1. Unfortunately, their method does not give an effective means of calculating it. Recently, Alexander and Zagier [2] showed how to calculate the "information dimension" in the case where θ =1/golden ratio and $p = \frac{1}{2}$. As it turns out, the Hausdorff and information dimensions coincide for the measures we consider here, but Alexander and Zagier did not prove this. Their argument is very elegant, relating the dimension to properties of the "Fibonacci tree" and thence to the Euclidean algorithm, but it does not seem to generalize easily. (Although the authors claim that "it seems likely that the particulars will extend to β_n^{-1}" for β_n as defined above, this is not readily apparent.)

The purpose of this paper is to characterize the Hausdorff dimension of $F_{\theta,p}$ as the top Lyapunov exponent of a certain natural sequence of random matrix products. This characterization is quite general: it is valid for every θ such that $\beta = 1/\theta \in (1,2)$ is a Pisot number, and for all values of the Bernoulli parameter p. (The method is applicable even more generally when the process $\epsilon_1, \epsilon_2, \ldots$ is a $k-$step Markov chain taking values in an arbitrary finite set of integers, but we carry out the details only in the case of Bernoulli processes.) Moreover, although the matrices involved may have large dimensions (typically increasing exponentially in m, where m is the degree of the minimal polynomial), they are sparse, so numerical computation of the Lyapunov exponent masy be feasible for $1/\theta$ of moderate degree. Small simulations give "nonrigorous" estimates that agree with the estimate of Alexander and Zagier (which is accurate to the 4th decimal place) to 3 decimal places, and for values of p ranging from .1 to .5 give estimates with accuracy to $\pm.005$. Some of these numerical results are reported in section 8 below.

Acknowledgments. Thanks to IRENE HUETER for help with some of the numerical computations and valuable conversations. Thanks to YUVAL PERES for the reference to [2], and to BORIS SOLOMYAK for pointing the author to [9] after seeing an earlier version of this manuscript.

2. DIMENSION AND ENTROPY: INFINITE BERNOULLI CONVOLUTIONS

For any probability distribution F on a metric space, one may define its (Hausdorff) *dimension* $\delta(F)$ to be the infimum of all $d > 0$ such that F is supported by a set of Hausdorff dimension d. There is a simple and well known tool for calculating $\delta(F)$, which we shall call *Frostman's Lemma.*

Lemma 1. *(Frostman) If for $F-$almost every x*

$$\delta_1 \le \liminf_{r \downarrow 0} \frac{\log F(B_r(x))}{\log r} \le \delta_2$$

then $\delta_1 \le \delta(F) \le \delta_2$.

Here $B_r(x)$ denotes the ball of radius r centered at x. The proof is relatively easy: see [7], ch.1, problem 1.8 , or [22].

Application of the Frostman Lemma to a probability measure on the real line requires a handle on the probabilities of "typical" small intervals. In our applications F is the distribution of a random sum $X = \sum \varepsilon_n \theta^n$, and the value of X is (roughly) determined to

within r by the sum of the first n terms of the series, where $r \approx \theta^n$. Henceforth, for any infinite sequence $\varepsilon = \varepsilon_1\varepsilon_2\ldots$ of 0s and 1s we will write

$$x(\varepsilon) = \sum_{k=1}^{\infty} \varepsilon_k\theta^k \text{ and } x_n(\varepsilon) = \sum_{k=1}^{n} \varepsilon_k\theta^k,$$

and for any finite sequence $\varepsilon = \varepsilon_1\varepsilon_2\ldots\varepsilon_n$ of length $n \geq 1$ we will write $x_n(\varepsilon) = \sum_{k=1}^{n} \varepsilon_k\theta^k$. Observe that for any $n \geq 1$ and every infinite 0-1 sequence ε, $x(\varepsilon) - x_n(\varepsilon) \leq \theta^{n+1}/(1-\theta)$. Consequently, if $x_n(\varepsilon') = x_n(\varepsilon)$ then $|x(\varepsilon') - x(\varepsilon)| \leq \theta^{n+1}/(1-\theta)$. In general, the converse need not be true; however, if θ is the inverse of a Pisot number then there is a weak converse, discovered by Garsia [10]. Because this result is of central importance in the ensuing arguments, and because the proof is rather short, we include it here. For the remainder of the paper, the following assumption will be in force:

Assumption 1. *$\theta = \beta^{-1}$ for a Pisot number $\beta \in (1,2)$ with minimal integer monic polynomial $p(x)$ of degree m.*

Lemma 2. [10] *(Garsia) There exists a constant $C > 0$ (depending on β) such that for any integer $n \geq 1$ and any two sequences $\varepsilon = \epsilon_1, \epsilon_2, \ldots, \epsilon_n$ and $\varepsilon' = \epsilon_1', \epsilon_2', \ldots, \epsilon_n'$ of zeroes and ones, either*

$$x_n(\varepsilon) = x_n(\varepsilon')$$

or

$$|x_n(\varepsilon') - x_n(\varepsilon)| \geq C\theta^n.$$

Proof. Suppose that $\varepsilon, \varepsilon'$ are 0-1 sequences such that $\sum_{k=1}^{n} \varepsilon_k\theta^k \neq \sum_{k=1}^{n} \varepsilon_k'\theta^k$. Then $F(\beta) \neq 0$, where $F(x) = \sum_{k=1}^{n} \delta_k x^{n-k}$ and $\delta_k = \varepsilon_k - \varepsilon_k'$. Note that F is a polynomial with coefficients $0, \pm 1$. Let β_i be the algebraic conjugates of β; then for each i, $F(\beta_i) \neq 0$, and

$$F(\beta)\prod_i F(\beta_i) \in \mathbf{Z}.$$

Since each $|\beta_i| < 1$ (β is a Pisot number) and the coefficients of F are bounded by 1 in modulus, $|F(\beta_i)| \leq 1/(1-|\beta_i|)$. Consequently,

$$|F(\beta)| \geq \prod_i (1-|\beta_i|) = C.$$

□

It follows from this lemma that if $x(\varepsilon') \in B_r(x(\varepsilon))$ for $r = \theta^{n+1}/2(1-\theta)$ then there are at most $2C+1$ possible values for $x_n(\varepsilon')$. There are, in general, *many* pairs of sequences for which $x_n(\varepsilon) = x_n(\varepsilon')$: the separation bound in the lemma implies that there are only $O(\theta^{-n})$ possible values of the sum, but there are 2^n sequences of zeroes and ones of length n (and $\theta^{-1} < 2$.) For each n the possible values of $x_n(\varepsilon)$ partition the space Σ of all infinite sequences of 0s and 1s: for any two sequences $\varepsilon, \varepsilon' \in \Sigma$, ε and ε' are in the same element of the partition iff $x_n(\varepsilon') = x_n(\varepsilon)$. Call the resulting partition $\mathcal{P}_n$. For any $\varepsilon \in \Sigma$ let $U(\mathcal{P}_n, \varepsilon)$ be the element of the partition $\mathcal{P}_n$ containing ε. Then for each $\varepsilon \in \Sigma$ and each value of the Bernoulli parameter p the set $U(\mathcal{P}_n, \epsilon)$ has a probability $\pi_n(\epsilon)$ (for notational simplicity, the dependence on p is suppressed). In the subsequent sections we will prove the following

Theorem 1. *For any $p \in (0,1)$, if $\epsilon_1, \epsilon_2, \ldots$ are iid Bernoulli-p random variables then with probability 1*

$$\alpha = \lim_{n\to\infty} (\pi_n(\epsilon))^{\frac{1}{n}} \tag{2}$$

exists, is positive, and is constant.

The proof will exhibit the limit $\alpha = \alpha(p)$ as the top Lyapunov exponent of a certain sequence of random matrix products, providing an effective means of calculation. Note that when $p = \frac{1}{2}$, the probability $\pi_n(\alpha)$ is just 2^{-n} times the cardinality of the equivalence class. In this case the matrices may be chosen to have all entries 0 or 1, so numerical computations are easiest in this case.

Using the close connection between the partitions $\mathcal{P}_n$ and the neighborhood system $B_r(x(\varepsilon))$, $r = \kappa\theta^n$, we will prove the following formula for the dimension of the measure $F_{\theta,p}$.

Theorem 2. *The Hausdorff dimension of $F_{\theta,p}$ is*

$$\delta = \frac{\log \alpha}{\log \theta}. \tag{3}$$

This formula is an instance of the by now well known general principle "dimension × expansion rate = entropy": see [22] for more. However, the proof is not entirely trivial, even given the result of Theorem 1: it relies on Garsia's lemma, and therefore on the algebraic nature of the ratio $\beta = 1/\theta$. In the special case $p = \frac{1}{2}$, Przytycki and Urbanski proved the inequality $\delta \le \log \alpha / \log \theta$ and used it to deduce that $\delta < 1$, but did not establish equality. Alexander and Yorke proved, again in the special case $p = \frac{1}{2}$, that $\log \alpha / \log \theta$ equals the "information dimension" (also called the Renyi dimension) of F, which always dominates the Hausdorff dimension, but did not prove equality with the Hausdorff dimension. Theorem 2 follows directly from Theorem 1 and Propositions 3-4 below.

It is worth noting here that the dimension $\delta(F_{\theta,p})$, considered as a function of p, is symmetric about 1/2, i.e.,

$$\delta(F_{\theta,p}) = \delta(F_{\theta,1-p}).$$

The proof is simple: If $\varepsilon_1, \varepsilon_2, \ldots$ are i.i.d. Bernoulli (p), then $\varepsilon_1', \varepsilon_2', \ldots$ are i.i.d. Bernoulli $(1-p)$, where $\varepsilon_j' = 1 - \varepsilon_j$. Consequently, if $X = x(\varepsilon)$ has distribution $F_{\theta,p}$ then $Y = \theta/(1-\theta) - X$ has distribution $F_{\theta,1-p}$. But the distributions of X and Y clearly have the same dimensions, because Y is obtained from X by an isometry of the real line.

Proposition 1. *If $F_{\theta,p}$ is singular with respect to Lebesgue measure, then $\delta < 1$.*

In fact this holds in even greater generality – see Proposition 5 below.

Proposition 2. *If β is a Pisot number between 1 and 2, then $F_{\theta,p}$ is singular for every $p \in (0,1)$.*

Proof. This follows by Erdös' original argument. The Fourier transform of $F_{\theta,p}$ is easily computed as an infinite product:

$$\hat{F}_{\theta,p}(t) = \int e^{itx} F_{\theta,p}(dx) = \prod_{k=1}^{\infty} (1 + p(\exp\{it\theta^k\} - 1)).$$

It is obvious from this that for every $t \in (-\infty, \infty)$ and every $n \ge 0$,

$$\hat{F}_{\theta,p}(\beta^n t) = \prod_{k=1}^{n} (1 + p(\exp\{it\beta^{n-k}\} - 1)) \hat{F}_{\theta,p}(t).$$

Since β is a Pisot number, distance$(\beta^n, \mathbf{Z}) \to 0$ as $n \to \infty$ at an exponential rate (see, e.g., [18], ch. 1). Consequently, for any t such that $\hat{F}_{\theta,p}(t) \neq 0$ and such that $\exp\{it\beta^n\} \neq 1$ for all $n \geq 0$,

$$\lim_{n\to\infty} \hat{F}_{\theta,p}(\beta^n t) = \prod_{k=0}^{\infty} (1 + p(\exp\{it\beta^k\} - 1))\hat{F}_{\theta,p}(t) \neq 0.$$

Therefore, by the Riemann-Lebesgue lemma, $F_{\theta,p}$ is singular. □

Theorems 1-2 reduce the problem of computing the Hausdorff dimension of $F_{\theta,p}$ to that of computing the "entropy" $\log \alpha$. This will be carried out in sections 4-8. It should be noted that the entropy arises in connection with several other dimensional quantities, such as the "information dimension": see [1] and [2] for more on this. Before coming to grips with the entropy, however, we will prove Theorem 2 and discuss certain generalizations of the results of this section to a class of measures $F_{\theta,\mu}$ including the $F_{\theta,p}$.

3. Dimension and Entropy: Stationary Measures

Let μ be an ergodic, shift-invariant measure on the space Σ of infinite sequences $\varepsilon_1, \varepsilon_2, \ldots$ of 0s and 1s. Define $F_{\theta,\mu}$ to be the distribution under μ of $X = \sum_{n=1}^{\infty} \varepsilon_n \theta^n$. Observe that if μ is the Bernoulli-p product measure, then $F_{\theta,\mu} = F_{\theta,p}$. The Bernoulli measures are not the only measures of interest, however—for instance, there are measures μ for which $F_{\theta,\mu}$ is absolutely continuous relative to Lebesgue measure. These measures were discovered by Renyi [17] for the case $\theta = 1/\text{golden ratio}$ and by Parry [13], [14] for the other Pisot numbers.

As in the previous section, for each n let $\mathcal{P}_n$ be the partition of Σ induced by the equivalence relation $\varepsilon \sim \varepsilon'$ iff $x_n(\varepsilon) = x_n(\varepsilon')$. Note that these partitions are *not* nested. For any sequence ε let $U_n(\varepsilon)$ be the element of $\mathcal{P}_n$ that contains ε, and let $\pi_n(\varepsilon) = \mu(U_n(\varepsilon))$. Similarly, for an arbitrary measurable partition $\mathcal{P}$ let $U(\mathcal{P}, \varepsilon)$ be the element of $\mathcal{P}$ that contains ε, and let $\pi(\mathcal{P}, \varepsilon) = \mu(U(\mathcal{P}, \varepsilon))$. Define the *entropy* of the partition $\mathcal{P}$ by

$$H(\mathcal{P}) = -E_\mu \log \pi(\mathcal{P}, \varepsilon) = \sum_{F\in\mathcal{P}} -\mu(F) \log \mu(F).$$

Lemma 3. $\lim_{n\to\infty} H(\mathcal{P}_n)/n = -\log \alpha$ *exists.*

Proof. For any integers $n, m \geq 1$, the partition $\mathcal{P}_n \vee \sigma^{-n}\mathcal{P}_m$ is a refinement of $\mathcal{P}_{n+m}$ (here σ is the shift), because the values of $x_n(\varepsilon)$ and $x_m(\sigma^n\varepsilon)$ determine the value of $x_{n+m}(\varepsilon)$. Hence, as a refinement, it has the larger entropy. But by an elementary property of the entropy function,

$$H(\mathcal{P}_n \vee \sigma^{-n}\mathcal{P}_m) \leq H(\mathcal{P}_n) + H(\mathcal{P}_m)$$

(see [21], sec. 4.3). Thus, $H(\mathcal{P}_n)$ is a subadditive sequence. □

Lemma 4. *For every* $n \geq 1$, $\liminf_{k\to\infty} \pi_{nk}(\varepsilon)^{\frac{1}{k}} \geq e^{-H(\mathcal{P}_n)}$ *a.e.* μ.

Proof. First, recall that the entropy $h(\sigma^n, \mathcal{P}_n)$ of the partition $\mathcal{P}_n$ with respect to the measure-preserving transformation σ^n is defined by

$$h(\sigma^n, \mathcal{P}_n) = \lim_{k\to\infty} \frac{1}{k} H\left(\vee_{i=0}^{k-1} \sigma^{-ni}\mathcal{P}_n\right) \leq H(\mathcal{P}_n);$$

see [21], ch. 4. for details. By the Shannon-MacMillan-Breiman theorem ([15],),

$$\lim_{k\to\infty} \frac{1}{k} \log \pi(\vee_{i=0}^{k-1} \sigma^{-ni}\mathcal{P}_n, \varepsilon) = h(\sigma^n, \mathcal{P}_n)$$

a.e. (μ). But $\vee_{i=0}^{k-1}\sigma^{-ni}\mathcal{P}_n$ is a refinement of $\mathcal{P}_{nk}$, so

$$\pi_{nk}(\varepsilon) \geq \pi(\vee_{i=0}^{k-1}\sigma^{-ni}\mathcal{P}_n, \varepsilon),$$

proving the lemma. □

Conjecture 1. *For any ergodic, shift-invariant probability measure μ on Σ,*

$$\lim \pi_n(\varepsilon)^{\frac{1}{n}} = \alpha \text{ almost surely}(\mu). \tag{4}$$

For the special case in which μ is the product Bernoulli-p measure, this is Theorem 1, and will be proved in sections 4-8 below. A modification of the argument (which we shall omit) shows that the conjecture is also true for any μ making the coordinate process $\varepsilon_1, \varepsilon_2, \ldots$ a stationary $k-$step Markov chain.

Proposition 3. *For each ergodic, shift-invariant probability measure μ on Σ, the dimension δ of $F_{\theta,\mu}$ satisfies*

$$\delta \leq \frac{\log \alpha}{\log \theta}.$$

Proof. By the Frostman lemma, it suffices to show that for x in a set of $F_{\theta,\mu}-$measure 1,

$$\liminf_{r\to 0} \frac{\log F_{\theta,\mu}(B_r(x))}{\log r} \leq \frac{\log \alpha}{\log \theta}, \tag{5}$$

and, by a routine argument, it suffices to consider only $r = \kappa\theta^{nk}$ for some fixed constant $\kappa > 0$, a fixed integer n, and $k = 1, 2, \cdots$. If $\varepsilon, \varepsilon'$ are in the same element of the partition $\mathcal{P}_{nk}$, then $|x(\varepsilon) - x(\varepsilon')| \leq \theta^{nk+1}/(1-\theta)$; consequently, if $\kappa = 2\theta/(1-\theta)$ then for every ε' in $U(\mathcal{P}_{nk}, \varepsilon)$, $x(\varepsilon') \in B_r(x(\varepsilon))$. Therefore, $F_{\theta,\mu}(B_r(x(\varepsilon))) \geq \pi_{nk}(\varepsilon)$, and so (5) follows from the two preceding lemmas. □

Proposition 4. *For any ergodic, shift-invariant measure μ on Σ for which (4) holds almost surely with respect to μ, the Hausdorff dimension of $F_{\theta,\mu}$ satisfies*

$$\delta = \frac{\log \alpha}{\log \theta}.$$

Proof. The inequality $\delta \leq \log\alpha/\log\theta$ has been proved in Proposition 3 above. Thus, it is enough to prove the reverse inequality. By Frostman's lemma, it suffices to show that for all x in a set of full $F_{\theta,\mu}-$measure,

$$\liminf_{r\to 0} \frac{\log F_{\theta,\mu}(B_r(x))}{\log r} \geq \frac{\log \alpha}{\log \theta}. \tag{6}$$

By a routine argument, it suffices to prove this for the sequence $r = \theta^n$.

Let $\varepsilon, \varepsilon'$ be arbitrary sequences of 0s and 1s. In order that $x(\varepsilon') \in B_r(x(\varepsilon))$ for $r = \theta^n$, it is necessary that $|x_n(\varepsilon') - x_n(\varepsilon)| \leq \kappa\theta^n$, where $\kappa = 1 + 2/(1-\theta)$. Consequently, for any fixed sequence $\varepsilon = \varepsilon_1\varepsilon_2\ldots$ of 0s and 1s and for any $r = \theta^n$,

$$F_{\theta,\mu}(B_r(x(\varepsilon))) \leq \rho_n(\varepsilon),$$

where

$$\rho_n(\varepsilon) = \mu\{\varepsilon' : |x_n(\varepsilon') - x_n(\varepsilon)| \leq \kappa\theta^n\}.$$

Here ε is fixed (nonrandom). Note that $\rho_n(\varepsilon) \geq \pi_n(\varepsilon)$.

Now let $\varepsilon = \varepsilon_1\varepsilon_2\cdots \in \Sigma$ be random, with distribution μ. By the Chebyshev inequality, for each $\eta > 0$ and each $n \geq 1$,

$$\mu\{\rho_n(\varepsilon) \geq (1+\eta)^n \pi_n(\varepsilon)\} \leq (1+\eta)^{-n} E_\mu\left(\frac{\rho_n(\varepsilon)}{\pi_n(\varepsilon)}\right).$$

We will argue below that there is a constant $C_* < \infty$ such that for all n, $E_\mu(\rho_n(\varepsilon)/\pi_n(\varepsilon)) < C_*$. It will then follow that for every $\eta > 0$, $\sum_n \mu\{\rho_n(\varepsilon) \geq (1+\eta)^n\pi_n(\varepsilon)\} < \infty$, and consequently, by the Borel-Cantelli lemma, that with $\mu-$probability 1, $\rho_n(\varepsilon) \geq (1+\eta)^n\pi_n(\varepsilon)$ for at most finitely many n. But this will imply that, with probability 1,

$$\lim_{n\to\infty} \rho_n(\varepsilon)^{1/n} = \lim_{n\to\infty} \pi_n(\varepsilon)^{1/n} = \alpha.$$

Since $\rho_n(\varepsilon)$ is an upper bound for $F_{\theta,\mu}(B_r(x(\varepsilon)))$, $r = \theta^n$, this will prove (6).

So consider $E_\mu(\rho_n(\varepsilon)/\pi_n(\varepsilon))$. By definition of $\rho_n(\varepsilon)$,

$$E_\mu\left(\frac{\rho_n(\varepsilon)}{\pi_n(\varepsilon)}\right) = \sum_{x_n(\varepsilon)} \sum_{x_n(\varepsilon')} \pi_n(\varepsilon'),$$

where the outer sum is over all possible values of $x_n(\varepsilon)$ and the inner sum is over those values of $x_n(\varepsilon')$ such that $|x_n(\varepsilon') - x_n(\varepsilon)| \leq \kappa\theta^n$ (only one representative sequence ε' is taken for each such value). But by Garsia's lemma, each value of $x_n(\varepsilon')$ appears in the inner sum for *at most* C_* distinct $x_n(\varepsilon)$, for some $C_* < \infty$ independent of n. Therefore,

$$E_\mu\left(\frac{\rho_n(\varepsilon)}{\pi_n(\varepsilon)}\right) \leq C_* \sum_{x_n(\varepsilon)} \pi_n(\varepsilon) = C_*.$$

□

Proposition 5. *Let μ be an ergodic, shift-invariant probability measure on Σ. If $F_{\theta,\mu}$ is singular with respect to Lebesgue measure, then $\delta < 1$.*

Proof. This is an adaptation of the arguments of [11] and [16]. By Proposition 3, it suffices to show that $-\log\alpha < -\log\theta$. For this it suffices to show that for some $n \geq 1$,

$$H(\mathcal{P}_n)/n < -\log\theta,$$

because (see Lemma 3 and its proof) $H(\mathcal{P}_n)$ is a subadditive sequence such that $H(\mathcal{P}_n)/n$ converges to $-\log\alpha$ as $n \to \infty$.

Recall that the elements of the partition $\mathcal{P}_n$ are in 1-1 correspondence with the possible values of the sum $x_n(\varepsilon) = \sum_{j=1}^n \varepsilon_j\theta^j$, where $\varepsilon_1\varepsilon_2\ldots\varepsilon_n$ is a 0-1 sequence. Hence, by Garsia's lemma, there are at most $C'\theta^{-n}$ elements of $\mathcal{P}_n$, for some constant $C' < \infty$ independent of n. Now the hypothesis that $F_{\theta,\mu}$ is singular, together with Garsia's lemma, implies (by the same argument as in [16]) that the probability measure $\mu|\mathcal{P}_n$ is highly concentrated on a subset of $\mathcal{P}_n$ of much smaller than the cardinality of $\mathcal{P}_n$: in particular, for every $\eta > 0$ there exists an n and a subset $\mathcal{Q}_n \subset \mathcal{P}_n$ such that

$$\#\mathcal{Q}_n/\#\mathcal{P}_n < \eta$$

and

$$\mu(\cup_{F\in\mathcal{Q}_n} F) = 1 - \rho > 1 - \eta.$$

Consequently, with $\overline{\rho} = 1 - \rho$,

$$\begin{aligned} H(\mathcal{P}_n) &= -\sum_{F \in \mathcal{Q}_n} \mu(F) \log \mu(F) - \sum_{F \in \mathcal{P}_n - \mathcal{Q}_n} \mu(F) \log \mu(F) \\ &\le \overline{\rho} \log(\#\mathcal{Q}_n) - \overline{\rho} \log \overline{\rho} + \rho \log(\#(\mathcal{P}_n - \mathcal{Q}_n)) - \rho \log \rho \\ &\le \log \theta^{-n} + \overline{\rho} \log \eta + \log C' - \overline{\rho} \log \overline{\rho} - \rho \log \rho. \end{aligned}$$

Here we have used the estimates $\#\mathcal{P}_n \le C'\theta^{-n}$ and $\#\mathcal{Q}_n \le C'\eta\theta^{-n}$, and also the fact that for a given partition, the probability measure that maximizes entropy is the uniform distribution on the partition. Now $\rho < \eta$ and $\eta > 0$ may be chosen arbitrarily small. Since $x \log x + (1-x)\log(1-x) \to 0$ as $x \to 0$, the last two terms of the upper bound may be made arbitrarily small. The term $\log C'$ is independent of η and ρ. Hence, by choosing $0 < \rho < \eta$ very small, we may make $\overline{\rho} \log \eta + \log C'$ much less than 0. It follows that for sufficiently large n, $H(\mathcal{P}_n) < -n \log \theta$. □

4. Equivalent Sequences

In section 2 we reduced the problem of computing the Hausdorff dimension of the measure $F_{\theta,p}$ to that of computing the entropy α, which is essentially the same as problem of estimating probabilities of equivalence classes. Thus, we must find an effective way to tell when two sequences are equivalent. Recall that for fixed θ the equivalence relation is defined as follows: two sequences ϵ and ϵ' of zeroes and ones of length n are equivalent iff $x_n(\varepsilon) = x_n(\varepsilon')$. More generally, say that two length-n sequences $\varepsilon, \varepsilon'$ of *integers* are equivalent iff $\sum_{k=1}^n \epsilon_k \theta^k = \sum_{k=1}^n \epsilon'_k \theta^k$.

Let $p(z)$ be the minimal polynomial of $\beta = 1/\theta$, and assume that it has degree m and leading coefficient 1. The relation $p(\beta) = 0$ may be rewritten by moving all terms with negative coefficients to the other side, yielding an identity between two polynomials in β of degree m with nonnegative integer coefficients. This equation translates to an equivalence between two distinct sequences ϵ and ϵ' of nonnegative integers of length $m+1$: we will call this equivalence the *fundamental relation*. Thus, for example, if β is the golden ratio, the fundamental relation is

$$100 \sim 011,$$

reflecting the fact that the minimal polynomial of the golden ratio is $x^2 - x - 1$. Note that there are Pisot numbers β between 1 and 2 such that the sequences in the fundamental relation have entries other than 0 and 1: for instance, the leading root $\beta = 1.755\cdots$ of the cubic $p(x) = x^3 - 2x^2 + x + 1$ is a Pisot number with fundamental relation $1011 \sim 0200$. But observe that if $\varepsilon \sim \varepsilon'$ is the fundamental relation, then the first entry of ε is always 1 and the first entry of ε' is 0.

Let $\gamma = \gamma_1\gamma_2\ldots\gamma_n$ and $\gamma' = \gamma'_1\gamma'_2\ldots\gamma'_n$ be arbitrary sequences of integers, and let $\varepsilon \sim \varepsilon'$ be the fundamental relation. We will say that γ' can be obtained from γ by applying the fundamental relation k times in the $(m+1)-$block starting at the lth entry if $\gamma_j = \gamma'_j$ for all j except $j = l, l+1, \ldots, l+m$, and $\gamma_{l+j} + k\varepsilon'_j = \gamma'_{l+j} + k\varepsilon_j$ for all $j = 0, 1, \ldots, m$. Here $\varepsilon = \varepsilon_0\varepsilon_1\ldots\varepsilon_m$ and $\varepsilon' = \varepsilon'_0\varepsilon'_1\ldots\varepsilon'_m$, and k may be any integer. In general, if a 0-1 sequence ϵ of arbitrary length $n \ge m+1$ may be obtained from another sequence ϵ' of the same length by repeatedly applying the fundamental relation to various $(m+1)-$blocks (strings of $m+1$ consecutive entries), then $\epsilon \sim \epsilon'$. For example, when β is the golden ratio the sequences 100011 and 011100 are equivalent, because one may obtain the second from the first by the chain of substitutions $100011 \to 100100 \to 011100$.

Proposition 6. *Two sequences ϵ and ϵ' of length n are equivalent iff ϵ' may be obtained from ϵ by applying the fundamental relation repeatedly to blocks of length $m+1$, starting at the left and ending at the right.*

The stipulation that the substitutions be made left to right will be crucial. However, it should be noted that in general it is not possible to make the substitutions in order *and* have all of the intermediate sequences be sequences of zeroes and ones, even when the sequences $\varepsilon, \varepsilon'$ in the fundamental relation $\varepsilon \sim \varepsilon'$ are 0-1 sequences. For example, when the fundamental relation is $100 \sim 011$ the sequences 10100 and 01111 are equivalent, the chain of substitutions being

$$10100 \to 01200 \to 01111.$$

Moreover, the proposition does not state that the fundamental relation is applied only once at each block of length $m+1$: in fact, it may be that one must use it more than once in a given direction, e.g, changing 200 to 0(-1)(-1). Later we will prove certain restrictions on the substitutions that can occur at a given $(m+1)-$block. Finally, one should bear in mind that the sequences are merely shorthand for sums $\sum_{j=1}^{k} \epsilon_j \theta^j$.

Proof. The sequences ϵ and ϵ' are equivalent iff θ satisfies the polynomial equation

$$\sum_{k=1}^{n} \delta_k x^k = 0,$$

where $\delta_k = \epsilon_k - \epsilon'_k$. This happens iff β is a root of the polynomial $f(x) = \sum_{k=1}^{n} \delta_k x^{n-k}$. Since $p(x)$ is the *minimal* polynomial of β, p divides f in the polynomial ring $\mathbf{Z}[x]$, i.e., there exists a polynomial $g(x) = \sum_{j=0}^{n-m} b_j x^j$ with integer coefficients b_j such that $f = pg$.

The coefficients of $g(x)$ provide the schedule of substitutions. The jth coefficient b_j tells how many times to apply the fundamental relation to the $(m+1)-$block starting at the $(j+m+1)$th entry from the right; the sign (+ or −) tells whether the fundamental relation should be applied in the forward or the backward direction. This is best illustrated by a simple example: let the fundamental relation be $100 \sim 011$ (β = golden ratio), and let $\epsilon = 10100$ and $\epsilon' = 01111$. Then $f(x) = x^4 - x^3 - x - 1$ and $g(x) = x^2 + 1$. The quotient polynomial g has coefficients +1 in the 0th and 2nd positions; these indicate that the substitution $100 \to 011$ should be made once to the block at the extreme left and once to the block two positions in, i.e., $10100 \to 01200 \to 01111$.

That the quotient polynomial $g(x)$ does in fact provide a left-right sequence of substitutions in *every* case is easily proved by induction on n. We prove a more general statement: if $u_1, u_2, \ldots, u_n$ and $v_1, v_2, \ldots, v_n$ are equivalent sequences of *integers* (meaning that $f(\theta) = \sum_{j=1}^{n} (u_j - v_j)\theta^n = 0$) then $g(x) = f(x)/p(x)$ provides a correct left-right sequence of substitutions. The statement is clearly true if $n = m$, because in this case f is an integer multiple of p, as p is the minimal polynomial of θ. Suppose it is true whenever $n < N$, and let $u_1, u_2, \ldots, u_n$ and $v_1, v_2, \ldots, v_n$ be equivalent sequences of length $n = N$. If the leading term of $g(x)$ is b_{n-m-1} then clearly $u_1 - v_1 = b_{n-m-1}$. (NOTE: Here we use the fact that β is an algebraic *integer* – this guarantees that the leading term in the minimal polynomial has coefficient 1.) Consequently, if one makes b_{n-m} substitutions of the fundamental relation in the leading $(m+1)-$block of $u_1, u_2, \ldots, u_n$ one obtains an equivalent sequence $w_1, w_2, \ldots, w_n$ whose leading entry is v_1. But $w_1, w_2, \ldots, w_n$ and $v_1, v_2, \ldots, v_n$ are equivalent sequences with the same first entry, so $w_2, w_3, \ldots w_n$ and $v_2, v_3, \ldots, v_n$ are equivalent sequences of length $n-1$. The induction hypothesis now implies the result. □

Corollary 1. *Suppose that $\varepsilon \sim \varepsilon'$ are equivalent sequences of length n. Let $f(x) = \sum_{k=1}^{n} \delta_k x^{n-k}$, where $\delta_k = \varepsilon_k - \varepsilon'_k$, and let $g(x) = \sum_{j=0}^{n-m} b_j x^j = f(x)/p(x)$. Then ε' may be obtained from ε by applying the fundamental relation b_j times to the $(m+1)-$block starting at the $(j+m+1)$th entry from the right, in the order $j = n-m, n-m-1, \ldots, 0$.*

We will call the sequence of transformations described in this corollary the *canonical transformation* taking ε to ε'. It produces $(n-m-1)$ intermediate sequences

$$\varepsilon^{(1)}, \varepsilon^{(2)}, \ldots, \varepsilon^{(n-m-1)}.$$

In modifying $\varepsilon^{(i)}$ to obtain $\varepsilon^{(i+1)}$, only entries in the $(m+1)-$block starting at the $(i+1)$th entry are changed; since these $(m+1)-$blocks move left to right one unit at a time, it follows that

(a) the first i entries of $\varepsilon^{(i)}$ agree with corresponding entries of ε'; and
(b) the final $n-i-m$ entries of $\varepsilon^{(i)}$ agree with the corresponding entries of ε.

In particular, only those entries of $\varepsilon^{(i)}$ in the $m-$block starting at the $(i+1)$th entry can be different from 0 or 1. Neither Proposition 8 nor Corollary 2 implies that the entries of $m-$blocks arising in intermediate sequences of canonical transformations are bounded — in principle, there could be infinitely many possible such $m-$blocks. In the next section, we shall show that in fact there are only finitely many possibilities, and give an algorithm for identifying them.

5. Admissible $m-$Blocks

Define an *admissible $m-$block* to be an $m-$block (a sequence of m integers) that occurs in some intermediate sequence in a canonical transformation of some sequence ε of 0s and 1s to an equivalent sequence ε' of 0s and 1s, and define

$$\mathcal{A} = \{ \text{ admissible } m-\text{blocks } \}.$$

The matrices M_0, M_1 that will appear in the random matrix products used to characterize the entropy α (see Theorems 3-4 below) will have rows and columns indexed by the admissible $m-$blocks. In this section we shall verify that $\mathcal{A}$ is finite and give an effective procedure for identifying its elements.

Proposition 7. $|\mathcal{A}| < \infty$.

We know two proofs, one based on Garsia's lemma, the other on the following result, a special case of Proposition 2.5 of [9] (which is attributed to J. P. Bezevin). We reproduce the proof, because it provides an explicit bound for the size of entries of admissible $m-$blocks (see Corollary 2 below).

Proposition 8. *Assume that β is a Pisot number. Then there exists a positive integer $C < \infty$ with the following property: For any polynomial $f(x)$ with coefficients $0, 1, -1$ such that $p|f$ in the ring $\mathbf{Z}[x]$, the coefficients b_j of the quotient polynomial $g(x) = f(x)/p(x) = \sum_{j=0}^{n-m} b_j x^j$ are bounded in modulus by C. Furthermore, if $\beta = \beta_1, \beta_2, \ldots, \beta_m$ are the roots of $p(x)$, then we may take $C = [C']$, where $[\cdot]$ denotes greatest integer and*

$$C' = (1-\beta^{-1})^{-1} \prod_{i=2}^{m} (1-|\beta_i|)^{-1}. \tag{7}$$

Note: Since the coefficients b_j of any such quotient polynomial are integers, it follows that they are elements of the finite set $\{-C, -C+1, \ldots, C-1, C\}$.

Proof. The argument, in brief, is as follows. Say that a polynomial $P(x)$ has the *bounded division property* if for each real $a > 0$ there exists a constant $C_a < \infty$ such that for any polynomial $F(x) = \sum_{i=0}^{n} a_i x^i \in \mathbf{C}[x]$ with coefficients a_i bounded in modulus by a, if $P(x)|F(x)$ in the ring $\mathbf{C}[x]$, then the coefficients of the quotient polynomial $F(x)/P(x)$ are bounded in modulus by C_a. Call the assignment $a \to C_a$ an *expansion function* for $P(x)$. It is easily seen that if $P_1(x)$ and $P_2(x)$ both have the bounded division property, then so does their product P_1P_2, and the product has expansion function $a \to C(a)$ satisfying $C(a) \leq C^{(1)} \circ C^{(2)}(a)$, where $C^{(1)}(\cdot)$ and $C^{(2)}(\cdot)$ are expansion functions for P_1 and P_2, respectively. It is also easily seen that any *linear* polynomial of the form $x - \alpha$ has the bounded division property iff $|\alpha| \neq 1$, and that in this case an expansion function is

$$\begin{aligned} C_a &= (1 - |\alpha|)^{-1} a \qquad \text{if } |\alpha| < 1; \\ &= (1 - |\alpha|^{-1})^{-1} a \qquad \text{if } |\alpha| > 1. \end{aligned}$$

Therefore, a polynomial $P(x) \in \mathbf{C}[x]$ has the bounded division property iff it has no roots on the unit circle.

Since β is a Pisot number, its minimal polynomial $p(x)$ has no roots on the unit circle. Thus, it has the bounded division property. In fact, $\beta = \beta_0 > 1$ is the only root outside the unit circle, so the results of the previous paragraph imply the bound (7) for the constant $C = C_1$. □

Corollary 2. *Assume that β is a Pisot number. Let H be the maximum of the absolute values of the coefficients of the minimal polynomial $p(x)$. Then the entries of admissible $m-$blocks are bounded in modulus by C_*, where*

$$C_* = CHm + CH + 1; \tag{8}$$

here C is the constant in (7) and m is the degree of the minimal polynomial $p(x)$ of β.

Proof. Consider the sequence of transformations specified in Corollary 1. The fundamental relation is applied b_{n-m} times at the leftmost $(m+1)-$block of ε; then b_{n-m+1} times at the leftmost-but-one $(m+1)-$block of the resulting sequence; etc. The integers b_j are the coefficients of a quotient polynomial f/p, where f has all coefficients in $0, -1, 1$, so by Proposition 8 each b_j satisfies $|b_j| \leq C$. Applying the fundamental relation once (either in the forward direction or the backward direction) changes entries by at most H (in absolute value), so applying it b_j times changes entries by at most CH. Moreover, an entry is changed only if its position is in the current $(m+1)-$block. Since no position is in the current $(m+1)-$block more than $m+1$ times, it follows that the maximum amount by which it can be modified is no more than $CH(m+1)$. Therefore, since the entries of the original sequence ε are either 0 or 1, entries in intermediate sequences cannot be more than $1 + CH(m+1)$ or less than $-CH(m+1)$.

□

Next, we discuss the problem of explicitly enumerating the set $\mathcal{A}$ of admissible $m-$blocks. By Corollary 2, $\mathcal{A}$ is contained in the set of $m-$tuples with entries in $\{-C_*, -C_*+1, \cdots, C_*\}$. Unfortunately, even for Pisot numbers of small degree m, C_*^m may be fairly large compared to the cardinality of $\mathcal{A}$ (see the table below) and so a brute force search of the set of all such

$m-$tuples may be needlessly lengthy. A much more useful test seems to be that provided by the following lemma.

Lemma 5. *If $\varepsilon = \varepsilon_1\varepsilon_2\ldots\varepsilon_m$ is an admissible $m-$block, then*

$$-(1-\theta)^{-1} \le \sum_{i=1}^{m} \varepsilon_i\theta^i \le \theta^m(1-\theta)^{-1}.$$

Proof. If an application of the fundamental relation changes a sequence $\delta_1\delta_2\ldots\delta_n$ of integers to an equivalent sequence $\delta_1'\delta_2'\ldots\delta_n'$, then $\sum\delta_j\theta^j = \sum\delta_j'\theta^j$. Now recall that admissible $m-$blocks occur only in intermediate sequences of canonical transformations of equivalent sequences of 0s and 1s, and that in these canonical transformations, the fundamental relation is applied repeatedly, left to right. Consequently, if $\varepsilon = \varepsilon_1\varepsilon_2\ldots\varepsilon_m$ is an admissible $m-$block such that $\sum_{j=1}^m \varepsilon_j\theta^j > \sum_{j=1}^m \theta^j$, then the "excess" $\sum_{j=1}^m \varepsilon_j\theta^j - \sum_{j=1}^m \theta^j$ cannot be greater than $\sum_{j=0}^\infty \theta^{-j}$, because this excess must eventually be "transferred" to the right of the $m-$block. Similarly, if $\sum_{j=1}^m \varepsilon_j\theta^j < 0$, then this "deficiency" cannot be less than $-\sum_{j=0}^\infty \theta^{-j}$, because it must eventually be compensated by the terms to the right of the $m-$block. □

Thus, $\mathcal{A} \subset \mathcal{B}$, where $\mathcal{B}$ is the set of all $m-$blocks with entries bounded in modulus by C_* such that the inequalities of Lemma 5 are satisfied. Let $\mathbf{b} = b_1b_2\ldots b_m$ and $\mathbf{b}' = b_1'b_2'\ldots b_m'$ be elements of $\mathcal{B}$. Write

$$\mathbf{b} \to \mathbf{b}'$$

if $\mathbf{b}'$ can be obtained from $\mathbf{b}$ by (1) appending either a 0 or a 1 to the *end* of $\mathbf{b}$, then (2) applying the fundamental relation to the resulting $(m+1)-$block either $-b_1$ or $-b_1+1$ times, and finally (3) deleting the entry at the *beginning* of the transformed $(m+1)-$block. For example, if the fundamental relation is $100 \sim 011$, then $20 \to 12$ (20:201:112:12), but $21 \not\to 01$. Observe that for a given $m-$block $\mathbf{b}$ there are at most 4 $m-$blocks $\mathbf{b}'$ such that $\mathbf{b} \to \mathbf{b}'$.

Lemma 6. *An element $\mathbf{b}$ of $\mathcal{B}$ is an element of $\mathcal{A}$ if and only if there is a finite chain*

$$\mathbf{b}^{(1)} \to \mathbf{b}^{(2)} \to \cdots \to \mathbf{b} \to \cdots \to \mathbf{b}^{(K)}$$

in which both endpoints $\mathbf{b}^{(1)}$ and $\mathbf{b}^{(K)}$ are sequences of 0s and 1s.

Proof. This is a direct consequence of Corollary 1 and the discussion following. The fundamental relation can be applied either $-b_1$ or $-b_1+1$ times, because the resulting $(m+1)-$block will begin with either a 0 or a 1. □

Similarly, write $\mathbf{b} \xrightarrow{r} \mathbf{b}'$ if $\mathbf{b}' \to \mathbf{b}$. Note that $\mathbf{b} \xrightarrow{r} \mathbf{b}'$ iff $\mathbf{b}'$ can be obtained from $\mathbf{b}$ by (1) prepending either a 0 or a 1 to the *beginning* of $\mathbf{b}$, then (2) applying the fundamental relation to the resulting $(m+1)-$block either $-b_m$ or $-b_m+1$ times, and finally (3) deleting the entry at the *end* of the transformed $(m+1)-$block. Again, for a given $m-$block $\mathbf{b}$ there are at most 4 $m-$blocks $\mathbf{b}'$ such that $\mathbf{b} \xrightarrow{r} \mathbf{b}'$. Moreover, the conclusion of the preceding lemma holds with all the arrows $\to$ replaced by $\xrightarrow{r}$.

We may now give an algorithm for determining the admissible $m-$blocks.

Algorithm: *Set $\mathcal{A}'$ to be the set of all $m-$blocks with 0-1 entries. Update $\mathcal{A}'$ by appending to it the set of all $\mathbf{b}' \in \mathcal{B}$ such that for some $\mathbf{b} \in \mathcal{A}'$, $\mathbf{b} \to \mathbf{b}'$. Continue updating $\mathcal{A}'$ in this manner until it stabilizes. Next, set $\mathcal{A}''$ to be the set of all 0,1 $m-$blocks. Update $\mathcal{A}''$ by appending to it the set of all $\mathbf{b}' \in \mathcal{B} \cap \mathcal{A}'$ such that for some $\mathbf{b} \in \mathcal{A}''$, $\mathbf{b} \xrightarrow{r} \mathbf{b}'$. Continue updating $\mathcal{A}''$ in this manner until it stabilizes. Finally, $\mathcal{A} = \mathcal{A}''$.*

Following is a table of all Pisot numbers between 1 and 2 of degree ≤ 4 whose minimal polynomial has coefficients ≤ 2 in absolute value, together with the cardinality of $\mathcal{A}$ and the value of C_*.

β	$p(x)$	$\lvert\mathcal{A}\rvert$	$C_*(\beta)$
1.618	$x^2 - x - 1$	8	21.6
1.325	$x^3 - x - 1$	200	949.5
1.466	$x^3 - x^2 - 1$	68	417.0
1.755	$x^3 - 2x^2 + x - 1$	28	310.5
1.839	$x^3 - x^2 - x - 1$	14	128.1
1.380	$x^4 - x^3 - 1$	1702	28288.8
1.866	$x^4 - 2x^3 + x - 1$	92	3535.6
1.905	$x^4 - x^3 - 2x^2 + 1$	154	4790.6
1.928	$x^4 - x^3 - x^2 - x - 1$	24	1398.2

The following result will not be needed in the analysis to follow, but it is essentially equivalent to Proposition 7.

Proposition 9. *There is a deterministic finite automaton $\mathcal{M}$ such that the language accepted by $\mathcal{M}$ is the set of all finite sequences $\gamma = \gamma_1\gamma_2\ldots\gamma_n$ with entries $\gamma_i = (\varepsilon_i, \varepsilon_i')$ such that $\varepsilon \sim \varepsilon'$, where*

$$\begin{aligned} \varepsilon &= \varepsilon_1\varepsilon_2\ldots\varepsilon_n, \\ \varepsilon' &= \varepsilon_1'\varepsilon_2'\ldots\varepsilon_n'. \end{aligned}$$

See [12], sec. 2.2 for the definition of a finite automaton and the language it accepts.

6. Counting Equivalence Classes by Matrix Multiplication

We will now show that the cardinality of the equivalence class of a given sequence of zeroes and ones may be represented as a certain natural matrix product. Two matrices, which we shall call M_0 and M_1, are involved in these products. The rows and columns are indexed by the set $\mathcal{A}$ of admissible $m-$blocks. By the results of the preceding section, if $\beta = \beta_m$ is one of the simple Pisot numbers, then the admissible $m-$blocks are the sequences of length m with entries in $\{-1, 0, 1, 2\}$, subject to the restrictions (1) any 2 that occurs must be followed only by 0s and preceeded only by 1s; and (2) any -1 that occurs must be followed only by 1s and preceeded only by 0s. Note that for the case of the golden ratio β_2 there are 8 such sequences:

$$\begin{array}{ll} (0,0) & (1,0) \\ (0,1) & (1,1) \\ (1,2) & (0,-1) \\ (2,0) & (-1,1) \end{array}$$

Entries of the matrices M_i are zeroes and ones; they indicate whether transitions between the various $m-$blocks can be made when i is the entry immediately to the right of the current $m-$block. Allowable transitions may be described as follows: (1) If the leading entry of the current $m-$block is a 0 or 1, the transition that concatenates i to the $m-$block at the end and deletes the initial entry is allowable. Thus, for example, when $\beta = \beta_2$, the transition

$(1,0) \to (0,i)$ is allowable. (2) A transition that concatenates i to the $m-$block at the end, then applies the fundamental relation an integral number of times to the resulting $(m+1)-$block, then deletes the initial entry is allowable iff the deleted entry is a 0 or 1 *and* the resulting $m-$block is admissible. For example, when $\beta = \beta_2$, if $i = 1$ then the transition $2,0 \to 1,2$ is allowable, but $2,0 \to 0,1$ is not, because the deleted initial entry in the latter case would be a 2, nor is $1,1 \to 2,2$, because $2,2$ is not an admissible $m-$block.

Note that although the matrices are large (8 by 8 in the case of the golden ratio and correspondingly larger for the other β_m) they are "sparse': for each block b and each $i = 0,1$ there are at most two blocks b' such that the entry $M_i(b,b')$ is 1. Thus, to describe M_i it is easier to make a list of allowable transitions than to write out the entire matrix.

Example: $\beta = \beta_2$.

Following is a table showing the allowable transitions.

M_0			M_1		
From	To		From	To	
(0,0)	(0,0)		(0,0)	(0,1)	
(0,1)	(1,0)	(0,-1)	(0,1)	(1,1)	(0,0)
(1,0)	(0,0)	(1,1)	(1,0)	(0,1)	(1,2)
(1,1)	(1,0)		(1,1)	(1,1)	
(2,0)	(1,1)		(2,0)	(1,2)	
(1,2)	(2,0)		(1,2)		
(0,-1)			(0,-1)	(-1,1)	
(-1,1)	(0,-1)		(-1,1)	(0,0)	

Notice that in M_0 there are no allowable transitions from the $m-$block $(0,-1)$. This is because a -1 must always be followed by a 1 in an $m-$block. But notice also that there is an allowable transition from $(0,-1)$ in the matrix M_1.

Proposition 10. *Let $\epsilon = \epsilon_1\epsilon_2\ldots\epsilon_{n+m}$ be an arbitrary sequence of zeroes and ones. Then the number of sequences of zeroes and ones equivalent to ϵ and ending in the $m-$block $b_1 = \epsilon'_1\epsilon'_2\ldots\epsilon'_m$ is the (b_0,b_1)th entry of the matrix product $M_{\epsilon_{m+1}}M_{\epsilon_{m+2}}\ldots M_{\epsilon_{n+m}}$, where $b_0 = \epsilon_1\epsilon_2\ldots\epsilon_m$.*

It obviously follows that the total number of 0-1 sequences equivalent to ϵ is the sum over all $m-$blocks b_1 of the (b_0,b_1)th entries of the matrix product $M_{\epsilon_{m+1}}M_{\epsilon_{m+2}}\ldots M_{\epsilon_{n+m}}$.

Proof. We prove a slightly more general statement, specifically, that for *any* admissible $m-$blocks b_0, b_1 the b_0, b_1th entry of $M_{\epsilon_{m+1}}M_{\epsilon_{m+2}}\ldots M_{\epsilon_{n+m}}$ is the number of allowable left-right transformations of the sequence $b_0\epsilon_{m+1}\epsilon_{m+2}\ldots\epsilon_{n+m}$ ending in a sequence whose first n entries are 0s and 1s and whose last m entries are the $m-$block b_1. The proof is by induction on $n \geq 1$. The case $n = 1$ is easily checked: the matrices M_0 and M_1 were defined in such a way that this would be true.

The induction step is also easy. Assume that it is true for all integers $< n$ and let $\epsilon = \epsilon_1\epsilon_2\ldots\epsilon_{n+m}$ be a sequence of 0s and 1s. Then we know by the induction hypothesis that the number of canonical (left-right) transformations of the sequence $b_0\epsilon_{m+1}\epsilon_{m+2}\ldots\epsilon_{n+m-1}$ that move the "cursor" $n-1$ steps to the right and result in a sequence beginning with $n-1$ 0s and 1s and ending in a given block b is the b_0, bth entry of $M_{\epsilon_{m+1}}\ldots M_{\epsilon_{n+m-1}}$. To obtain the number of canonical transformations of $b_0\epsilon_{m+1}\epsilon_{m+2}\ldots\epsilon_n$ that move the cursor

n steps to the right and transform the final $m-$block to b_1, partition the count by the possible values of the next-to-last $m-$block b (when the cursor is $n-1$ units to the left). Regardless of the steps taken to reach b the number of ways to go from b to b_1 is given by the b, b_1th entry of $M_{\epsilon_{n+m}}$, again by the induction hypothesis. (Note: Once the first $n-1$ steps of the substitution sequence have been made, the first $n-1$ entries of the resulting transformed sequence play no further role, since the cursor has now moved to their right. Therefore, even though different substitution sequences may result in transformed sequences with different initial $(n-1)-$blocks, as long as they result in the same $m-$block b to the right of the cursor the number of ways to complete the transformation will be the same for each.) Finally, summing over all possible $m-$blocks b and using the definition of matrix multiplication one obtains the desired equality, completing the inductive phase of the proof. □

It now follows that the asymptotic behavior of the random variables $\pi_n(\epsilon)$ when $\epsilon_1, \epsilon_2, \ldots$ is a sequence of iid Bernoulli-$\frac{1}{2}$ random variables is determined by that of the random matrix product

$$\Pi_n = M_{\epsilon_{m+1}} M_{\epsilon_{m+2}} \ldots M_{\epsilon_n}.$$

The asymptotic behavior of these random matrix products is in turn described by the Furstenberg-Kesten theorem ([8]; also [3], ch. 1). This theorem implies that

$$\lim_{n\to\infty} \frac{1}{n} \log \|\Pi_n\| = \lambda \tag{9}$$

where

$$\lambda = \lim_{n\to\infty} \frac{1}{n} E \log \|\Pi_n\|$$

is the "top Lyapunov exponent" of the sequence Π_n. Neither the convergence nor the value of λ is affected by the choice of norm.

If the entries of Π_n were eventually positive, with probability 1, then not only the norm but also the individual entries would grow exponentially at the rate λ. This is not the case, however. For example, examination of the entries of M_0, M_1 in the case $\beta = \beta_2$ (see above) shows that the (1,1) and (1,0) columns of Π_n cannot both have positive entries. But the (Euclidean) norms of the *rows* do grow exponentially at rate λ, as the following result shows. For any $\mathbf{b} \in \mathcal{A}$, let $u_{\mathbf{b}}$ be the vector with $\mathbf{b}$th entry 1 and all other entries 0.

Proposition 11. *For each* $\mathbf{b} \in \mathcal{A}$,

$$\lim_{n\to\infty} \|u'_{\mathbf{b}} \Pi_n\|^{\frac{1}{n}} = e^{\lambda}.$$

Proof. First, we will argue that for each pair $\mathbf{b}, \mathbf{b}'$ of admissible $m-$blocks there exists a sequence $i_1 i_2 \ldots i_k$ of 0s and 1s such that the $\mathbf{b}, \mathbf{b}'$ entry of $M_{i_1} M_{i_2} \ldots M_{i_k}$ is positive. Write $\mathbf{b} \to \mathbf{b}'$ if there is such a sequence. Let $\mathbf{b}, \mathbf{b}'$ be arbitrary admissible $m-$blocks. By definition of $\mathcal{A}$, for each $\mathbf{b} \in \mathcal{A}$ there exist $m-$blocks $\mathbf{b}'', \mathbf{b}'''$ with 0-1 entries such that $\mathbf{b} \to \mathbf{b}''$ and $\mathbf{b}''' \to \mathbf{b}$. Consequently, it suffices to prove the contention only for pairs of $m-$blocks $\mathbf{b}, \mathbf{b}'$ with all entries in $\{0, 1\}$. But if $\mathbf{b}, \mathbf{b}'$ have only 0-1 entries, then it is certainly true that $\mathbf{b} \to \mathbf{b}'$, because all the $m-$blocks in the concatenation $\mathbf{b}\mathbf{b}'$ are 0-1 $m-$blocks and are therefore admissible, and hence, if $\mathbf{b}' = i_1 i_2 \ldots i_m$, then the $\mathbf{b}, \mathbf{b}'$ entry of $M_{i_1} M_{i_2} \ldots M_{i_m}$ is positive.

Next, we will argue that with probability 1 there exists an admissible $m-$block $\mathbf{b}$ (possibly random) so that

$$\liminf_{n\to\infty} \|u'_{\mathbf{b}}\Pi_n\|^{\frac{1}{n}} \geq e^{\lambda}. \tag{10}$$

Since the entries of Π_n and $u_{\mathbf{b}}$ are all nonnegative, and since $\sum_{\mathbf{b}\in\mathcal{A}} u_{\mathbf{b}}$ is the vector with all entries 1, it follows that $\|\sum_{\mathbf{b}\in\mathcal{A}} u'_{\mathbf{b}}\Pi_n\| \geq \|\Pi_n\|$. Consequently, by the Furstenberg-Kesten theorem, there exists an admissible $m-$block $\mathbf{b}$ such that

$$\limsup_{n\to\infty} \|u'_{\mathbf{b}}\Pi_n\|^{\frac{1}{n}} \geq e^{\lambda}.$$

But since $\|\Pi_n\|^{\frac{1}{n}} \to e^{\lambda}$, an elementary argument shows that lim sup may be replaced by lim inf in the above equation.

The proposition follows from the results of the last two paragraphs. Choose an admissible $m-$block $\mathbf{b}_*$ such that, with positive probability, (10) holds for $\mathbf{b} = \mathbf{b}_*$. Fix any $\mathbf{b} \in \mathcal{A}$; then by the result of the first paragraph (and the fact that the matrices in the product Π_n are iid) there exist (random) $N_1 < N_2 < \cdots$ such that the $\mathbf{b}_*$th entry of $u'_{\mathbf{b}}\Pi_{N_j}$ is ≥ 1. Hence, for each $j = 1, 2, \ldots$ and each $n \geq 1$,

$$\|u'_{\mathbf{b}}\Pi_{N_j+n}\| \geq \|u'_{\mathbf{b}_*}\Pi_{N_j}^{-1}\Pi_{N_j+n}\|.$$

Now consider the sequence of sequences $\{u'_{\mathbf{b}_*}\Pi_{N_j}^{-1}\Pi_{N_j+n}\}_{n\geq 1}$, for $j = 1, 2, \ldots$. By the Kolmogorov 0-1 Law, this sequence is ergodic, since the invariant $\sigma-$algebra is contained in the tail $\sigma-$algebra of the sequence $\varepsilon_1, \varepsilon_2, \ldots$. Moreover, by the choice of $\mathbf{b}_*$, the probability that (10) holds for any one of these sequences is $p > 0$. Consequently, by the Birkhoff ergodic theorem, with probability 1 there exists $j \geq 1$ such that

$$\liminf_{n\to\infty} \|u'_{\mathbf{b}_*}\Pi_{N_j}^{-1}\Pi_{N_j+n}\|^{\frac{1}{n}} \geq e^{\lambda}.$$

The proposition now follows directly from this and the last displayed inequality. □

Let $\mathcal{A}_0$ be the set of all admissible $m-$blocks with only 0-1 entries, and let $v = \sum_{\mathbf{b}\in\mathcal{A}_0} u_{\mathbf{b}}$. Let $\mathbf{1} = \sum_{\mathbf{b}\in\mathcal{A}} u_{\mathbf{b}}$ be the vector with all entries 1, and let $w = \mathbf{1} - v$.

Corollary 3. *For each* $\mathbf{b} \in \mathcal{A}_0$,

$$\lim_{n\to\infty} (u'_{\mathbf{b}}\Pi_n v)^{\frac{1}{n}} = e^{\lambda}.$$

Proof. Recall that for each $\mathbf{b} \in \mathcal{A}$ there exist a finite 0-1 sequence $\mathbf{i}_{\mathbf{b}} = i_1 i_2 \ldots i_J$ and an admissible $m-$block $\mathbf{b}'$ with 0-1 entries such that the $\mathbf{b}, \mathbf{b}'$ entry of $M_{i_1} M_{i_2} \cdots M_{i_J}$ is ≥ 1. Moreover, there exists $K < \infty$ such that the length J of the string $\mathbf{i}_{\mathbf{b}}$ satisfies $J \leq K$ for all $\mathbf{b} \in \mathcal{A}$, because there are only finitely many $\mathbf{b} \in \mathcal{A}$.

Let u be a vector with nonnegative entries, not all 0. For each $k \geq 1$ let $\mathbf{b}^{(k)} \in \mathcal{A}$ be such that the $\mathbf{b}^{(k)}$ entry of $u'\Pi_k$ is maximal. (Note that $\mathbf{b}^{(k)}$ is random.) For each $k \geq 1$ it may happen, with probability at least 2^{-K}, that $\varepsilon_{k+j} = i_j \; \forall 1 \leq j \leq J$, where $i_1 i_2 \ldots i_J$ is a 0-1 sequence as in the previous paragraph, for $\mathbf{b} = \mathbf{b}^{(k)}$. Denote this event by G_k. Since $P(G_k) \geq 2^{-K}$ for each k, and since the events $G_K, G_{2K}, \ldots$ are independent, the Borel-Cantelli lemma implies that

$$P\left(\left(\cup_{k=n-\sqrt{n}}^{n-K} G_k\right)^c \; i.o.\right) = 0.$$

We will argue that on the event $\cup_{k=n-\sqrt{n}}^{n-K} G_k$,

$$u'\Pi_n v \geq \frac{\|u'\Pi_n\|}{|\mathcal{A}|^3 \max_{n-\sqrt{n}\leq k\leq n} \|\Pi_k^{-1}\Pi_n\|}. \tag{11}$$

This will prove the corollary, because by the preceding proposition $\|u'\Pi_n\|^{1/n} \to e^\lambda$, and by Lemma 7 below,

$$\limsup_{n\to\infty} \max_{n-\sqrt{n}\leq k\leq n} \|\Pi_k^{-1}\Pi_n\|^{\frac{1}{n}} \leq 1.$$

On the event G_k, there is at least one $\mathbf{b}' \in \mathcal{A}_0$ and $1 \leq J \leq K$ such that the $\mathbf{b}'$ entry of $u'\Pi_{k+J}$ is at least as large as the $\mathbf{b}^{(k)}$ entry of $u'\Pi_k$. Since $\mathbf{b}' \in \mathcal{A}_0$, there is at least one $\mathbf{b}'' \in \mathcal{A}_0$ such that the $\mathbf{b}', \mathbf{b}''$ entry of $\Pi_{k+J}^{-1}\Pi_n$ is ≥ 1. Consequently,

$$\begin{aligned} u'\Pi_n v &\geq u'\Pi_{k+J} u_{\mathbf{b}'} \\ &\geq u'\Pi_k u_{\mathbf{b}^{(k)}} \\ &\geq \|u'\Pi_k\|/|\mathcal{A}|, \end{aligned}$$

the last inequality because of the choice of $\mathbf{b}^{(k)}$. On the other hand, no entry of $u'\Pi_n$ can be larger than $\|u'\Pi_k\|\,\|\Pi_k^{-1}\Pi_n\|$, and hence

$$\|u'\Pi_n\| \leq |\mathcal{A}|^2\,\|u'\Pi_k\|\,\|\Pi_k^{-1}\Pi_n\|.$$

The inequality (11) follows from the last two displayed inequalities. □

Lemma 7. $\limsup_{n\to\infty} \max_{n-\sqrt{n}\leq k\leq n} \|\Pi_k^{-1}\Pi_n\|^{\frac{1}{n}} \leq 1.$

Proof. Recall that the random matrices M_{ε_i} can assume only two values, so $E\|M_{\varepsilon_i}\| = \rho < \infty$. Now each $\Pi_k^{-1}\Pi_n$ is the product of independent copies of M_{ε_1}, so for any $n-\sqrt{n} \leq k \leq n$ and any $\varepsilon > 0$, Markov's inequality implies that

$$P\{\|\Pi_k^{-1}\Pi_n\| > (1+\varepsilon)^n\} \leq \rho^{\sqrt{n}}(1+\varepsilon)^{-n}.$$

The probability that $\|\Pi_k^{-1}\Pi_n\| > (1+\varepsilon)^n$ for *some* $n-\sqrt{n} \leq k \leq n$ cannot be more than $\sqrt{n}$ times this. Because $\sum_n \sqrt{n}\rho^{\sqrt{n}}(1+\varepsilon)^{-n} < \infty$, it follows from the Borel-Cantelli lemma that with probability 1,

$$\max_{n-\sqrt{n}\leq k\leq n} \|\Pi_k^{-1}\Pi_n\| > (1+\varepsilon)^n$$

for only finitely many n. □

Theorem 3. *Let λ be the top Lyapunov exponent for the sequence Π_n of random matrix products defined above. If $\varepsilon = \varepsilon_1\varepsilon_2\ldots$ where $\varepsilon_1, \varepsilon_2, \ldots$ are i.i.d. Bernoulli-$\frac{1}{2}$ then with probability 1*

$$\lim_{n\to\infty} \pi_n(\varepsilon)^{\frac{1}{n}} = \frac{1}{2}e^\lambda = \alpha. \tag{12}$$

Note: This proves Theorem 1 in the special case $p = \frac{1}{2}$.

Proof. By Proposition 10, the cardinality of the equivalence class of $\varepsilon_1\varepsilon_2\ldots\varepsilon_n$ is $u'_{\mathbf{b}}\Pi_n v$, where $\mathbf{b} = \varepsilon_1\varepsilon_2\ldots\varepsilon_m$. Consequently, $\pi_n(\varepsilon) = 2^{-n}u'_{\mathbf{b}}\Pi_n v$. By the preceding corollary, $\pi_n(\varepsilon)^{1/n} \to e^\lambda/2$. □

7. Calculating the Weight of an Equivalence Class

For values of p other than $\frac{1}{2}$ the probabilities $\pi_n(\epsilon)$ cannot be obtained merely by enumerating the equivalence classes of the various sequences. Nevertheless, it is still possible to represent them in terms of a matrix product, with matrices similar structurally to those used to enumerate the classes.

Let $\epsilon = \epsilon_1\epsilon_2\ldots\epsilon_n$ be a given sequence of 0s and 1s; its probability under the Bernoulli-p measure is $p^{S(\varepsilon)}q^{n-S(\varepsilon)}$, where $S(\varepsilon) = \sum_{i=1}^{n}\varepsilon_i$ is the number of 1s in ε. Observe that not all sequences of length n equivalent to ϵ have the same probabilities: for instance, when $\beta = \beta_2$ the sequences 100 and 011 are equivalent, but the first has probability pq^2 while the second has probability p^2q.

Suppose in general that $\epsilon \sim \epsilon'$; then by Corollary 1 there is a sequence of left-right substitutions based on the fundamental relation, some positive and some negative, transforming ϵ to ϵ'. Every time a single positive substitution is made, the likelihood of the sequence is multiplied by the factor $(p/q)^\kappa$, where κ is the number of 1s on the left side of the fundamental relation minus the number of 1s on the right side. This is because each time the fundamental relation is applied (in the positive direction) there is a net increase of κ in the number of 1s and a net decrease of κ in the number of 0s. Similarly, every time a single negative substitution is made, the likelihood is multiplied by $(q/p)^\kappa$. Hence the likelihood of ϵ' is ρ times the likelihood of ϵ, where $\rho = (p/q)^{N\kappa}$ and N is the total number of positive substitutions minus the total number of negative substitutions in the transformation from ϵ to ϵ'.

In the last section we defined matrices M_i with 0-1 entries to indicate whether transformations between different $m-$blocks could occur when the entry to the right of the block was i. We found that the entries of products of these matrices give the number of ways to do substitutions left to right on the original sequence and arrive at given $m-$blocks. If we want to tally total likelihood (relative to the parameter p) instead of cardinality, then instead of 1s as entries we should put a positive number indicating the (multiplicative) effect on likelihood. Thus, we redefine the matrices M_0 and M_1 as follows: For admissible $m-$blocks b and b' let the (b,b')th entry of M_i be zero when the transition $b \to b'$ is not allowable; p^iq^{1-i} when the transition is allowable and no substitution is made in the transition; and $p^iq^{1-i} \times \nu$ when the transition is allowable and a substitution is made, where $\nu = (p/q)^{l\kappa}$ and l is the number of substitutions made in the transition (l may be positive or negative). For example, when $\beta = \beta_2$, $\kappa = -1$, so the nonzero entries of M_0 and M_1 are as given in the following table:

M_0			M_1		
From	To	Entry	From	To	Entry
(0,0)	(0,0)	q	(0,0)	(0,1)	p
(0,1)	(1,0)	q	(0,1)	(1,1)	p
(0,1)	(0,-1)	q^2/p	(0,1)	(1,0)	q
(1,0)	(0,0)	q	(1,0)	(0,1)	p
(1,0)	(1,1)	p	(1,0)	(1,2)	p^2/q
(1,1)	(1,0)	q	(1,1)	(1,1)	p
(2,0)	(1,1)	p	(2,0)	(1,2)	p^2/q
(1,2)	(2,0)	q	(1,2)		
(0,-1)			(0,-1)	(-1,1)	p
(-1,1)	(0,-1)	q^2/p	(-1,1)	(0,0)	q

Note that the positive entries occur in exactly the same locations as for the matrices defined in the previous section. The $(0,-1)$ row of M_0 has no nonzero entries, since there are no allowable transitions from this 2−block when the next entry of the sequence is a 0, nor does the $(1,2)$row of M_1.

For any admissible m−block $b = \epsilon_1\epsilon_2\ldots\epsilon_m$ define the (row) vector v_b to have bth entry $p^i q^{m-i}$, where i is the number of 1s in $\epsilon_1\epsilon_2\ldots\epsilon_m$, and all other entries zero. Define w to be the column vector with entries 1 for admissible m−blocks with no -1s or 2s, and all other entries 0.

Proposition 12. *Let $\epsilon = \epsilon_1\epsilon_2\ldots\epsilon_n$ be a sequence of 0s and 1s, and let $b_0 = \epsilon_1\epsilon_2\ldots\epsilon_m$. Then*

$$\pi_n(\epsilon) = v_{b_0} M_{\epsilon_{m+1}} M_{\epsilon_{m+2}} \ldots M_{\epsilon_n} w.$$

The proof is virtually the same as that of the corresponding result in the previous section. There are analogues of all the results of the preceding section. Let λ be the top Lyapunov exponent of the sequence $\Pi_n = M_{\epsilon_{m+1}} M_{\epsilon_{m+2}} \cdots M_{\epsilon_n}$, and let $u_{\mathbf{b}}$, $\mathbf{b} \in \mathcal{A}$ and $v = \sum_{\mathcal{A}} u_{\mathbf{b}}$.

Proposition 13. *For each $\mathbf{b} \in \mathcal{A}_0$,*

$$\lim_{n\to\infty} (u'_{\mathbf{b}} \Pi_n v)^{\frac{1}{n}} = e^{\lambda}.$$

Theorem 4. *Let λ be the top Lyapunov exponent for the sequence Π_n of random matrix products defined above. If $\varepsilon = \varepsilon_1\varepsilon_2\ldots$ where $\varepsilon_1, \varepsilon_2, \ldots$ are i.i.d. Bernoulli-p then with probability 1*

$$\lim_{n\to\infty} \pi_n(\varepsilon)^{\frac{1}{n}} = e^{\lambda} = \alpha. \tag{13}$$

This completes the proof of Theorem 1.

It should now be clear how to modify the approach for other 0-1 stochastic sequences, in particular, k−step Markov chains. If $k \geq m$ then it is necessary to index the rows and columns by $(k+1)$−blocks rather than m−blocks, to keep track of the likelihoods involved. It should also be clear that the whole approach could be adapted to sequences of random variables valued in other finite subsets of $\mathbf{Z}$ than $\{0,1\}$.

8. Computing the Lyapunov exponent

Computation of Lyapunov exponents is a difficult problem, even for small matrices. However, because of the special structure of the matrices arising in sections 6-7, computation to a reasonable degree of accuracy is possible for simple Pisot numbers β_m of small degree m. In this section, we restrict our discussion to the cases $\beta = \beta_m$, $m \geq 2$.

Fix a value of the Bernoulli parameter p and let $\epsilon_1, \epsilon_2, \ldots$ be iid Bernoulli-p random variables. Let M_0 and M_1 be the matrices defined in the preceding section. Our problem is to calculate the top Lyapunov exponent of the sequence

$$\Pi_n = M_{\epsilon_1} M_{\epsilon_2} \ldots M_{\epsilon_n}.$$

By Proposition 13, for any vector u with nonnegative entries, not all 0,

$$\lambda = \lim_{n\to\infty} \frac{1}{n} \log \|u\Pi_n\|. \tag{14}$$

A propitious choice is the vector u with $(0,0,\ldots,0)$ entry q, $(1,1,\ldots,1)$ entry p, and all other entries 0. (Recall that the rows and columns of the matrices M_i are indexed by admissible m−blocks, hence so are the entries of row vectors.)

The reason for the peculiar choice of u is that the random sequence of vectors $u\Pi_n$ visits the ray through u infinitely often.

Lemma 8. *Let T be the infimum of the set of integers $n \geq 1$ such that $u\Pi_n$ is a scalar multiple of u. Then*

$$ET < \infty;$$

in particular, $T < \infty$ with probability 1.

The proof will be given later. In fact, we will obtain an explicit algebraic (in p) expression for ET: thus, for instance, when $m = 2$ and $p = \frac{1}{2}$, $ET = 12$.

The existence of such a stopping time allows one to re-express the Lyapunov exponent in a simpler form. The vector $u\Pi_T$ is a scalar multiple of the starting vector u. Since scalars may be factored out of matrix products, this implies that the process $u\Pi_n$ effectively "regenerates" at time T: specifically, the law of the sequence

$$\frac{u\Pi_{T+n}}{\|u\Pi_T\|}, \ n \geq 0$$

is the same as that of the original sequence $u\Pi_n$, $n \geq 0$. It follows that there are *infinitely many* times at which $u\Pi_n$ is a scalar multiple of u; label these times $T_0 = 0 < T_1 = T < T_2 < T_3 < \cdots < \infty$. At each of these times the sequence regenerates. For each $n \geq 1$,

$$\frac{1}{T_n} \log \|u\Pi_{T_n}\| = \frac{1}{T_n} \sum_{j=1}^{n} \log \frac{\|u\Pi_{T_j}\|}{\|u\Pi_{T_{j-1}}\|},$$

and the summands are iid with finite expectation (this because $ET < \infty$). Dividing numerator and denominator by n and using the strong law of large numbers on each, one obtains the limit in (14) along the subsequence T_n as a ratio of expectations. But we know the limit in (14) exists, by the Furstenberg-Kesten theorem, so the limit of the whole sequence must be given by the same ratio of expectations. (Note: A standard argument shows that the limit exists without reference to the Furstenberg-Kesten theorem.) In summary,

Corollary 4.

$$\lambda = \frac{E \log \|u\Pi_T\|}{ET} \tag{15}$$

and

$$\delta = \frac{E \log \|u\Pi_T\|}{(\log \theta) ET}. \tag{16}$$

Since we have an exact expression for ET, only the expectation in the numerator requires estimation. Unfortunately, no further simplification seems possible: the numerator can only be estimated by simulation or summation over paths. It should be remarked, however, that this is a significant improvement on the crude representation of λ as $\lim_{n\to\infty} E \log \|\Pi_n\|$, because in general one cannot say how fast the convergence in this limit is.

The numerator may, in general, be estimated by simulation with very little effort. Greater or lesser precision may be obtained by adjusting the number of replications. To obtain estimates with accuracy (to confidence level .99) to within $\pm .002$ requires on the order of one million replications (depending on the values of p, m). Results for $m = 2$ and various values of p are reported in the table below.

p	dimension	estimated error
.5	.9954	.0008
.4	.9868	.001
.3	.9501	.002
.2	.8499	.004
.1	.6085	.008
.05	.3877	.003

We have conducted simulations for all rational values of p between 0 and .5 with denominators less than 12 and also for $.01, .02, \ldots, .09$; these seem to indicate that the dimension is a strictly increasing function of $p \in (0, .5]$. As yet we have no rigorous argument for this conjecture.

Proof of Lemma 8. We shall discuss only the case $m = 2$, the general case being completely similar but requiring more cumbersome notation. Thus, u is the vector with $(0,0)$ entry q, $(1,1)$ entry p, and all other entries 0.

Let $u_n = u\Pi_n$.We will say that any admissible 2−block for which the corresponding entry of u_n is positive *occurs* in u_n. Not all admissible 2−blocks may occur *simultaneously* in u_n: for instance, 11 and 10 cannot occur simultaneously. However, all admissible m−blocks may occur in some u_n (see the proof of Proposition 11).

Since the random variables $\epsilon_1, \epsilon_2, \ldots$ are iid Bernoulli-p, the sequence must, with probability 1, contain arbitrarily long blocks of 1s. Consider what happens to the vectors u_n when such a long block of 1s occurs. If either of the blocks $(0,-1)$ or $(-1,1)$ occurs in u_n, then after one or two successive 1s these blocks will be converted to the block $(0,0)$. If either of the blocks $(2,0)$ or $(1,2)$ occurs in u_n, then after one or two successive 1s they will be "killed". If any of the four blocks with no 2 or -1 entries occurs in u_n then after either one or two successive 1s all will be converted to one of the blocks $(0,0),(0,1),(1,1)$. Consequently, regardless of which blocks occur in u_n, if $\epsilon_{n+1} = \epsilon_{n+2} = 1$ then only 00, 01, and 11 can occur in u_{n+2}. Moreover, the blocks 00 and 01 cannot occur simultaneously. Note that if the two blocks that occur in u_{n+2} are 00 and 11 then if $\epsilon_{n+3} = 1$, the blocks that occur in u_{n+3} must be 01 and 11. Thus, with probability 1, for some n the vector u_n will have positive entries in the 01 and 11 entries and all other entries zero.

Now consider what happens when u_n has positive entries in the 01 and 11 entries and all other entries zero, and $\epsilon_{n+1} = \varepsilon_{n+2} = 0$. Using the table in section 6, one easily verifies that u_{n+2} must have 11 entry $pq(u_n(01) + u_n(11))$, 00 entry $q^2(u_n(01) + u_n(11))$, and all other entries zero. Thus, u_{n+2} is a scalar multiple of u.

The arguments of the last two paragraphs show that, depending on the composition of u_n, a regeneration will occur if either (a) $\varepsilon_{n+1} = 1$ and $\varepsilon_{n+2} = \varepsilon_{n+3} = 0$, or (b)$\varepsilon_{n+1} = \varepsilon_{n+2} = 1$ and $\varepsilon_{n+3} = \varepsilon_{n+4} = 0$. Elementary arguments show that this must happen eventually, with probability 1; in fact, that the expected time until it happens is finite.

□

The preceding argument shows that the regeneration event is determined by the set of admissible m−blocks that occur in a given u_n and the subsequent pattern of 0s and 1s in the sequence ϵ_j, but *not* on the actual coefficients of the m−blocks that occur in u_n. It follows that T is a stopping time for the Markov chain whose state at any time n is the set of admissible m−blocks that occur in u_n. It follows from elementary Markov chain theory that ET may be computed by solving a simple matrix equation. In the special case $m = 2$

there are 7 distinct sets of admissible $m-$blocks that may occur simultaneously in u_n; the transition probabilities between these sets are easy to write down, and the resulting matrix equation is easily solved via *Mathematica*. This yields the identity (when $m = 2$)

$$ET = \frac{-2 - p + p^2}{-p + 2p^2 - 2p^3 + p^4}. \tag{17}$$

Similar formulas may be obtained for arbitrary m, although the size of the matrix equation that must be solved grows with m.

9. The Associated Graph

In this section we indicate another approach to the representation of the probabilities $\pi_n(\varepsilon)$ by matrix products. This approach is essentially geometric in nature, relying on simple properties of a natural graph associated with the Pisot number $\beta \in (1,2)$. In the special case β =golden ratio, this graph is the "Fibonacci tree" exploited by Alexander and Zagier [2].

The (directed) graph $\Gamma = (\mathcal{V}, \mathcal{E})$ is defined as follows. The vertex set $\mathcal{V}$ is the union of countably many finite sets $\mathcal{V}_n$, $n \geq 0$; the elements of $\mathcal{V}_n$ are the possible sums $x_n(\varepsilon) = \sum_{k=1}^n \varepsilon_k \theta^k$, where ε is a 0-1 sequence. The (directed) edges connect vertices at depths n and $n+1$: there are edges from $x_n(\varepsilon)$to $x_{n+1}(\varepsilon)$ for every $\varepsilon \in \Sigma$, and no others. To each edge from $x_n(\varepsilon)$ to $x_n(\varepsilon) + \theta^{n+1}$ attach weight p, and to each edge from $x_n(\varepsilon)$ to $x_n(\varepsilon)$ attach weight $q = 1 - p$.

For any two vertices $x_n(\varepsilon), x_n(\varepsilon')$ at the same depth n, define their distance $\rho(x_n(\varepsilon), x_n(\varepsilon'))$ by

$$\rho(x_n(\varepsilon), x_n(\varepsilon')) = \beta^n |x_n(\varepsilon) - x_n(\varepsilon')|.$$

Fix a constant $\kappa > 2\theta/(1-\theta)$. For any vertex $x_n(\varepsilon)$ define its *neighborhood* $\mathcal{N}(x_n(\varepsilon)) = \mathcal{N}_\kappa(x_n(\varepsilon))$ to be the set of vertices $x_n(\varepsilon')$ at the same depth such that $\rho(x_n(\varepsilon), x_n(\varepsilon')) < \kappa$. Say that two vertices $x_n(\varepsilon), x_k(\varepsilon')$ (not necessarily at the same depth) have the same *neighborhood type* if there is a bijective mapping between $\mathcal{N}(x_n(\varepsilon))$ and $\mathcal{N}(x_k(\varepsilon'))$ that preserves the distance function ρ.

Proposition 14. *There are only finitely many neighborhood types, i.e., there is a finite set of vertices $\mathcal{S}$ such that every vertex of Γ has the same neighborhood type as one of the vertices in $\mathcal{S}$.*

Remark: The reader should notice the similarity with Th. of [4], concerning the Cayley graph of a finitely generated, discontinuous group of isometries of a hyperbolic space.

Proof. This follows from Garsia's lemma, which implies that there is a lower bound d on the $\rho-$distance between distinct vertices in any neighborhood. Consider the neighborhoods $\mathcal{N}$ of a vertex $x_{n+k}(\varepsilon)$ and $\mathcal{N}'$ of the vertex $x_k(\sigma^n\varepsilon)$ (here $\sigma$is the shift operator). There is a distance-preserving injection $\mathcal{N}' \to \mathcal{N}$, because there is a copy Γ' of the graph Γ embedded in Γ emanating from the vertex $x_n(\sigma^n\varepsilon)$. Consequently, for any sequence $\varepsilon = \varepsilon_1\varepsilon_2\ldots \in \Sigma$ and each $n \geq 1$ there is a chain of distance-preserving injections

$$\mathcal{N}(x_1(\sigma^{n-1}\varepsilon)) \to \mathcal{N}(x_2(\sigma^{n-2}\varepsilon)) \to \ldots \mathcal{N}(x_n(\varepsilon)).$$

By Garsia's lemma, all sufficiently long chains must stabilize, i.e., there is a finite integer k such that all the injections after the kth in any such chain must be *bijections* (if not, there

would be neighborhoods with arbitrarily large cardinalities). It follows that there are only finitely many neighborhood types. □

Let $\mathcal{T}$ be the (finite) set of possible neighborhood types.

For any vertex $x_{n+1}(\varepsilon)$ at depth $n+1 \geq 1$, let $B(x_{n+1}(\varepsilon)) \subset \mathcal{V}_n$ be the set of all vertices at depth n from which emanate directed edges of Γ leading into $\mathcal{N}(x_{n+1}(\varepsilon))$.

Lemma 9. $B(x_{n+1}(\varepsilon)) \subset \mathcal{N}(x_n(\varepsilon))$.

Proof. Observe that for any vertex $x_{n+1}(\varepsilon')$ at depth $n+1$ there are at most 2, and at least 1, directed edges from V_n to $x_{n+1}(\varepsilon')$. If $\varepsilon'_{n+1} = 1$ then there must be a p–edge from $x_n(\varepsilon')$ to $x_{n+1}(\varepsilon')$, and there may be a q–edge from some $x_n(\varepsilon'')$ to $x_{n+1}(\varepsilon')$; if $\varepsilon'_{n+1} = 0$ then there must be a q–edge from $x_n(\varepsilon')$ to $x_{n+1}(\varepsilon')$, and there may be a p–edge from some $x_n(\varepsilon'')$ to $x_{n+1}(\varepsilon')$. Consequently, $B(x_{n+1}(\varepsilon))$ is finite.

Now consider any directed edge from a vertex $x_n(\varepsilon') \in \mathrm{V}_n$ leading into $B(x_{n+1}(\varepsilon))$. Depending on whether it is a p–edge or a q–edge,

$$|x_n(\varepsilon') + \theta^{n+1} - x_n(\varepsilon) + \varepsilon_{n+1}\theta^{n+1}| < \kappa\theta^{n+1} < 2\theta^{n+2}/(1-\theta)$$

or

$$|x_n(\varepsilon') - x_n(\varepsilon) + \varepsilon_{n+1}\theta^{n+1}| < \kappa\theta^{n+1} < 2\theta^{n+2}/(1-\theta).$$

In either case,

$$\beta^n|x_n(\varepsilon') - x_n(\varepsilon)| < \theta + 2\theta^2/(1-\theta) < 2\theta/(1-\theta) < \kappa,$$

since $\kappa > 2\theta/(1-\theta)$. □

Corollary 5. *For any 0-1 sequence $\varepsilon = \varepsilon_1\varepsilon_2\ldots$, the neighborhood type $\mathcal{N}(x_{n+1}(\varepsilon))$ is determined by the neighborhood type $\mathcal{N}(x_n(\varepsilon))$ and the entry ε_{n+1}.*

For any infinite 0-1 sequence ε and each $n \geq 1$, define $Y_n(\varepsilon)$ to be the vector of probabilities

$$Y_n(\varepsilon) = (\pi_n(\varepsilon'))_{\mathcal{N}(x_n(\varepsilon))},$$

where the entries are indexed by the elements of $\mathcal{N}(x_n(\varepsilon))$ (for each element $x_n(\varepsilon')$ of $\mathcal{N}(x_n(\varepsilon))$, choose *one* representative ε' and let $\pi_n(\varepsilon')$ be the entry for that element). The probabilities $\pi_n(\varepsilon)$ are computed under the Bernoulli-p measure (the measure on sequence space Σ making $\varepsilon_1, \varepsilon_2, \ldots$ i.i.d. Bernoulli-p random variables).

Lemma 10. *There exist nonegative matrices $L_{\mathcal{N},i}$, for $\mathcal{N} \in \mathcal{T}$ and $i = 0, 1$, such that for every 0-1 sequence ε and every $n \geq 1$,*

$$Y_{n+1}(\varepsilon) = L_{\mathcal{N}(x_n(\varepsilon),\varepsilon_{n+1}}Y_n(\varepsilon). \tag{18}$$

Proof. For any sequence $\varepsilon' = \varepsilon'_1\varepsilon'_2\ldots$ the probability $\pi_{n+1}(\varepsilon)$ is gotten by summing the probabilities of all sequences ε'' such that $x_{n+1}(\varepsilon'') = x_{n+1}(\varepsilon')$. In geometric terms, $\pi_{n+1}(\varepsilon)$ is obtained by summing all $\pi_n(\varepsilon')w$ for depth-n vertices $x_n(\varepsilon')$ such that there is an edge from $x_n(\varepsilon')$ to $x_{n+1}(\varepsilon)$; $w = p$ or q, depending on the weight attached to the edge. Note that there are at most two, and at least one, terms in this sum. Moreover, by (9), the factors $\pi_n(\varepsilon')$ in the two terms are both entries of the vector $Y_n(\varepsilon)$. It is clear that these equations may be written in the matrix form (18), with the matrix $L_{\mathcal{N}(x_n(\varepsilon),\varepsilon_{n+1}}$ having rows indexed by elements of $\mathcal{N}(x_{n+1}(\varepsilon))$, columns indexed by elements of $\mathcal{N}(x_n(\varepsilon))$, and entries $0, p, 1-p$. □

Note that this argument used the assumption that μ is a Bernoulli measure. Generalization to arbitrary shift-invariant measures on Σ would seem to be problematic. However, if μ is a Markov measure (i.e., under μ the coordinate process $\varepsilon_1, \varepsilon_2, \ldots$ is a $k-$step Markov chain for some $k < \infty$), then there is an obvious generalization of Lemma 10. We refrain from giving the details.

Lemma 10 provides a representation of the probability vectors $Y_n(\varepsilon)$ in terms of matrix products, but, unfortunately, the matrices need not be square. Nevertheless, it is possible to use the theory of random matrix products to determine the asymptotic behavior of the sequence $Y_n(\varepsilon)$. The idea is to embed the vectors $Y_n(\varepsilon)$ into vectors $\overline{Y}_n(\varepsilon)$ whose entries are indexed by elements of the union $\mathcal{G}$ of all possible neighborhood types $\tau \in \mathcal{T}$, and to embed the matrices $L_{\mathcal{N},i}$ (as blocks) into matrices $\mathcal{L}_i$ with rows and columns indexed by elements of $\mathcal{G}$. The entries of $\overline{Y}_n(\varepsilon)$ are all zero except for those indexed by elements of the neighborhood type $\mathcal{N}(x_n(\varepsilon)$; these entries have the same values as the corresponding entries of $Y_n(\varepsilon)$. Then Lemma 10 implies that $\overline{Y}_{n+1}(\varepsilon) = \mathcal{L}_{\varepsilon_{n+1}} \overline{Y}_n(\varepsilon)$, and consequently,

Proposition 15. *If $n > k$,*

$$\overline{Y}_n(\varepsilon) = \mathcal{L}_{\varepsilon_n} \mathcal{L}_{\varepsilon_{n-1}} \cdots \mathcal{L}_{\varepsilon_{k+1}} \overline{Y}_k(\varepsilon). \tag{19}$$

Proposition 15 provides another representation of $\pi_n(\varepsilon)$ in terms of a random matrix product (recall Propositions 10 and 12), as the value of $\pi_n(\varepsilon)$ is one of the entries of the vector $\overline{Y}_n(\varepsilon)$. Although this representation is in some ways more natural, and easier to derive, it seems less useful for actual computations. This is because the matrices $\mathcal{L}_i$ are, in general, much larger than the matrices in the products in Propositions 10 and 12, and enumeration of the neighborhood types may be practically more difficult than the enumeration of the admissible $m-$blocks for a given β.

References

[1] Alexander, J. C., and Yorke, J. A. (1984) Fat baker's transformations. *Ergodic Th. Dyn. Syst.* **4**, 1-23.

[2] Alexander, J. C., and Zagier, D. (1991) The entropy of a certain infinitely convolved Bernoulli measure. *J. London Math. Soc. (2)* **44**, 121-134.

[3] Bougerol, and LaCroix (198) *Products of Random Matrices.* Birkhauser.

[4] Cannon, J. (1984) The combinatorial structure of co-compact discrete hyperbolic groups. *Geom. Dedicata* **16**, 123-148.

[5] Erdös , P. (1939) On a family of symmetric Bernoulli convolutions. *American J. of Math.* **61**, 974-976.

[6] Erdös , P. (1940) On the smoothness properties of a family of Bernoulli convolutions. *American J. of Math.* **62** , 180-186.

[7] Falconer, K. (1985) *The Geometry of Fractal Sets.* Cambridge Univ. Press.

[8] Furstenberg, H., and Kesten, H. (1963) Products of random matrices. *Annals of Mathematical Statistics*

[9] Frougny, C. (1992) Representations of numbers and finite automata. *Math. Systems Th.***25**, 37-60.

[10] Garsia, A. (1962) Arithmetic properties of Bernoulli convolutions. *Trans. Amer. Math. Soc.* **102**, 409-432.

[11] Garsia, A. (1963) Entropy and singularity of infinite convolutions. *Pacific J. Math.* **13**, 1159-1169.

[12] Hopcroft, J. , and Ullman, J. (1979) *Introduction to Automata Theory, Languages, and Computation.* Addison-Wesley, Reading.

[13] Parry, W. On the β expansions of real numbers. *Acta Math. Acad. Sci. Hung.* **11**, 401-416.

[14] Parry, W. (1964) Intrinsic Markov chains. *Trans. Amer. Math. Soc.* **112**, 55-64.

[15] Peterson, (1983) *Ergodic Theory.* Cambridge Univ. Press.

[16] Przytycki, F., and Urbanski, M. (1989) On the Hausdorff dimension of some fractal sets. *Studia Math.* **93**, 155-186.

[17] Renyi, A. (1957) Representations for real numbers and their ergodic properties. *Acta Math. Acad. Sci. Hung.* **8**, 477-493.

[18] Salem, R. (1963) *Algebraic Numbers and Fourier Analysis.* Heath, Lexington.
[19] Siegel, C. L. (1944) Algebraic integers whose conjugates lie in the unit circle. *Duke Math. J.* **11**,597-602.
[20] Solomyak, B. (1994) Preprint.
[21] Walters, P. (1982) *An Introduction to Ergodic Theory.* Springer, New York.
[22] Young, L.-S. (1982) Dimension, entropy, and Lyapunov exponents. *Ergodic Th. Dyn. Syst.* **2**, 109-124.

Department of Statistics, Mathematical Sciences Bldg., Purdue University
E-mail address: lalley@stat.purdue.edu