Normality in non-integer bases and polynomial time randomness

Javier Almarza Santiago Figueira

April 13, 2015

Abstract

It is known that if $x \in [0,1]$ is polynomial time random (i.e. no polynomial time computable martingale succeeds on the binary fractional expansion of x) then x is normal in any integer base greater than one. We show that if x is polynomial time random and $\beta > 1$ is Pisot, then x is "normal in base β", in the sense that the sequence $(x\beta^n)_{n\in\mathbb{N}}$ is uniformly distributed modulo one. We work with the notion of *P-martingale*, a generalization of martingales to non-uniform distributions, and show that a sequence over a finite alphabet is distributed according to an irreducible, invariant Markov measure P if an only if no P-martingale whose betting factors are computed by a deterministic finite automaton succeeds on it. This is a generalization of Schnorr and Stimm's characterization of normal sequences in integer bases. Our results use tools and techniques from symbolic dynamics, together with automata theory and algorithmic randomness.

1 Introduction

A weak notion of randomness for sequences over a finite alphabet $\Sigma = \{0,\dots,b-1\}$ ($b \in \mathbb{N}$) is *normality*, introduced by Borel in 1909. Normality may be regarded as a "law of large numbers" for blocks of events, in the sense that the average occurrences of a block $\sigma \in \Sigma^*$ of length n converges to $|\Sigma|^{-n}$. A real number x is called *normal in base b* ($b \in \mathbb{N}$) if its expansion in base b is normal. While almost all numbers are normal to all bases it is not too difficult to see that this notion is not base invariant. In fact for any multiplicatively independent bases b and b' the set of numbers normal to b but not normal to b' has full Hausdorff dimension [14]. We say a number x is *absolutely normal* if it is normal in all integer bases greater than one. It is not difficult to see that x is normal in base b if and only if the sequence $(xb^n)_{n\in\mathbb{N}}$ is u.d. modulo one, and then x is absolutely normal if and only if $(xb^n)_{n\in\mathbb{N}}$ is uniformly distributed (u.d.) modulo one for all integer $b > 1$.

Polynomial time randomness is another weak notion of randomness. We say that x is *polynomial time random in base b* if no martingale (a formaliza-

tion of *betting strategy*) on the alphabet $\{0, \ldots, b-1\}$ which is computable in polynomial time succeeds on the expansion of x in base b. A result of Schnorr [16] states that if x is polynomial time random in base b then x is normal in base b.

It was recently shown [6] that polynomial time randomness is base invariant, so that being polynomial time random in a single base implies being normal for all bases, i.e. being absolutely normal. The converse is not true, since there are absolutely normal numbers which are computable in polynomial time [1, 6, 10], and these cannot be polynomial time random. The following question was left open in [6]:

Question 1.1. Suppose that x is polynomial time random. Is the sequence $(x\beta^n)_{n\in\mathbb{N}}$ u.d. modulo one for all rational $\beta > 1$?

The distribution of $(x\beta^n)_{n\in\mathbb{N}}$ modulo one for rational β seems, however, fairly intractable. It is unknown, for instance, if $((3/2)^n)_{n\in\mathbb{N}}$ is u.d. modulo one. Our first main result is that there is a class of algebraic reals for which the question may be readily handled:

Theorem 1.2. If x is polynomial time random then the sequence $(x\beta^n)_{n\in\mathbb{N}}$ is u.d. modulo one for all Pisot $\beta > 1$.

Observe that any non-integer Pisot β is irrational, and as a consequence of a result of Brown, Moran and Pearce [4, Theorem 2], there are uncountably many reals which are absolutely normal but $(x\beta^n)_{n\in\mathbb{N}}$ is not u.d. modulo one.

The formulation of normality to integer bases β in terms of modulo one uniform distribution allows us to understand normality as equivalent to what ergodic theory calls *genericity*, an equivalence which boils down to two facts: 1) the map $T_\beta(x) = (\beta x) \mod 1$ on $[0, 1)$ is equivalent to a "shift" rightwards in the space of sequences $\{0, \ldots, \beta - 1\}^{\mathbb{N}}$ when x is mapped to its base β expansion; 2) $(x\beta^n) \mod 1 = T_\beta^n(x)$.

When a non-integer base β is considered, 2) is immediately false, while 1) has no clear reformulation, since there is no obvious candidate for a space of sequences that "represent" numbers in base β. It is here that the theory of β-shifts and β-representations, developed, among others, by Parry [12] and Bertrand [2], helps fill in the missing pieces.

Once the space of sequences that represent numbers in the base β (using symbols from $\Sigma = \{0, \ldots, \lceil\beta\rceil - 1\}$) is defined, it is equipped with a natural shift transformation and a measure P_β called the *Parry measure*, which plays the same role that the uniform or Lebesgue measure played in integer representation. Indeed, a result by Bertrand says that, when β is Pisot, if a real number x has a β-expansion that is distributed according to P_β (this is the analogue notion to being "normal in base β"), then $(x\beta^n)_{n\in\mathbb{N}}$ is u.d. modulo one.

To see how this is useful for the proof of Theorem 1.2, let us say we have a number z such that $(z\beta^n)_{n\in\mathbb{N}}$ is not u.d. modulo one. Then, by Bertrand's theorem, its β-representation would have some block σ whose average occurrences do not converge to $P_\beta(\sigma)$. We would then want to construct a polynomial time martingale that succeeds by betting on that block, as is done in the integer base case.

However, this cannot be done in a straightforward manner, since the martingale condition as used in the algorithmic randomness literature, assumes outcomes should be distributed according to the uniform measure.

We work with a generalized definition of martingales which captures the idea of a "fair" betting strategy when expansions are supposed to obey some non-uniform distribution P. Indeed, this definition of a *P-martingale* will capture the broader sense of *martingale* as it is used in probability theory. In this setting, not only may the probability of the next symbol be different from $|\Sigma|^{-1}$, it may also show all forms of conditional dependence on the preceding symbols. It should be noted that randomness notions under measures different from Lebesgue have already been considered in, for example, [15].

Schnorr and Stimm [17] show that a sequence is normal in base b if and only if no martingale on the alphabet of b digits whose betting factors are computed by a deterministic finite automaton (DFA) succeeds on the expansion of x in base b. Our second main result is a generalization of this last statement in terms of P-martingales:

Theorem 1.3. A sequence is distributed according to an irreducible, invariant Markov measure P if an only if no P-martingale whose betting factors are computed by a DFA succeeds on it.

The importance of Markov measures is that they exhibit enough memorylessness to make them compatible with the memoryless structure of a DFA.

As regards β-representations, a second result by Bertrand establishes that for β Pisot P_β, the natural measure on β-expansions, is "hidden" Markov. By extending Theorem 1.3 to hidden Markov measures we are able to construct a P_β-martingale generated by a DFA that succeeds on the β-expansion of z. We use the polynomial time computability of the β-expansion and of the measure P_β to show that an integer base (i.e. classical) martingale which succeeds on z can be constructed from our P_β-martingale, following the same ideas used in [6].

1.1 Outline

The paper is organized as follows. In §2 we introduce some basics from symbolic dynamics, mainly the definition of Markov and sofic subshifts, and the notion of sequences distributed according to invariant measures P

over the shift. In §3 we introduce the notion of P-(super)martigales and show the characterization given by Theorem 1.3. In §4 we introduce some definitions and results regarded to representation of reals in non-integer bases, in particular, Pisot bases. Finally, in §5 we put all pieces together to get Theorem 1.2.

2 Subshifts and measures

Throughout this work Σ will denote an alphabet of finitely many symbols, which will be denoted by a, b, c, etc. The set of all words over the alphabet Σ will be denoted by Σ^*, and the set of all words of length k over the alphabet Σ will be denoted Σ^k (so $\Sigma^* = \bigcup_k \Sigma^k$). Greek letters σ, τ and so on will be used for finite words in Σ^* Letters s, s' will be used for infinite sequences in $\Sigma^{\mathbb{N}}$. The i-th symbol of the sequence s will be denoted s_i. Concatenation will bear no special symbol, so we may write $\sigma = ab$, $s = as'$, $\rho = \sigma\tau$, etc. For a word σ and $k \in \mathbb{N}$ we denote with σ^k the string of length $k|\sigma|$ which consists of the k-times repetition of σ, and with σ^∞ to the infinite sequence which consist of the repetition of σ infinitely may times. For any sequence $s \in \Sigma^{\mathbb{N}}$ we will denote by $s\restriction_N$ the word that consists of the first N symbols of s, and by $\langle s : k\rangle$ the same sequence s when regarded as a sequence in Σ^k. For a word σ and a non-negative integer k, let $\sigma\restriction_k$ denote the subword of σ consisting of its last k symbols (in case $l < k$ then $\sigma\restriction_k$ is just σ). By $\sigma \preceq \tau$ we will denote that σ is a prefix of τ, and by $\sigma \prec \tau$ we will denote that σ is a strict prefix of τ. We will use the same notation ($\sigma \prec s$ and $\sigma \preceq s$) for sequences s. By $[\sigma]$ we will denote the cylinder set consisting of all infinite sequences extending σ, i.e. $[\sigma] = \{s \in \Sigma^{\mathbb{N}} : \sigma \prec s\}$.

Definition 2.1. Given a finite alphabet Σ, a *subshift* is a tuple (X, T) where

1. X is some closed (hence, compact) subset of $\Sigma^{\mathbb{N}}$ with the product topology
2. X is invariant under T (that is, $T(X) \subseteq X$); and
3. T is the continuous mapping defined by $(T(s))_n = s_{n+1}$.

If $X = \Sigma^{\mathbb{N}}$ we say that (X, T) is the full $|\Sigma|$-shift. The language associated to (X, T), denoted $L(X) \subseteq \Sigma^*$, consists of all words appearing in the sequences of X.

Notice that $L(X)$ is a factorial and prolongable language, that is, it contains all subwords of its words and, if $\sigma \in L(X)$, then there exists a non-empty word τ in Σ^* such that $\sigma\tau \in L(X)$. Conversely, given any language, there is a corresponding closed subset of sequences. For a language $L \subseteq \Sigma^*$ we define

$$X_L = \{s \in \Sigma^{\mathbb{N}} : \forall N,\ s\restriction_N \in L\}.$$

Observe that X_L is closed and that if L is factorial then X_L is shift invariant, hence, it is a subshift of $\Sigma^{\mathbb{N}}$. Moreover, if L is factorial and prolongable, then $L(X_L) = L$

Definition 2.2. A *k-step Markov shift* (also known as a *subshift of finite type*, or *SFT*) is a subshift (X, T) of $\Sigma^{\mathbb{N}}$ such that there exists a set G (called a *grammar*) of admissible words of length k satisfying

$$X = \{s \in \Sigma^A \colon (\forall i \in A)\ s_i s_{i+1} \dots s_{i+k-1} \in G\}.$$

These are called Markov shifts by analogy with the Markov processes of probability theory. For these, looking back at the last k values of the process (say, the last k flipped coins) is enough to know the probabilities of the next value (looking further backwards does not change these conditional probabilities). In the case of Markov shifts, looking at the last $k-1$ symbols is enough to know if the next symbol is admissible.

Definition 2.3. A probability measure P on $\Sigma^{\mathbb{N}}$ is called *k-step Markov* for some fixed $k \in \mathbb{N}$ if for all $\sigma, \tau \in \Sigma^*$, $|\sigma| \geq k$, $P([\sigma\tau] \mid [\sigma]) = P([\rho\tau] \mid [\rho])$ where $\rho = \sigma\restriction_k$.

The above condition is actually called k-step *homogenous* Markov. A strict Markovian condition would read $P([\sigma\tau] \mid [\sigma]) = P(T^{-(l-k)}([\rho\tau]) \mid T^{-(l-k)}([\rho]))$. Since we will never consider non-homogenous Markov processes, we can spare the reader this extra terminology.

From now on we will simplify notation and write $P(\sigma)$ instead of $P([\sigma])$ for any word $\sigma \in \Sigma^*$. Given a 1-step Markov probability measure P on $\Sigma^{\mathbb{N}}$ we define its transition matrix $(p_{a,b})_{a,b\in\Sigma}$ to be

$$p_{a,b} = P(ab \mid a).$$

An *invariant* measure on a subshift (X, T) is a probability measure P on X (with its Borel σ-algebra $\mathcal{B}$) such that $P \circ T^{-1} = P$. Notice that, by definition of T, $P \circ T^{-1}(\sigma) = \sum_{a\in\Sigma} P(a\sigma)$ for words σ, and invariance need only be checked for such word cylinders.

Let P be a 1-step Markov measure on $\Sigma^{\mathbb{N}}$ with transition matrix $M = (p_{a,b})_{a,b\in\Sigma}$. Define the vector $v \in \mathbb{R}^\Sigma$, $v_a = P(a)$. Then P is invariant if and only if v is a left eigenvector of M. Let P be a k-step Markov on $\Sigma^{\mathbb{N}}$. Let $\Theta_k = \{\tau \in \Sigma^k \colon P(\tau) > 0\}$. Then P induces a 1-step Markov measure P^k on $\Theta_k^{\mathbb{N}}$ with transition matrix $(p^k_{\sigma,\tau})_{\sigma,\tau\in\Theta_k} = P(\sigma\tau \mid \sigma)$.

A probability measure P on $\Sigma^{\mathbb{N}}$ is called *irreducible* if for any words σ, τ such that $P(\sigma) > 0$, $P(\tau) > 0$ there is some word ρ such that $P(\sigma\rho\tau) > 0$. A nonnegative $n \times n$ matrix A is *irreducible* when the associated directed graph G_A, which has n nodes and in which there is an edge from node i to node j if and only if $A_{ij} > 0$, is strongly connected. A 1-step Markov measure is

irreducible if its transition matrix is irreducible, a k-step invariant Markov measure is irreducible if the matrix $(p^k_{\sigma,\tau})_{\sigma,\tau\in\Theta_k}$ is irreducible.

The following is the Perron-Frobenius Theorem for Markov chains in the finite state case (see [7, Theorem 1.3.5] and [11, Theorems 1.7.5-7 and Exercise 1.7.5]):

Theorem 2.4. Let P and P' be two invariant, irreducible 1-step Markov measures on $\Sigma^{\mathbb{N}}$ such that their transition matrices are the same. Then $P = P'$.

Given two subshifts (X, T) and (X', T) with $X \subseteq \Sigma^A, X' \subseteq \Sigma'^A$, a *factor map* is an onto map $\psi : X \to X'$ which commutes with the shift operator, that is, $\psi \circ T = T \circ \psi$. Markov shifts are not closed under factor maps, but the following class of subshifts is.

Definition 2.5. A *sofic subshift* is the image of a Markov shift under a factor map.

Example 2.6. Let us consider X to be the 2-step Markov shift on $\{0,1\}^{\mathbb{N}}$ with grammar $G = \{00, 10, 01\}$. For each s in X, let $\psi(s)$ be such that $(\psi(s))_i = 0$ if $s_i = s_{i+1} = 0$ and $(\psi(s))_i = 1$ otherwise. The image of ψ is the set of infinite sequences such that all blocks of consecutive 1's are of even length (blocks of 0's are of arbitrary length). This corresponds to the regular expression $((11)^*0^*)^*$. Notice that this is not a Markov shift, since no matter how big k is, looking back at the last k values is not enough to determine whether a 0 is admissible next.

Definition 2.7. Given a subshift (X, T), $s \in X$ and an invariant measure μ on X, we will say s is *μ-distributed* if for all continuous $f\colon X \to \mathbb{R}$ we have

$$\lim_{N\to\infty} \frac{\sum_{n=0}^{N-1} f(T^n s)}{N} = \int f d\mu.$$

Notice that the above condition need only be checked on the characteristic functions of word cylinders (this is because characteristic functions of cylinders are dense in $C(\Sigma^{\mathbb{N}})$, since they form an algebra that separates points). Then it is immediate that if $X = X_k$, the full k-shift for some integer $k > 1$, and μ is the uniform or Lebesgue measure on X with $\mu(i) = k^{-1}$ for $i \in \Sigma$, then s is μ-distributed if and only if the real number $\sum_{j>0} s_j k^{-j}$ is normal in base k.

There is a notion of entropy for dynamical systems called *metric entropy* or *Kolmogorov-Sinai entropy*, which is a natural extension of the Shannon entropy, and which assigns an entropy value $h_\mu(X)$ to any invariant measure μ^* on a system X. A measure μ^* has maximal entropy if $h_{\mu^*}(X) \geq h_\mu(X)$ for all invariant measures μ on X. An important result concerning invariant measures for Markov shifts is the following, due to Parry [13]:[1]

[1] An earlier and independent proof, in a somewhat different language, was already formulated in [18].

Theorem 2.8. Given an irreducible Markov shift (X, T) with a grammar of wordlength $k-1$, there is a unique invariant probability measure $\tilde{P}$ on X of maximal metric entropy. Moreover, this measure is k-step Markov.

3 P-martingales and P-distributed sequences

In the algorithmic randomness literature, given a martingale f on Σ^*, one often constructs a (semi)measure $\mu_f(\sigma) = f(\sigma)|\Sigma|^{-|\sigma|}$, which may be alternatively written as

$$\mu_f(\sigma) = f(\sigma)\lambda([\sigma]), \tag{1}$$

where λ is the Lebesgue or uniform measure on $\Sigma^{\mathbb{N}}$, which is taken to be the natural or "fair" measure on sequences of digits.

As was hinted in the introduction and by the mention of Theorem 2.8, we will be interested in measures different from Lebesgue, i.e. we would like to substitute some arbitrary P for λ in the right hand side of (1). This forces us to change the definition of a martingale f, if we still want to make μ_f an additive measure. Given an alphabet Σ and a language $L \subseteq \Sigma^*$, a probability measure P on $\Sigma^{\mathbb{N}}$ is called *L-supported* if $P(\sigma) = 0 \Leftrightarrow \sigma \in \Sigma^* \backslash L$. Equivalently, P has full support on X_L.

Definition 3.1. Given an alphabet Σ, a language $L \subseteq \Sigma^*$ and some L-supported probability measure P on $\Sigma^{\mathbb{N}}$, a *P-supermartingale* on L is a function $f\colon L \to \mathbb{R}$ satisfying

$$f(\sigma) \geq \sum_{\substack{a \in \Sigma \\ \sigma a \in L}} P(\sigma a \mid \sigma) f(\sigma a) \tag{2}$$

for all σ in L. The function f is called a *P-martingale* if the above inequality can be replaced by an equality for all $\sigma \in L$. We say that f *succeeds* on $s \in \Sigma^{\mathbb{N}}$ if $\limsup_N f(s\restriction_N) = \infty$. The ratios $f(\sigma a)/f(\sigma)$ are called *betting factors* of f.

Notice that the conditional probabilities in (2) are always well-defined since P is L-supported and $\sigma \in L$. Of course, when P is λ as in Definition 3.1, the classical definition of a martingale is recovered, since $\lambda(\sigma a \mid \sigma) = \lambda(a) = |\Sigma|^{-1}$. This generalized definition is somewhat more intuitive in the sense that it makes explicit the real-life fact that the odds offered by a bookie at some gamble are the inverse of some implied probability (conditional on the available information) on the outcomes of the gamble. Classical martingales then just capture the case when these probabilities are uniform and independent of previous outcomes.

We define now the notion of P-martingale generated by a deterministic finite automaton (DFA). This is a generalization of the notion of a classical betting strategy generated by a DFA, introduced in [17]. We will write

automata in the usual form $M = \langle Q, \Sigma, \delta, q_0, Q_f \rangle$, where Q is a finite set of states, Σ is the input alphabet, δ is the transition function, q_0 is the initial state and $Q_f \subseteq Q$ is the set of accepting states. Also, we will use the notation δ^* for the natural extension of the transition function δ from symbols to words in Σ.

Definition 3.2. A P-martingale f on a language L is *generated by a DFA* if there is a DFA M accepting L, and a function $b\colon Q \times \Sigma \to \mathbb{R}$ such that

$$f(\sigma a) = b(\delta^*(\sigma, q_0), a) f(\sigma)$$

for any word σ and symbol a such that $\sigma a \in L$.

The main result of this section is that any sequence is distributed according to an irreducible, invariant Markov measure P if an only if no P-martingale generated by a DFA succeeds on it. The rest of the section is devoted to show it. In §3.1 we show the 'if' implication and in §3.2 we show the 'only if' implication. For the case of P being a measure on a sofic shift, we extend the 'if' direction in §3.1.1. This generalization will be needed for §5.

3.1 P-martingales on a DFA can beat sequences that are not P-distributed

Theorem 3.3. Let Σ be an alphabet, (X, T) a subshift of $\Sigma^{\mathbb{N}}$, and let P be a $L(X)$-supported k-step Markov invariant measure on $\Sigma^{\mathbb{N}}$ such that $(p^k_{\sigma,\tau})_{\sigma,\tau\in\Theta_k}$, the Markov transition matrix induced on $\Theta_k^{\mathbb{N}}$, is irreducible. Suppose $s \in X$ is not P-distributed. Then there is a P-martingale generated by a DFA which succeeds on s. Moreover, the only betting factors of this martingale are 1, $(1+\delta)$ and $(1 - \delta p^*/(1-p^*))$, where δ is rational and $p^* = P(\tau\rho \mid \tau)$ or $p^* = 1 - P(\tau\rho \mid \tau)$ for some $\tau, \rho \in \Sigma^*$.

Before proceeding to the proof of the theorem we present some useful notation and auxiliary lemmas. For words $\sigma, \tau \in \Sigma^*$, we let $\mathsf{occ}(\tau, \sigma)$ be the number of occurrences of τ in σ, that is

$$\mathsf{occ}(\tau, \sigma) = |\{i\colon 0 \le i \le |\sigma| - |\tau|, \tau = \sigma_i \dots \sigma_{i+|\tau|-1}\}|.$$

For k an integer, P a measure on $\Sigma^{\mathbb{N}}$, $\sigma \in \Sigma^*$ and $A \subseteq \Sigma^*$ we write

$$\mathsf{Prec}_k(\sigma) = \{\tau \in \Sigma^k\colon\ P(\tau\sigma) > 0\}, \qquad \text{and} \qquad \mathsf{Prec}_k(A) = \bigcup_{\sigma\in A} \mathsf{Prec}_k(\sigma).$$

We will also make use of the following functions M_s and m_s defined on Σ^*

$$M_s(\sigma) = \limsup_{N\to\infty} \frac{\mathsf{occ}(\sigma, s\restriction_N)}{N}, \qquad \text{and} \qquad m_s(\sigma) = \liminf_{N\to\infty} \frac{\mathsf{occ}(\sigma, s\restriction_N)}{N},$$

for some fixed $s \in \Sigma^{\mathbb{N}}$. The subscript s will often be omitted from M_s and m_s when it is understood from context.

For the sake of simplicity, since the step k is fixed, we will write $p_{\sigma,\tau} = p^k_{\sigma,\tau}$. To prove Theorem 3.3 we will first need some auxiliary lemmas. For the rest of this section, the measure P is assumed to satisfy the conditions of Theorem 3.3.

Lemma 3.4. Suppose $s \in X$ is not P-distributed and that $m_s(\tau^*) \neq 0$ for some $\tau^* \in \Theta_k$. Then there is some $\sigma^* \in L(X)$ with $|\sigma^*| \geq k$ and $b \in \Sigma$ such that

$$\frac{\mathsf{occ}(\sigma^* b, s\restriction_N)}{\mathsf{occ}(\sigma^*, s\restriction_N)} \not\to P(\sigma^* b \mid \sigma^*) \tag{3}$$

when $N \to \infty$. Moreover, σ^* can be chosen so that $m_s(\sigma^*) > 0$.

Proof. Let $s \in X$ not be P-distributed, and let $M = M_s$ and $m = m_s$. Let us define

$$r_{\sigma,\tau} = \lim_{N\to\infty} \frac{\mathsf{occ}(\sigma\tau, s\restriction_N)}{\mathsf{occ}(\sigma, s\restriction_N)}$$

for any words $\sigma, \tau \in \Sigma^*$, whenever the limit exists.

The proof follows by contradiction, so let us assume that for all words σ with $|\sigma| \geq k$ and $P(\sigma) > 0$, and any $b \in \Sigma$ we have $r_{\sigma,b} = P(\sigma b \mid \sigma)$. The core of our proof consists in showing that these assumptions imply that M is actually a Markov measure with the same transition matrix as P. This will be carried out through the following Propositions 3.5, 3.6 and 3.7.

Proposition 3.5. For all words σ with $|\sigma| \geq k$, $P(\sigma) > 0$, and any word $\tau = b_1 \dots b_m \in \Sigma^*$ we have $r_{\sigma,\tau} = P(\sigma\tau \mid \sigma)$.

Proof. Given τ, we first take the largest j such that $P(\sigma b_1 \dots b_{j-1}) > 0$. Then, by an iterated use of $r_{\sigma,b} = P(\sigma b \mid \sigma)$ we get

$$\begin{aligned} r_{\sigma,b_1\dots b_j} &= \prod_{i=1}^{j-1} r_{\sigma b_1\dots b_i, b_{i+1}} \\ &= \prod_{i=1}^{j-1} P(\sigma b_1 \dots b_{i+1} \mid \sigma b_1 \dots b_i) = P(\sigma b_1 \dots b_j \mid \sigma). \end{aligned}$$

If $j = m$ we are done. Otherwise, we have

$$0 = P(\sigma b_1 \dots b_{j+1}) = P(\sigma b_1 \dots b_{j+1} \mid \sigma b_1 \dots b_j) = r_{\sigma b_1\dots b_j, b_{j+1}}.$$

Notice that

$$0 = r_{\sigma b_1\dots b_j, b_{j+1}} = \lim_{N\to\infty} \frac{\mathsf{occ}(\sigma b_1 \dots b_{j+1}, s\restriction_N)}{\mathsf{occ}(\sigma b_1 \dots b_j, s\restriction_N)} \geq \limsup_{N\to\infty} \frac{\mathsf{occ}(\sigma b_1 \dots b_m, s\restriction_N)}{\mathsf{occ}(\sigma, s\restriction_N)},$$

and hence $= r_{\sigma,b_1\dots b_m}$ exists and is equal to 0. Then

$$r_{\sigma,b_1\dots b_m} = 0 = P(\sigma b_1 \dots b_{j+1}) \geq P(\sigma b_1 \dots b_m) \geq 0$$

implies $P(\sigma\tau \mid \sigma) = 0$, which finishes our proof of Proposition 3.5. □

Proposition 3.6. For any $\tau \in \Theta_k$ we have

$$M(\tau) = \lim_{N\to\infty} \frac{\mathsf{occ}(\tau, s\restriction_N)}{N}. \tag{4}$$

Proof. Fix some $\tau^* \in \Theta_k$ such that $m(\tau^*) \neq 0$ and let $\tau_1 \dots \tau_\ell$ be an enumeration of all the other words in Θ_k. It should be noted that if we define $\Theta_k^M = \{\tau \in \Sigma^k \mid \ M(\tau) > 0\}$, then from Proposition 3.5 and the fact that P is $L(X)$-supported and $s \in X$ it is easy to deduce that $\Theta_k^M \subseteq \Theta_k$. This fact will be implicit in the following calculations. Then, for any $i \leq \ell$,

$$\begin{aligned}
\limsup_{N\to\infty} \frac{\mathsf{occ}(\tau_i, s\restriction_N)}{\mathsf{occ}(\tau^*, s\restriction_N)} &= \limsup_{N\to\infty} \sum_{\tau\in\Theta_k} \frac{\mathsf{occ}(\tau\tau_i, s\restriction_N)}{\mathsf{occ}(\tau^*, s\restriction_N)} \\
&= p_{\tau^*,\tau_i} + \limsup_{N\to\infty} \sum_{j=1}^{\ell} \frac{\mathsf{occ}(\tau_j\tau_i, s\restriction_N)}{\mathsf{occ}(\tau_j, s\restriction_N)} \frac{\mathsf{occ}(\tau_j, s\restriction_N)}{\mathsf{occ}(\tau^*, s\restriction_N)} \\
&\leq p_{\tau^*,\tau_i} + \sum_{j=1}^{\ell} p_{\tau_j,\tau_i} \limsup_{N\to\infty} \frac{\mathsf{occ}(\tau_j, s\restriction_N)}{\mathsf{occ}(\tau^*, s\restriction_N)}.
\end{aligned} \tag{5}$$

Notice that

$$\begin{aligned}
\limsup_{N\to\infty} \frac{\mathsf{occ}(\tau_i, s\restriction_N)}{\mathsf{occ}(\tau^*, s\restriction_N)} &\leq \limsup_{N\to\infty} \frac{\mathsf{occ}(\tau_i, s\restriction_N)}{N} \left(\liminf_{N\to\infty} \frac{\mathsf{occ}(\tau^*, s\restriction_N)}{N}\right)^{-1} \\
&= \frac{M(\tau_i)}{m(\tau^*)} < \infty.
\end{aligned}$$

Hence can write

$$x_i = \limsup_{N\to\infty} \frac{\mathsf{occ}(\tau_i, s\restriction_N)}{\mathsf{occ}(\tau^*, s\restriction_N)}$$

and $x = (x_1, \dots, x_\ell)$, and reformulate (5) in matrix form as follows

$$(\mathbf{id} - R^*)x \leq p^*, \tag{6}$$

where $\leq$ is the product order on $\mathbb{R}^\ell$, R^* is the transpose of the Markov transition matrix $p_{\sigma,\tau}$ restricted to $\Theta_k \setminus \{\tau^*\}$ and $p^* = (p_{\tau^*,\tau_1}, \dots, p_{\tau^*,\tau_\ell})$.

Similarly, if

$$y_i = \liminf_{N\to\infty} \frac{\mathsf{occ}(\tau_i, s\restriction_N)}{\mathsf{occ}(\tau^*, s\restriction_N)}$$

and $y = (y_1, \dots, y_\ell)$, then the same reasoning used in (5) shows

$$(\mathbf{id} - R^*)y \geq p^*. \tag{7}$$

Let us write $A = (\mathbf{id} - R^*)$ and show that $Ax = Ay$. Equations (6) and (7) imply $Ax \leq Ay$, so it suffices to show that a contradiction follows from assuming $Ax < Ay$. Indeed, $Ax < Ay$ means that, for all i, $\sum_j A_{ij}x_j \leq \sum_j A_{ij}y_j$, where the inequality is strict for some i. This, in turn, implies

$$\sum_j \left(\sum_i A_{ij}\right) x_j = \sum_i \sum_j A_{ij}x_j < \sum_i \sum_j A_{ij}y_j = \sum_j \left(\sum_i A_{ij}\right) y_j,$$

which is impossible since $x_j \geq y_j$ for all j and $\sum_i A_{ij} = 1 - \sum_i p_{\tau_j,\tau_i} \geq 0$. Thus, $Ax = Ay$.

Now, if A were invertible then it would follow that $x = y$, which means $\lim_{N\to\infty} \mathsf{occ}(\tau_i, s\restriction_N)/\mathsf{occ}(\tau^*, s\restriction_N)$ exists for all i, and this in turn implies that (4) is true for τ^*, that is, $\mathsf{occ}(\tau^*, s\restriction_N)/N$ converges (to $M(\tau^*)$, its lim sup), since

$$\sum_i \lim_{N\to\infty} \frac{\mathsf{occ}(\tau_i, s\restriction_N)}{\mathsf{occ}(\tau^*, s\restriction_N)} + 1 = \lim_{N\to\infty} \frac{N}{\mathsf{occ}(\tau^*, s\restriction_N)}$$

and from this convergence for τ^* we derive that of τ_i for all i using

$$\lim_{N\to\infty} \frac{\mathsf{occ}(\tau_i, s\restriction_N)}{\mathsf{occ}(\tau^*, s\restriction_N)} \lim_{N\to\infty} \mathsf{occ}(\tau^*, s\restriction_N) = \lim_{N\to\infty} \frac{\mathsf{occ}(\tau_i, s\restriction_N)}{N}.$$

So it remains to show that A is indeed invertible. If it were not, then R^* would have 1 as an eigenvalue, and the Perron-Frobenius theorem, together with the fact that the column sums of R^* are smaller than 1, imply 1 has a unique nonnegative eigenvector $z = (z_1, \ldots, z_\ell)$. That is,

$$\sum_{i=1}^{\ell} z_i R^*_{ji} = \sum_{i=1}^{\ell} z_i p_{\tau_i,\tau_j} = z_j. \tag{8}$$

Now, $\tau^* \in \Theta_k$ is excluded from the enumeration $(\tau_i)_{1\leq i\leq \ell}$ and the irreducibility of the matrix $p_{\sigma,\tau}$ implies that $\mathsf{Prec}_k(\tau^*) \cap (\Theta_k \setminus \{\tau^*\})$ is not empty. Hence, there is some $\tau_i \in \mathsf{Prec}_k(\tau^*)$ and for each such i we have $\sum_{j=1}^{\ell} p_{\tau_i,\tau_j} < 1$, so that if $z_i \neq 0$ then (8) implies

$$\sum_{j=1}^{\ell} z_j = \sum_{i,j=1}^{\ell} z_i p_{\tau_i,\tau_j} = \sum_{i=1}^{\ell} z_i \sum_{j=1}^{\ell} p_{\tau_i,\tau_j} < \sum_{i=1}^{\ell} z_i,$$

which is a contradiction.

Thus, $z_i = 0$ for all i such that $\tau_i \in \mathsf{Prec}_k(\tau^*)$. This in turn implies $p_{\tau_j,\tau_i} = 0$ for all j such that $z_j \neq 0$, since $0 = z_i = \sum_j z_j p_{\tau_j,\tau_i}$. Equivalently, $z_j = 0$ for all j such that $\tau_j \in \mathsf{Prec}_k(\tau^i)$.

We then repeat this reasoning to show $z_k = 0$ for all k such that $\tau_k \in \mathsf{Prec}_k(\tau^j)$ and keep repeating the same reasoning until all entries in

z have been shown to be 0 (this is guaranteed by irreducibility). Hence, $z = 0$, which contradicts the assumption that z is an eigenvector of eigenvalue 1. It follows that A must be invertible. This concludes the proof of Proposition 3.6. □

Proposition 3.7. *M is equal to P restricted to Θ_k.*

Proof. Now, M is a probability measure on Θ_k, since

$$\sum_{\tau\in\Theta_k} M(\tau) = \sum_{\tau\in\Theta_k} \lim_{N\to\infty} \frac{\mathsf{occ}(\tau, s\restriction_N)}{N} = \lim_{N\to\infty} \frac{\sum_{\tau\in\Theta_k} \mathsf{occ}(\tau, s\restriction_N)}{N} = 1.$$

Together with the Markov transition matrix $p_{\sigma,\tau}$, M defines a probability measure ν on $\Theta_k^{\mathbb{N}}$ in a natural way. First, ν is defined inductively on word cylinders

$$\begin{aligned} \nu([\tau]) &= M(\tau) \\ \nu([\tau_1 \dots \tau_j]) &= \nu([\tau_1 \dots \tau_{j-1}])p_{\tau_{j-1},\tau_j} \end{aligned}$$

then extended naturally to all cylinders and finally to the Borel σ-algebra $\mathcal{B}$ of $\Theta_k^{\mathbb{N}}$ via Caratheodory's extension theorem.

As with P, we will drop the brackets for word cylinders. To show that ν is invariant, it is enough to show it for word cylinders, that is, it is enough to show

$$\nu(T^{-1}(\tau_1 \dots \tau_j)) = \sum_{\sigma\in\Theta_k} \nu(\sigma\tau_1 \dots \tau_j) = \nu(\tau_1 \dots \tau_j).$$

By our construction of ν,

$$\begin{aligned} \sum_{\sigma\in\Theta_k} \nu(\sigma\tau_1 \dots \tau_j) &= \sum_{\sigma\in\Theta_k} \nu(\sigma)\nu(\tau_1 \mid \sigma) \prod_{i=1}^{j-1} \nu(\tau_{i+1} \mid \tau_i) \\ &= \sum_{\sigma\in\Theta_k} M(\sigma)p_{\sigma,\tau_1} \prod_{i=1}^{j-1} p_{\tau_i,\tau_{i+1}} = \left(\prod_{i=1}^{j-1} p_{\tau_i,\tau_{i+1}}\right) \sum_{\sigma\in\Theta_k} M(\sigma)r_{\sigma,\tau_1} \\ &= \left(\prod_{i=1}^{j-1} p_{\tau_i,\tau_{i+1}}\right) \sum_{\sigma\in\Theta_k} \lim_{N\to\infty} \frac{\mathsf{occ}(\sigma, s\restriction_N)}{N} \lim_{N\to\infty} \frac{\mathsf{occ}(\sigma\tau_1, s\restriction_N)}{\mathsf{occ}(\sigma, s\restriction_N)} \\ &= \left(\prod_{i=1}^{j-1} p_{\tau_i,\tau_{i+1}}\right) \sum_{\sigma\in\Theta_k} \lim_{N\to\infty} \frac{\mathsf{occ}(\sigma\tau_1, s\restriction_N)}{N} \\ &= \left(\prod_{i=1}^{j-1} p_{\tau_i,\tau_{i+1}}\right) \lim_{N\to\infty} \frac{\mathsf{occ}(\tau_1, s\restriction_N)}{N} \\ &= \left(\prod_{i=1}^{j-1} p_{\tau_i,\tau_{i+1}}\right) M(\tau_1) = \nu(\tau_1 \dots \tau_j). \end{aligned} \tag{9}$$

Thus, ν is invariant, 1-step Markov and has the irreducible Markov transition matrix $p_{\sigma,\tau}$. Theorem 2.4 implies that $P = \nu$ and M is equal to P restricted to Θ_k, and this concludes the proof of Proposition 3.7. □

Finally, we will now show that s is P-distributed, leading to a contradiction. In (9) we show that

$$\nu(\sigma\tau) = M(\sigma)p_{\sigma,\tau} = P(\sigma)r_{\sigma,\tau} = \lim_{N\to\infty} \frac{\mathsf{occ}(\sigma\tau, s\restriction_N)}{N}$$

for $\tau \in \Theta_k$. This extends trivially to $\tau \in \Sigma^k$, for $M(\tau) = 0$ if and only if $P(\tau) = 0$ and $P = \nu$. Moreover, the same is valid if we substitute any word ρ for τ in the above equations, since all we need is that $r_{\sigma,\rho}$ exist and be equal to $P(\sigma\rho \mid \sigma)$. Since σ must be of length k, this means that

$$\lim_{N\to\infty} \frac{\mathsf{occ}(\rho, s\restriction_N)}{N} = \nu(\rho) = P(\rho) \tag{10}$$

for all words ρ of length at least k. But then (10) must also be true for words ρ of length smaller than k, since

$$\begin{aligned} P(\rho) &= \sum_{\substack{\tau\in\Theta_k \\ \rho\prec\tau}} P(\tau) = \sum_{\substack{\tau\in\Theta_k \\ \rho\prec\tau}} \lim_{N\to\infty} \frac{\mathsf{occ}(\tau, s\restriction_N)}{N} \\ &= \lim_{N\to\infty} \frac{\sum_{\substack{\tau\in\Theta_k \\ \rho\prec\tau}} \mathsf{occ}(\tau, s\restriction_N)}{N} = \lim_{N\to\infty} \frac{\mathsf{occ}(\rho, s\restriction_N)}{N}. \end{aligned}$$

P-distribution need only be checked on word cylinders, so this completes the proof that some σ^* satisfies (3).

It only remains to show that such a σ^* can be chosen so that $m_s(\sigma^*) > 0$. Again, we prove this by contradiction. That is, let us suppose that for all $\sigma \in L(X)$ ($|\sigma| \geq k$) such that $m(\sigma) > 0$, we have that $r_{\sigma,b}$ exists for any symbol b and is equal to $P(\sigma b \mid \sigma)$. As before, this implies $r_{\sigma,\tau}$ exists for all words τ and is equal to $P(\sigma\tau \mid \sigma)$.

Take some σ^* that satisfies (3). Then $m(\sigma^*) = 0$. Take some τ such that $P(\tau\sigma^*) > 0$ (irreducibility implies this can be done by finding some $(\tau_i)_{1\leq i\leq l} \subseteq \Theta_k$ such that $\sigma^* \prec \tau_1 \dots \tau_l \in L(X)$ and then finding some $\tau \in \mathsf{Prec}_k(\tau_i)$). If $m(\tau) > 0$ then r_{τ,σ^*} exists and is equal to $P(\tau\sigma^* \mid \tau) > 0$. But this contradicts the fact that $m(\tau\sigma^*) \leq m(\sigma^*) = 0$. So $m(\tau) = 0$ for all $\tau \in \mathsf{Prec}_k(\sigma^*)$.

Similarly, for all $\sigma \in \mathsf{Prec}_k(\mathsf{Prec}_k(\sigma^*))$ we have $m(\sigma) = 0$ and the same reasoning can be repeated until $m(\sigma) = 0$ has been shown for all $\sigma \in \Theta$ (irreducibility guarantees this), which contradicts the condition that $m(\tau^*) > 0$ for some $\tau^* \in \Theta_k$. This concludes the proof of Lemma 3.4. □

Lemma 3.8. Given $s \in X$, if there is some $\sigma^* \in \Theta_k$ that satisfies $m(\sigma^*) = 0$, then there are some $d > 0$, $\rho \in \Theta_k$, $\sigma \in \mathsf{Prec}_k(\rho)$ and a strictly increasing sequence $(N_j)_{j\in\mathbb{N}}$ of natural numbers such that $\lim_{j\to\infty} \mathsf{occ}(\rho, s\restriction_{N_j})/N_j = 0$ and $\limsup_{j\to\infty} \sum_{\tau\in\Theta_k\setminus\{\rho\}} \mathsf{occ}(\sigma\tau, s\restriction_{N_j})/N_j \geq d$.

Proof. Let $\sigma^* \in \Theta_k$ satisfy $m(\sigma^*) = 0$ and let $(N_j)_{j\in\mathbb{N}}$ be a strictly increasing sequence of natural numbers such that $\lim_{j\to\infty} \mathsf{occ}(\sigma^*, s\restriction_{N_j})/N_j = 0$. If for some $\sigma \in \mathsf{Prec}_k(\sigma^*)$ and some $d > 0$ we have $\limsup_{j\to\infty} \sum_{\tau\in\Theta_k\setminus\{\sigma^*\}} \mathsf{occ}(\sigma\tau, s\restriction_{N_j})/N_j \geq d$, then we set $\rho = \sigma^*$ and we are done.

Otherwise, we have, for $\epsilon > 0$, a j_0 such that for all $j \geq j_0$,

$$\sum_{\substack{\sigma\in\mathsf{Prec}_k(\sigma^*)\\ \tau\in\Theta_k\setminus\{\sigma^*\}}} \frac{\mathsf{occ}(\sigma\tau, s\restriction_{N_j})}{N_j} < \frac{\epsilon}{|\Theta_k|},$$

and since $\lim_{j\to\infty} \mathsf{occ}(\sigma\sigma^*, s\restriction_{N_j})/N_j \leq \lim_{j\to\infty} \mathsf{occ}(\sigma^*, s\restriction_{N_j})/N_j = 0$, we conclude, for all j greater than some j_0,

$$\frac{\sum_{\sigma\in\mathsf{Prec}_k(\sigma^*)} \mathsf{occ}(\sigma, s\restriction_{N_j}) + \mathsf{occ}(\sigma^*, s\restriction_{N_j})}{N_j} < 2\epsilon.$$

Hence, if we write $B_0 = \{\sigma^*\}$, $B_{t+1} = B_t \cup \mathsf{Prec}_k(B_t)$ and, for any finite $A \subseteq \Sigma^*$

$$\mathsf{occ}(A) = \limsup_{j\to\infty} \frac{\sum_{\sigma\in A} \mathsf{occ}(\sigma, s\restriction_{N_j})}{N_j},$$

then we have just shown that $\mathsf{occ}(B_1) = \mathsf{occ}(\mathsf{Prec}_k(\sigma^*)) = 0$.

Similarly, given t such that $\mathsf{occ}(B_t) = 0$ we can use the same reasoning to show that either $\limsup_{j\to\infty} \sum_{\tau\in\Theta_k\setminus\{\rho\}} \mathsf{occ}(\sigma\tau, s\restriction_{N_j})/N_j \geq d$ for some $d > 0$, $\rho \in B_t$ and some $\sigma \in \mathsf{Prec}_k(\rho) \subseteq B_{t+1}$, in which case we are done, or else $\mathsf{occ}(B_{t+1}) = 0$.

But irreducibility implies that, for some p, $B_p = \Theta_k$ and we cannot have $0 = \mathsf{occ}(B_p) = \mathsf{occ}(\Theta_k) = 1$. Hence, there is some t and some $d > 0$, $\rho \in B_t$ and $\sigma \in \mathsf{Prec}_k(\rho)$ such that $\lim_{j\to\infty} \mathsf{occ}(\rho, s\restriction_{N_j})/N_j \leq \lim_{j\to\infty} \sum_{\sigma\in B_t} \mathsf{occ}(\sigma, s\restriction_{N_j})/N_j$ and

$$\limsup_{j\to\infty} \sum_{\tau\in\Theta_k\setminus\{\rho\}} \mathsf{occ}(\sigma\tau, s\restriction_{N_j})/N_j \geq d.$$

This concludes the proof of Lemma 3.8. □

Proof of Theorem 3.3. We will split our proof in two cases.

Case I. *There is some $\tau^* \in \Theta_k$ such that $m(\tau^*) > 0$.*

From Lemma 3.4 we may assume that for some $\sigma \in L(X)$ satisfying $|\sigma| \geq k$ and $m(\sigma) > 0$, some $b \in \Sigma$ and some rational $\delta > 0$

$$\limsup_{N\to\infty} \frac{\mathsf{occ}(\sigma b, s\restriction_N)}{\mathsf{occ}(\sigma, s\restriction_N)} > (1+\delta)P(\sigma b \mid \sigma), \tag{11}$$

since (3) implies either (11) or

$$\liminf_{N\to\infty} \frac{\mathsf{occ}(\sigma b, s\restriction_N)}{\mathsf{occ}(\sigma, s\restriction_N)} < (1-\delta)P(\sigma b \mid \sigma),$$

but in the latter case it is easy to find some b' such that (11) is true for $\sigma b'$.

We define our P-martingale L by:

$$L(\emptyset) = 1$$

$$L(\rho c) = \begin{cases} (1+\delta)L(\rho) & \text{if } \rho\restriction_{|\sigma|} = \sigma \text{ and } c = b; \\ \left(1 - \frac{\delta p^*}{1-p^*}\right) L(\rho) & \text{if } \rho\restriction_{|\sigma|} = \sigma \text{ and } c \neq b; \\ L(\rho) & \text{otherwise.} \end{cases}$$

for any $c \in \Sigma$, $\rho c \in L(X)$, where $p^* = P(\sigma b|\sigma)$ and we further impose that $\delta < (1-p^*)/p^*$. Notice that for all ρ such that $\rho\restriction_{|\sigma|} = \sigma$ the k-step Markov property and $|\sigma| \geq k$ impliy that $p^* = P(\rho b|\rho)$. From this it is easy to see that L is a P-martingale, and it is also clearly generated by a DFA, since at each step the betting factor depends solely on the next symbol and the previous $|\sigma|$ symbols of ρ and since there are finitely many words of length σ, it suffices to consider the finite set of states $Q = |\Sigma|^{|\sigma|}$.

To see that L succeeds on s, we observe first that

$$L(\rho) = (1+\delta)^{\mathsf{occ}(\sigma b, \rho)} \prod_{c\neq b} \left(1 - \frac{\delta p^*}{1-p^*}\right)^{\mathsf{occ}(\sigma c, \rho)}$$

and that

$$\sum_{c\neq b} \mathsf{occ}(\sigma c, s\restriction_N) \leq \mathsf{occ}(\sigma, s\restriction_N) - \mathsf{occ}(\sigma b, s\restriction_N). \tag{12}$$

Let $r = 1+\delta$ and $q = 1 - \frac{\delta p^*}{1-p^*}$. Equation (12) then implies

$$\begin{aligned}
&\limsup_{N\to\infty} \frac{\log L(s\restriction_N)}{N} \\
&\quad \geq \limsup_{N\to\infty} \frac{\mathsf{occ}(\sigma b, s\restriction_N)}{N}(\log r - \log q) + \frac{\mathsf{occ}(\sigma, s\restriction_N)}{N}\log q \\
&\quad = \limsup_{N\to\infty} \frac{\mathsf{occ}(\sigma, s\restriction_N)}{N}\left[\frac{\mathsf{occ}(\sigma b, s\restriction_N)}{\mathsf{occ}(\sigma, s\restriction_N)}(\log r - \log q) + \log q\right] \\
&\quad \geq \liminf_{N\to\infty} \frac{\mathsf{occ}(\sigma, s\restriction_N)}{N} \limsup_{N\to\infty}\left[r\frac{\mathsf{occ}(\sigma b, s\restriction_N)}{\mathsf{occ}(\sigma, s\restriction_N)}(\log r - \log q) + \log q\right] \\
&\quad \geq m(\sigma)\left[rP(\sigma b \mid \sigma)(\log r - \log q) + \log q\right] \\
&\quad = p^* m(\sigma)\left[(1+\delta)\log(1+\sigma) + \left(p^{*-1} - (1+\sigma)\right)\log\left(1 - \frac{\delta p^*}{1-p^*}\right)\right].
\end{aligned} \tag{13}$$

Observe that $p^* > 0$, for otherwise $\sigma b \notin L(X)$, $\mathsf{occ}(\sigma b, s\restriction_N) = 0$ for all N and the inequality in (11) would not be obtained. Hence, the multiplying factor on the left is strictly positive. Now if in (13) we make the substitution $x = p^{*-1} - 1$ we may notice that the function $f(\delta) = (1+\delta)\log(1+\delta) + (x-\delta)\log(1-\delta/x)$ satisfies $f(0) = 0$ and $f'(\delta) = \log(1+\delta) - \log(1-\delta/x) > 0$ for $0 < \delta < x$. Then there is a $c > 0$ such that $\limsup_{N\to\infty} \log L(s\restriction_N)/N \geq c$. and there will be infinitely many N's such that $L(s\restriction_N) \geq 2^{cN}$, which implies $\limsup_{N\to\infty} L(s\restriction_N) = \infty$.

Case II. *For all $\tau \in \Theta_k$ we have $m(\tau) = 0$.*

Lemma 3.8 implies that there are some $d > 0$, $\rho \in \Theta_k$, $\sigma \in \mathsf{Prec}_k(\rho)$ and a strictly increasing sequence of natural numbers $(N_j)_{j\in\mathbb{N}}$ such that

$$\lim_{j\to\infty} \frac{\mathsf{occ}(\rho, s\restriction_{N_j})}{N_j} = 0 \qquad \text{and} \qquad \limsup_{j\to\infty} \frac{\sum_{\tau\in\Theta_k\setminus\{\rho\}} \mathsf{occ}(\sigma\tau, s\restriction_{N_j})}{N_j} \geq d. \tag{14}$$

Notice that $\langle s : k\rangle$ is actually a sequence in Θ_k (and not just Σ^k) since $s \in X$ and all words of length k in s must belong to $L(X)$. Also, as mentioned in §2, P induces an irreducible Markov measure P^k on $\Theta_k^{\mathbb{N}}$, so we will first construct a P^k-martingale on Θ_k^*. Let $p^* = 1 - P(\sigma\rho|\sigma) < 1$ (since $\sigma \in \mathsf{Prec}_k(\rho)$), $(1-p^*) > p^*\delta$, c be a symbol of Θ_k and $\delta > 0$. We define M as follows:

$$\begin{aligned}
M(\emptyset) &= 1; \\
M(\tau_1\ldots\tau_{l+1}) &= \begin{cases} (1+\delta)M(\tau_1\ldots\tau_l) & \text{if } \tau_l = \sigma \text{ and } \tau_{l+1} \neq \rho; \\ \left(1 - \frac{\delta p^*}{1-p^*}\right) M(\tau_1\ldots\tau_l) & \text{if } \tau_l = \sigma \text{ and } \tau_{l+1} = \rho; \\ M(\tau_1\ldots\tau_l) & \text{otherwise.} \end{cases}
\end{aligned}$$

It is easy to check that M is a P^k-martingale generated by a DFA.

Observe that

$$M(\tau_1 \dots \tau_l) = \left(1 - \frac{\delta p^*}{1-p^*}\right)^{\mathsf{occ}(\sigma\rho, \tau_1 \dots \tau_l)} \prod_{\substack{c \in \Theta_k \\ c \neq \rho}} (1+\delta)^{\mathsf{occ}(\sigma c, \tau_1 \dots \tau_l)}.$$

Write $q = 1 - \delta p^*/(1-p^*)$ and $r = 1+\delta$ and fix $\epsilon > 0$ such that $\epsilon < d \log r$. Then (14) implies there are infinitely many N's such that

$$\log r \sum_{\substack{\tau \in \Theta_k \\ \tau \neq \rho}} \frac{\mathsf{occ}(\sigma\tau, \langle s : k \rangle \restriction_N)}{N} + \log q \frac{\mathsf{occ}(\sigma\rho, \langle s : k \rangle \restriction_N)}{N} \geq d \log r - \epsilon.$$

Thus, for $K = d \log r - \epsilon > 0$ we have $\log M(\langle s : k \rangle \restriction_N)/N \geq K > 0$ for infinitely many N's. This implies the martingale succeeds on $\langle s : k \rangle$.

From this martingale M on Θ_k^* one uses the definition of a P-martingale to extend M to a P-martingale $\widehat{M}$ on Σ^* (it is a routine exercise to check $\widehat{M}$ is well defined as a P-martingale and that it is also generated by a DFA), and the fact that M succeeds on $\langle s : k \rangle$ implies that $\widehat{M}$ succeeds on s. □

3.1.1 An extension to sofic shifts

We now extend Theorem 3.3 to a more general class of measures on sofic subshifts. In order to do so, we need some of the standard results and definitions regarding sofic subshifts.

A *labelled directed graph* on alphabet Σ is a tuple $(G, \mathcal{L})$ where G is a directed graph with finite nodes $\mathcal{N}(G)$ and finite edges $\mathcal{E}(G)$ and $\mathcal{L}$ is a function assigning to each edge e in $\mathcal{E}(G)$ a symbol $\mathcal{L}(e) \in \Sigma$.

Given a labelled directed graph $(G, \mathcal{L})$, a *path on* $(G, \mathcal{L})$ through states $i_0, \dots, i_l \in \mathcal{N}(G)$ is a finite sequence of symbols $a_1 \dots a_l \in \Sigma^*$ for which there are edges $e_1, \dots, e_l \in \mathcal{E}(G)$ such that, for all $1 \leq j \leq l$, i_j is the destination node of e_j, i_{j-1} is the origin node of e_j and $\mathcal{L}(e_j) = a_j$. The set of paths on $(G, \mathcal{L})$ is denoted $\mathcal{P}_G$.

Notice that any labelled directed graph is equivalent to an automaton M_G on Σ without an initial state and with a single absorbing non-accepting state. The accepting states of M_G are given by the nodes of G, and a transition $\delta(i, a) = j$ whenever there is an edge between i and j labelled a. The following definition is then equivalent to this automaton being deterministic.

A labelled directed graph G on alphabet Σ is called *right-resolving* if for any symbol $a \in \Sigma$, and any node $i \in \mathcal{N}(G)$ there is at most one edge $e \in \mathcal{E}(G)$ such that e has i as its origin node and $\mathcal{L}(e) = a$.

Labelled graphs may be used to represent sofic subshifts in the following way:

Definition 3.9. A *labelled graph presentation* of a sofic subshift (X, T) on alphabet Σ is a labelled directed graph $(G, \mathcal{L})$ such that $L(X) = \mathcal{P}_G$.

Notice that, given a directed graph G there is a natural 1-step Markov shift (X_G, T) consisting of the admissible sequences of edges. Furthermore, when a labelling function $\mathcal{L}$ is defined on the edges of G, the one-block code that maps e to $\mathcal{L}(e)$ induces a factor map $\mathcal{L}^*$ from the Markov shift (X_G, T) to the sofic subshift represented by $(G, \mathcal{L})$.

A labelled graph presentation $(G, \mathcal{L})$ of a sofic subshift (X, T) on alphabet Σ is called *minimal* if there is no other presentation $(G', \mathcal{L}')$ of X with strictly fewer nodes, and it is called *irreducible* if the underlying directed graph is strongly connected. We say that a sofic subshift (X, T) is *irreducible* if for any words $\sigma, \tau \in L(X)$ there is a word ρ such that $\sigma\rho\tau \in L(X)$.

Theorem 3.10. ([9, Theorem 3.3.2]) Any irreducible sofic subshift has a unique (up to graph isomorphism) minimal, irreducible, right-resolving graph presentation.

Given a labelled directed graph $(G, \mathcal{L})$ on alphabet Σ, a *synchronizing word* for G is a word $\alpha = a_1 \dots a_l \in \Sigma^*$ for which the set

$$\{i_l \in \mathcal{N}(G) \mid (\exists i_0, \dots, i_{l-1} \in \mathcal{N}(G))\ \alpha \text{ is a path through } i_0, \dots, i_l\}$$

has a single node. We call that node the *synchronizing node* of α. That is, when regarding the graph as the equivalent automaton M_G, a synchronizing word is one that leaves the automaton in one and only one state after being read, regardless of the state on which its reading began. Finally, we have [9, Proposition 3.3.9 and Proposition 3.3.16]:

Theorem 3.11. A minimal, right-resolving labelled graph presentation of a sofic subshift (X, T) has a synchronizing word.

We are now ready to prove the extension of our previous result.

Theorem 3.12. Let X be an irreducible sofic subshift on alphabet Σ and let $(G, \mathcal{L})$ be its minimal, irreducible, right-resolving presentation. Let (X_G, T) be the Markov shift of edge sequences associated to G and let P be an irreducible, invariant, $L(X_G)$-supported, 1-step Markov measure on $\mathcal{E}(G)^{\mathbb{N}}$. Let $\mathcal{L}^*\colon X_G \to X$ be the natural factor map induced by $\mathcal{L}$ and let $\nu = P \circ \mathcal{L}^{*-1}$ be the pushforward measure on $\Sigma^{\mathbb{N}}$. Let $s \in X$ be not ν-distributed.

1. If a synchronizing word appears as a factor of s then there is a ν-martingale generated by a DFA which succeeds on s. Moreover, the only betting factors of this martingale are 1, $(1+\delta)$ and $(1-\delta p^*/(1-p^*))$, where δ is rational and $p^* = P(\tau\rho \mid \tau)$ or $p^* = 1 - P(\tau\rho \mid \tau)$ for some $\tau, \rho \in \Sigma^*$.

2. If no synchronizing word appears as a factor of s then there is a ν-supermartingale generated by a DFA which succeeds on s. Moreover, the only betting factors of this martingale are 1, $(1-\delta^*)$ and $(1+\delta)$, where δ^* and δ are rational.

Proof of item 1 of Theorem 3.12. Since any leftward extension in $L(X)$ of a synchronizing word is a synchronizing word, we may assume that some prefix of s, say $\rho = s\restriction_{N'}$ is a synchronizing word with synchronizing node i_0. Denote by $f(i,a)$ the unique edge whose origin node is i and whose label is a, whenever it exists (uniqueness is guaranteed because the presentation is right-resolvable). We construct by induction a sequence z in $\mathcal{E}(G)^{\mathbb{N}}$ that records the sequence of edges followed by s after it reaches the synchronizing node:

- Define $z_1 = f(i_0, s_{N'+1})$ and let i_1 be the destination node of z_1.
- Assume z_n and i_n are defined. Define $z_{n+1} = f(i_n, s_{N'+n})$ and let i_{n+1} be the destination node of z_{n+1} (notice that $f(i_n, s_{N'+n})$ must exist because $s \in X$ and $(G,\mathcal{L})$ is a presentation of X)

Notice that by construction $\mathcal{L}^*(z) = T^{N'}(s)$. We will use this fact to show by contradiction that z is not P-distributed, and then use the martingale on a DFA that succeeds on z (guaranteed by Theorem 3.3) to build an appropriate martingale on a DFA that succeeds on s.

For any word $a_1 \dots a_n \in \Sigma^*$ let $A(a_1 \dots a_n)$ be the set of words $e_1 \dots e_n \in \mathcal{E}(G)^*$ such that $\mathcal{L}(e_i) = a_i$ for $1 \le i \le n$ (equivalently, $A(\alpha) = \mathcal{L}^{*-1}(\alpha)$). Clearly, we have $\nu(\alpha) = \sum_{\sigma \in A(\alpha)} P(\sigma)$ and, by construction of z,

$$\mathsf{occ}(\alpha, T^{N'}(s)\restriction_N) = \sum_{\sigma \in A(\alpha)} \mathsf{occ}(\sigma, z\restriction_N).$$

Then, if z is P-distributed, $T^{N'}(s)$ must be ν-distributed and $T^{N'}(s)$ is ν-distributed if and only if s is so. Hence, z cannot be P-distributed. By Theorem 3.3 there is a martingale $\widehat{L}$ generated by a DFA $\widehat{M} = \langle \widehat{Q}, \mathcal{E}(G), \hat{\delta}, \hat{q}_0, \hat{Q}_f \rangle$ and a function $\hat{b}\colon \hat{Q} \times \mathcal{E}(G) \to \mathbb{R}$, and such that $\limsup_N \hat{L}(s\restriction_N) = \infty$.

Given an edge e, we write $d(e)$ for its destination node and $o(e)$ for its origin node. Let M' be a DFA on Σ having $Q' = (\hat{Q} \times \mathcal{N}(G)) \cup \{q_g\}$ (for some unused garbage state q_g) as its set of states, $Q'_f = \hat{Q}_f \times \mathcal{N}(G)$ as its set of final states, $q'_0 = (\hat{q}_0, i_0)$ as its initial state and a transition function δ' defined by:

$$\delta'((q,i),a) = \begin{cases} (\hat{\delta}(q, f(i,a)), d(f(i,a))) & \text{if } f(i,a) \text{ exists;} \\ q_g & \text{otherwise.} \end{cases}$$
$$\delta'(q_g, a) = q_g \quad \text{for all } a.$$

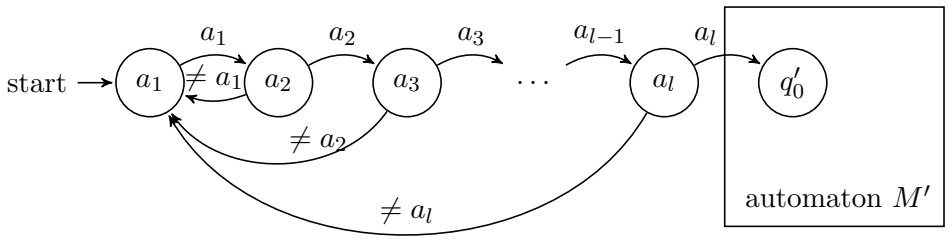

Figure 1: The automaton M. $\rho = a_1 \dots a_l$ is a synchronizing word, and M' "translates" $\widehat{M}$, which has inputs on the language of edges, $\mathcal{E}(G)^*$ to the language $L(X) \subseteq \Sigma^*$.

We also define $b' \colon Q' \times \Sigma \to \mathbb{R}$ as

$$b'((q,i),a) = \hat{b}(q, f(i,a))$$
$$b'(q_g, a) = 1.$$

Notice that, by construction, the function f defined by $f(\lambda) = 1$ and $f(\alpha a) = b'(\delta'^*(q'_0, \alpha), a) f(\alpha)$ satisfies

$$\limsup_N f(T^{N'}(s) \restriction_N) = \infty. \tag{15}$$

since the sequence of betting factors induced by $T^{N'}(s)$ for M' and b' is the same as that induced by z for $\hat{M}$ and $\hat{b}$.

Finally, write $\rho = a_1 \dots a_l$ and define the DFA $M = \langle Q, \Sigma, \delta, a_1, Q'_f \rangle$, where $Q = Q' \cup \{a_1, \dots, a_l\}$ and

$$\delta(q,a) = \begin{cases} \delta'(q,a) & \text{if } q \in Q'; \\ a_2 & \text{if } a = q = a_1; \\ a_{i+1} & \text{if } a = q = a_i, \text{ for } 2 \le i \le l-1; \\ q'_0 & \text{if } a = q = a_l; \\ a_1 & \text{otherwise.} \end{cases}$$

This automaton waits until the synchronizing word ρ is read. Once it finishes reading it the automaton transitions to the automaton M' and stays there (see Figure 1). The function $b \colon Q \times \Sigma \to \mathbb{R}$, computing the betting factors, must then be

$$b(q,a) = \begin{cases} 1 & \text{if } q \notin Q'; \\ b'(q,a) & \text{otherwise.} \end{cases}$$

From (15) and the construction of f, M and b, it follows that the function L defined by

$$L(\lambda) = 1$$
$$L(\alpha a) = b(\delta^*(a_1, \alpha), a) L(\alpha) \quad \text{when } \alpha a \in L(X)$$

satisfies $\limsup_N L(s\restriction_N) = \infty$. Also, by construction, L has the same betting factors as $\hat{L}$, which are 1, $(1+\delta)$ and $(1-\delta p^*/(1-p^*))$.

It remains to show that L is a ν-martingale, i.e. that $L(\alpha) = \sum_{a\in\Sigma} \nu(\alpha a \mid \alpha)L(\alpha a)$. By definition of L and b, this condition is trivially satisfied when ρ is not a prefix of α, since in that case $L(\alpha a) = L(\alpha) = 1$. Hence, we only need to show

$$1 = \sum_{a\in\Sigma} \nu(\alpha a \mid \alpha) b(\delta^*(a_1, \alpha), a) \tag{16}$$

when ρ is a prefix of α, say, $\rho = \alpha\restriction_{N_0}$.

Observe that

$$\nu(\alpha a \mid \alpha) = \frac{\sum_{\sigma\in A(\alpha)} P(\sigma f(d(\sigma_{|\sigma|}), a))}{\sum_{\sigma\in A(\alpha)} P(\sigma)}$$

and we may define $e = \sigma_{|\sigma|}$ and $h = f(d(e), a)$, independently of σ, since σ is the path of edges followed by α and α has a prefix that is a synchronizing word, so that σ_n is the same for all $\sigma \in A(\alpha)$ when $n \geq N_0$. Thus, the fact that P is 1-step Markov implies

$$\begin{aligned}\nu(\alpha a \mid \alpha) &= \frac{\sum_{\sigma\in A(\alpha)} P(\sigma h \mid \sigma)P(\sigma)}{\sum_{\sigma\in A(\alpha)} P(\sigma)} \\ &= P(eh \mid e)\frac{\sum_{\sigma\in A(\alpha)} P(\sigma)}{\sum_{\sigma\in A(\alpha)} P(\sigma)} = P(eh \mid e) = P(\sigma h \mid \sigma),\end{aligned}$$

and writing $\eta = \sigma\restriction_{|\alpha|-N_0}$ (which is the same for all $\sigma \in A(\alpha)$ by the preceding remark) and $\tau = \mathcal{L}^*(\eta)$ we have that (16) boils down to

$$\begin{aligned}1 &= \sum_{h:o(h)=d(e)} P(\sigma h \mid \sigma) b(\delta^*(a_1, \alpha), a) \\ &= \sum_{h:o(h)=d(e)} P(\eta h \mid \eta) b'(\delta'^*(q'_0, \tau), a) \\ &= \sum_{h:o(h)=d(e)} P(\eta h \mid \eta) \hat{b}(\hat{\delta}^*(\hat{q}_0, \eta), f(d(e), a)) \\ &= \sum_{h:o(h)=d(e)} P(\eta h \mid \eta) \hat{b}(\hat{\delta}^*(\hat{q}_0, \eta), h)\frac{\hat{L}(\eta)}{\hat{L}(\eta)} \\ &= \sum_{h:o(h)=d(e)} P(\eta h \mid \eta)\frac{\hat{L}(\eta h)}{\hat{L}(\eta)}.\end{aligned}$$

This last condition is met because $\hat{L}$ is a P-martingale. □

When no synchronizing word appears in s, the conditional probabilities that appear in the martingale condition may have infinitely many possible values, so that they will not be computed by a martingale generated by a DFA. Yet, we can still find a supermartingale on a DFA to handle this case.

Proof of item 2 Theorem 3.12. As before, define $A(\alpha) = \mathcal{L}^{*-1}(\alpha)$. By Theorem 3.11 $(G, \mathcal{L})$ has a synchronizing word α, and $\alpha \in X$. Hence, $A(\alpha)$ is not empty, and since P is $L(X_G)$-supported we have $P(\sigma) > 0$ for any $\sigma \in A(\alpha)$. Therefore, $\nu(\alpha) = \sum_{\sigma \in A(\alpha)} P(\sigma) > 0$.

By hypothesis, α is not a factor of s, hence $\mathsf{occ}(\alpha, s \restriction_N) = 0$ for all N. Let $N_\alpha = \max\{N \colon \alpha \restriction_N \text{ appears infinitely many times in } s\} \cup \{0\}$ and write $\eta bc = \alpha \restriction_{N_\alpha+1}$, when $N_\alpha > 0$ and $c = \alpha \restriction_1$ when $N_\alpha = 0$.

Take any rational $0 < \delta^* < 1$. For any word γ we have that

$$\nu(\gamma\eta bc \mid \gamma\eta b) = \sum_{ef \in A(bc)} \left(\sum_{\sigma \in A(\gamma\eta)} P(\sigma e \mid A(\gamma\eta b)) \right) P(ef \mid e) \tag{17}$$

when $N_\alpha > 0$, and

$$\nu(\gamma c \mid \gamma) = \sum_{f \in A(c)} \left(\sum_{\sigma \in A(\gamma)} P(\sigma \mid A(\gamma)) \right) P(\sigma f \mid \sigma)$$

when $N_\alpha = 0$.

Since there are finitely many $ef \in A(bc)$ we may set, in case $N_\alpha > 0$,

$$K = \min\{P(ef \mid e) \colon P(ef \mid e) \neq 0, ef \in A(bc)\},$$

and then, noticing (17) consists of nonnegative summands and choosing any f for which $P(ef \mid e) \neq 0$

$$\nu(\gamma\eta bc \mid \gamma\eta b) \geq K \sum_{\sigma e \in A(\gamma\eta b)} P(\sigma e \mid A(\gamma\eta b)) = K.$$

Taking any strictly positive rational $\delta \leq \delta^* K$ we get

$$\delta \leq \delta^* K \leq \frac{\delta^* \nu(\rho c \mid \rho)}{1 - \nu(\rho c \mid \rho)} \tag{18}$$

for any $\rho = \gamma\eta b$.

In case $N_\alpha = 0$ we set $K = \min\{P(\sigma f \mid \sigma) \colon \ P(\sigma f \mid \sigma) \neq 0; f \in A(c)\}$ (which exists because P is Markovian) and (18) holds. Then, the function defined by

$$\begin{aligned} L(\emptyset) &= 1; \\ L(\rho a) &= \begin{cases} (1+\delta)L(\rho) & \text{if } \rho \lfloor_{N_\alpha} = \alpha \restriction_{N_\alpha} \text{ and } a \neq c; \\ (1-\delta^*)\, L(\rho) & \text{if } \rho \lfloor_{N_\alpha} = \alpha \restriction_{N_\alpha} \text{ and } a = c; \\ L(\rho) & \text{otherwise.} \end{cases} \end{aligned}$$

satisfies the supermartingale inequality, since

$$1 \geq (1+\delta)(1 - \nu(\rho c \mid \rho)) + \nu(\rho c \mid \rho)(1 - \delta^*)$$

follows from (18) for all $\rho = \gamma\eta b$ (in case $N_\alpha > 0$) and for all ρ in case $N_\alpha = 0$.

Observe that

$$L(\rho) = (1-\delta^*)^{\mathsf{occ}((\alpha\restriction_{N_\alpha})c,\rho)} \prod_{\substack{a\in\Sigma \\ a\neq c}} (1+\delta)^{\mathsf{occ}((\alpha\restriction_{N_\alpha})a,\rho)} .$$

Since the word $(\alpha \restriction_{N_\alpha})c$ occurs finitely many times in s, while $\alpha \restriction_{N_\alpha+1}$ occurs infinitely many times in s, we conclude that the above function goes to infinity when evaluated on increasing prefixes of s. □

3.2 P-martingales on a DFA cannot beat P-distributed sequences

Our next goal is to prove a converse to Theorem 3.3, thus providing a complete characterization of sequences that are "normal" relative to some irreducible Markov measure, a characterization that generalizes the main result of Schnorr and Stimm in [17]. Our proof will mirror their ideas closely. The main intuition is that a sequence where the average occurrences of blocks converge to some measure on those blocks should also have the average number of visits to any state of a DFA converge to some measure on the states. The main differences are that, in our case, some states are not final and that the probability of symbols and states are not independent.

To ease notation, we will only consider 1-step Markov measures. The reader may check that there is no loss in generality in this, since, as mentioned in §2, any k-step irreducible Markov measure on $\Sigma^{\mathbb{N}}$ induces a 1-step irreducible Markov measure P^k on $\Theta_k^{\mathbb{N}}$ and any P-martingale M generated by a DFA on alphabet Σ that succeeds on a sequence s can be regarded as a P^k-martingale M' generated by a DFA on alphabet Θ_k. This martingale may not succeed on $\langle s : k\rangle$ but it must succeed on $\langle T^i(s) : k\rangle$ for some $i \leq k$. Since P-distribution is unaffected by the removal of finitely many symbols, this suffices.

The main result of this section is an analogue of part a) of Theorem 4.1 in [17]:

Theorem 3.13. Let $L \subseteq \Sigma^*$ be a prolongable and factorial language, let P be an L-supported irreducible 1-step Markov measure on $\Sigma^{\mathbb{N}}$ and let $s \in X_L$ be P-distributed. Then no P-martingale generated by a DFA succeeds on s.

Take some $M = \langle Q, \Sigma, \delta, q_0, Q_f\rangle$ accepting L. We may assume all accepting states in Q_f are reachable from the initial state q_0.

Definition 3.14. Let M be a DFA and $q \in Q_f$, then M_q is the DFA that has the same states, accepting states, alphabet and transition function as M but which has q as its initial state.

Notice first that, since our language L is factorial, if a word σ is such that there are states $q, q' \in Q_f$ and $\delta^*(\sigma, q) = q'$, then $\sigma \in L$. That is, words that transition between accepting states must belong to the language. Also, factoriality implies that transition to an accepting state is not possible once a state outside Q_f is reached (hence, $q_0 \in Q_f$). Thus, we may assume that the complement of Q_f consists of a single state $\tilde{q}$. Similarly, the fact that the language is prolongable implies that from any accepting state there is always a transition to an accepting state.

As in [17], we will define the relation $q \to q'$ when there is a word $\sigma \in L(X)$ such that $\delta^*(\sigma, q) = q'$ and $q \leftrightarrow q'$ if both $q \to q'$ and $q' \to q$. From the remarks in the preceding paragraph it is easy to see that $\leftrightarrow$ is an equivalence relation and that it allocates accepting states q to classes $[q]$ different from $[\tilde{q}]$. Also, it is easy to check that the relation $\to$ induces a relation $\geq$ in the equivalence classes of $Q/\leftrightarrow$, where $[q] \geq [q']$ if and only if $q \to q'$ (we write $[q] > [q']$ when this holds and $[q] \neq [q']$). We will call a class $[q]$ *ergodic* if $[q] \neq [\tilde{q}]$ and if there is no $q' \in Q_f$ such that $[q] > [q']$ (i.e., $[q]$ is minimal among the classes of accepting states).

In order to prove the main result of this section we must make some considerations regarding the interaction of Markov measures that are L-supported and a DFA that accepts L.

For any $q' \in Q_f$ consider the maps $\phi_{q'} \colon X_L \to Q^{\mathbb{N}}$ and $\Phi_{q'} \colon X_L \to (\Sigma \times Q)^{\mathbb{N}}$ defined by the following rules:

$$\begin{aligned} \phi_{q'}(x)_1 &= q' \\ \phi_{q'}(x)_{n+1} &= \delta(x_n, \phi_{q'}(x)_n) \\ \Phi_{q'}(x)_n &= (x_n, \phi_{q'}(x)_n). \end{aligned}$$

The invariant, L-supported, irreducible 1-step Markov measure P of Theorem 3.13 allows us to define some random processes, that is, random variables indexed by natural numbers n. For any $n \geq 1$ let W_n, Y_n^q and $Z_n^{q'}$ be measurable functions (i.e. random variables) on the probability space $(X_L, \mathcal{B}, P)$

$$\begin{aligned} W_n(x) &= x_n \\ Y_n^{q'}(x) &= \phi_{q'}(x)_n \\ Z_n^{q'}(x) &= \Phi_{q'}(x)_n = (W_n, Y_n^q). \end{aligned}$$

As is customary for random variables, we will omit the specification of the element x of the probability space on which the random variable is being evaluated. Also, when $q' = q_0$, we will drop the subscripts and superscripts and write ϕ, Φ, Y_n and Z_n.

It is a commonplace observation in the theory of Markov processes [11, Theorem 1.1.2] that the random process $(Z_n^{q'})_{n \in \mathbb{N}}$ is Markov of order 1 or, equivalently, that the measure $P \circ \Phi_{q'}^{-1}$ on $(\Sigma \times Q)^{\mathbb{N}}$ is 1-step Markov.

Moreover, while the overall measure depends on the choice of q', the transition matrix does not.

For any ergodic class $[q^*]$, write

$$\mathsf{A}_{q^*} = \{(a,q) \in \Sigma \times [q^*] \colon \ \delta(a,q) \notin [\tilde{q}]\}$$

for the tuples of "admisible" pairs in $\Sigma \times [q^*]$. And for any $(a,q) \in \mathsf{A}_{q^*}$ we define a measure $\widehat{P}_{a,q}$ on $(\Sigma \times Q)^{\mathbb{N}}$ by letting

$$\widehat{P}_{a,q}(z_1 \dots z_l) = P\left(\{x \colon \ \forall\, 1 \le i \le l,\ Z_{N+i} = z_i\} \mid \{x \colon \ Z_N = (a,q)\}\right)$$

for any N such that $P(\{x \colon \ Z_N = (a,q)\}) > 0$ (equivalently, any N for which there is a word σ such that $|\sigma| = N-1$ and $\delta^*(\sigma, q_0) = q$). This measure is the probability distribution of the process $(Z_{N+i})_{i \in \mathbb{N}}$ conditioned on $Z_n = (a,q)$ and by [11, Theorem 1.1.2] it is a Markov measure independent of N and having the same transition matrix as $P \circ \Phi_{q'}^{-1}$ regardless of (a,q). When restricted to a given ergodic class $[q^*]$ we will denote this transition matrix by

$$\widehat{P}^{q^*} = (\widehat{P}^{q^*}_{z,z'})_{z,z' \in \mathsf{A}_{q^*}}.$$

Observation 3.15. For $z = (a,q)$ and $z' = (a',q')$ we have

$$\widehat{P}_{z,z'} = \begin{cases} P(aa' \mid a) & \text{if } \delta(a,q) = q'; \\ 0 & \text{otherwise.} \end{cases}$$

Lemma 3.16. Let $[q^*]$ be an ergodic class and $(a,q) \in \mathsf{A}_{q^*}$. Then the measure $\widehat{P}_{a,q}$ is supported on $\mathsf{A}_{q^*}^{\mathbb{N}}$ and has an irreducible transition matrix

Proof. Suppose that, for some N_1, $\widehat{P}_{a,q}(\{z \in (\Sigma \times Q)^{\mathbb{N}} \colon \ z_{N_1} = (a',q')\}) > 0$. This means that there are words $\tau, \rho \in L$ such that $\delta^*(\tau, q_0) = q$, $|\rho| = N_1 - 1$, $\delta^*(\rho, \delta(a,q)) = q'$ and $P(\tau a \rho a') > 0$. But since P is $L(X)$-supported this means $\tau a \rho a' \in L(X)$ and therefore $\hat{q} = \delta^*(\tau a \rho a', q_0)$ is an accepting state and $\hat{q} = \delta(\rho a', q)$, so that $[\hat{q}] \le [q]$. But since $[q]$ is an ergodic class it must be the case that $[q'] = [q] = [q^*]$. Hence, $\widehat{P}_{a,q}$ is supported on $\mathsf{A}_{q^*}^{\mathbb{N}}$.

To see that the transition matrix of $\widehat{P}_{a,q}$ is irreducible, take any $z_1 = (a_1, q_1), z' = (a',q') \in \mathsf{A}_{q^*}$. We want to find some word $\rho = z_2 \dots z_m$ in $\mathsf{A}_{q^*}^*$ such that

$$\widehat{P}_{a,q}((a_1,q_1)\rho(a',q')) > 0. \tag{19}$$

Let $q_2 = \delta(a_1, q_1)$. Since $\widehat{P}_{a,q}$ is supported on $\mathsf{A}_{q^*}^{\mathbb{N}}$ it follows that q_2 is also in the ergodic class $[q^*]$, and since q' also belongs to the same class $[q^*]$ by hypothesis, then there must be a word $\sigma \in L(X)$ such that $\delta^*(\sigma, q_2) = q'$. Write $\sigma = a_2 \dots a_m$ and inductively define $q_{i+1} = \delta(a_i, q_i)$ for $2 \le i < m$. Let us show the word $z_2 \dots z_m$ for $z_i = (a_i, q_i)$ satisfies (19). Take any word τ such that $\delta^*(\tau, q_0) = q$. Then $\delta^*(\tau a a_1 \sigma, q_0) = q'$ and since $(a',q') \in \mathsf{A}_{q^*}^*$ then $\delta(a',q') \notin [\tilde{q}]$, so that $\delta^*(\tau a a_1 \sigma a', q_0) \in Q_f$ and therefore $\tau a a_1 \sigma a' \in L(X)$. Since P is $L(X)$-supported, this means $P(\tau a a_1 \sigma a') > 0$, and from $[\tau a a_1 \sigma a'] \subseteq \Phi^{-1}\left(T^{-|\tau|}[(a,q)(a_1,q_1)z_2 \dots z_m(a',q')]\right)$ we derive (19). $\square$

Now, $\widehat{P}_{a,q}$ is 1-step Markov with an irreducible transition matrix. Given a 1-step Markov measure with irreducible transition matrix, the ergodic theorem for Markov processes (Theorem 1.10.2 from [11]) ensures that the Cesaro averages of cylinder characteristic functions converge almost surely to a constant that depends only on the transition matrix. In our context, that result has to be restated in the following form:

Theorem 3.17. Let P and L be as in Theorem 3.13 and M be a DFA accepting L. Let $[q^*]$ be an ergodic class and $(a,q),(a',q') \in \mathsf{A}_{q^*}$. Then there is some constant $k_{a',q'}$ independent of (a,q) such that

$$\widehat{P}_{a,q}\left(z\colon \sum_{i=1}^{N} \frac{\mathcal{X}_{(a',q')}(z_i)}{N} \to k_{a',q'}\right) = 1.$$

Moreover, the vector $\psi_{q^*} = (k_{a',q'})_{(a',q')\in\mathsf{A}_{q^*}}$ is a distribution on A_{q^*} (that is, it has nonnegative entries that add up to 1) and is a left eigenvector of the transition matrix $\widehat{P}^{q^*}$, that is $\psi_{q^*}\widehat{P}^{q^*} = \psi_{q^*}$.

At this point, we would like to prove an analogue of Lemma 4.5 from [17], which states a precise formulation of the idea that if a sequence s is P-distributed then the joint sequence of visited states and symbols should also be distributed according to some measure derived from P. If the sequence of visited states were eventually concentrated in some ergodic class $[q^*]$, then $k_{a,q}$ would be the natural candidate for that derived measure. The following simple but useful result will allow us to make that assumption regarding an eventual ergodic class $[q^*]$.

Lemma 3.18. Given a DFA M accepting a factorial and prolongable language L and a class of accepting states $[q] \in Q_f/\leftrightarrow$, there is a word $\sigma \in L$ such that for all $s \in [q]$, the class $[\delta^*(\sigma,s)]$ is ergodic or is equal to $[\tilde{q}]$.

Proof. If $[q]$ is ergodic then any word $\sigma \in L$ will do, so let us assume $[q]$ is not ergodic. Let us write $[q] = \{q_0,\ldots,q_m\}$. The proof will show by induction on i that there are words $\sigma_i \in L$ such that, for all $j \le i$, $[\delta^*(\sigma_i,q_j)] = [\tilde{q}]$ or $[\delta^*(\sigma_i,q_j)]$ is ergodic. For $i=1$, since $[q]$ is not ergodic there must be some ergodic $[q']$ such that $[q] > [q']$. Hence, we can choose a word $\sigma_1 \in L$ such that $\delta^*(\sigma_1,q_1) \in [q']$.

For the inductive step, if $[\delta^*(\sigma_i,q_{i+1})] = [\tilde{q}]$ or $[\delta^*(\sigma_i,q_{i+1})]$ is ergodic then we are done. Otherwise, there must be some ergodic $[q']$ such that $[\delta^*(\sigma_i,q_{i+1})] > [q']$. Hence, there is a word σ such that $[\delta^*(\sigma,\delta^*(\sigma_i,q_{i+1}))] = [q']$ is ergodic, and we choose $\sigma_{i+1} = \sigma_i\sigma$, which belongs to L because L is factorial and σ_{i+1} transitions between accepting states q_{i+1} and q' (for some representative of $[q']$). □

We need some notation for the 2-tuples of letters and states visited by words of finite and fixed length. For this purpose, let $\Phi_q^k\colon \Sigma^k \to (\Sigma\times Q)^k$

be defined as

$$\Phi_q^1(a) = (a, q)$$
$$\Phi_q^{k+1}(\sigma a) = \Phi_q^k(\sigma)(a, \delta(\Phi_q^k(\sigma)_k.$$

We now get the desired generalization of Lemma 4.5 from [17].

Lemma 3.19. Let L, P and s be as in Theorem 3.13 and M be some DFA accepting L. Then there is some ergodic class $[q^*]$ such that, for all $(a', q') \in \mathsf{A}_{q^*}$, we have

$$\lim_N \sum_{n=1}^{N} \frac{\mathcal{X}_{(a',q')}(Z_n(s))}{N} = k_{a',q'}. \tag{20}$$

Proof. Notice that, since $s \in X$, M on input $s \restriction_N$, will never reach the garbage state $\tilde{q}$, since $s \in X_L$ and L is a factorial language. Moreover, we will show that there is some N_0 such that, for all $N > N_0$, M stays within some ergodic class $[q^*]$ after reading the first N symbols of s, i.e., $[\delta^*(s \restriction_N, q_0)] = [q^*]$ for all $N > N_0$.

Observe that if $M > N$, then $[\delta^*(s \restriction_M, q_0)] = [\delta^*(s \restriction_N, q_0)]$ or $[\delta^*(s \restriction_M, q_0)] > [\delta^*(s \restriction_N, q_0)]$, and since there are only finitely many classes in $Q_f/\leftrightarrow$ this means that there is some N_0 such that, for all $N > N_0$, M stays in the same class, that is,

$$[\delta^*(s \restriction_{N_0}, q_0)] = [\delta^*(s \restriction_N, q_0)]. \tag{21}$$

We will call this class $[q^*]$ and claim it is ergodic.

Indeed, by Lemma 3.18 we could choose some word $\sigma \in L(X)$ such that, after reading it from any state in $[q^*]$, M either reaches the garbage state $\tilde{q}$ or reaches a state within an ergodic class. Since $s \in L(X)$ and P is L-supported, we have $P(\tau) > 0$, and since L is P-distributed, σ occurs (infinitely many times) in s. Since $\tilde{q}$ cannot be reached when a subword of s is read as input, it follows that an ergodic class $[q']$ is reached at some $N > N_0$. By equation (21) it follows that $[q^*] = [q']$ and $[q^*]$ is ergodic.

Fix some $(a', q') \in \mathsf{A}_{q^*}$. Let us now apply Theorem 3.17 to $[q^*]$. We have

$$\widehat{P}_{a,q}\left(z : \lim_N \sum_{i=1}^{N} \frac{\mathcal{X}_{(a',q')}(z_i)}{N} = k_{a',q'}\right) = 1$$

for all $(a, q) \in \mathsf{A}_{q^*}$, which in turn implies that for all $\epsilon, \epsilon' > 0$ there is k_0 such that for all $k \geq k_0$ and $(a, q) \in \mathsf{A}_{q^*}$,

$$\widehat{P}_{a,q}\left((a_1, q_1)\dots(a_l, q_k) \in (\Sigma \times Q)^k : \left|\sum_{i=1}^{k} \frac{\mathcal{X}_{(a',q')}((a_i, q_i))}{N} - k_{a',q'}\right| < \epsilon\right)$$
$$> 1 - \epsilon'. \tag{22}$$

Notice that (22) can be rewritten as

$$P\left(a\sigma \in \Sigma^k : \left|\sum_{i=1}^{k} \frac{\mathcal{X}_{(a',q')}(\Phi_q^k(a\sigma)_i)}{N} - k_{a',q'}\right| < \epsilon\right)/P(a) > 1-\epsilon'. \quad (23)$$

Write

$$\mathsf{B}_{q^*}^k(\epsilon, a)$$
$$= \left\{a\sigma \in \Sigma^k : (\forall q \in [q^*], (a,q) \in \mathsf{A}_{q^*}) \left|\sum_{n=1}^{k} \frac{\mathcal{X}_{(a',q')}(\Phi_q^k(a\sigma)_n)}{N} - k_{a',q'}\right| < \epsilon\right\}$$
$$\mathsf{B}_{q^*}^k(\epsilon) = \bigcup_{a\in\Sigma} \mathsf{B}_{q^*}^k(\epsilon, a).$$

These are just the words of length k for which the average occurrences of all pairs $(a,q), \in \mathsf{A}_{q^*}$ are within an ϵ distance of their limit. Then it is a standard exercise in probability theory to see that (23) and the finiteness of A_{q^*} and $[q^*]$ imply that for all a $(a,q) \in \mathsf{A}_{q^*}$ for some q, and for all $\epsilon, \epsilon' > 0$ there k_0 such that for all $k \geq k_0$ we have

$$P\left(\bigcup\{[\sigma] \colon \sigma \in \mathsf{B}_{q^*}^k(\epsilon, a)\}\right)/P(a) > 1-\epsilon'. \quad (24)$$

Let us see that this last inequality holds for all a, that is, for all a there is some q satisfying $(a,q) \in \mathsf{A}_{q^*}$. Indeed, the irreducibility of the transition matrix of the L-supported measure P implies that for all $a \in \Sigma$ and all $\sigma \in \Sigma^*$ there is some $\rho \in \Sigma^*$ such that $\sigma\rho a \in L$. Take any σ such that $\delta^*(\sigma, q_0) \in [q^*]$ and then take some ρ such that $\sigma\rho a \in L$. This implies $\delta^*(\sigma\rho a, q_0) \notin [\tilde{q}]$ and since $[q^*]$ is an ergodic class this also implies that $\delta^*(\sigma\rho, q_0) \in [q^*]$. So $(a, \delta^*(\sigma\rho, q_0)) \in \mathsf{A}_{q^*}$.

Then from (24) we derive that for all $\epsilon, \epsilon' > 0$ there is k_0 such that for all $k \geq k_0$ we have

$$P\left(\bigcup\{[\sigma] \colon \sigma \in \mathsf{B}_{q^*}^k(\epsilon)\}\right) > 1-\epsilon'. \quad (25)$$

Now, remember $\langle s : k\rangle$ is the sequence s read as a sequence in $(\Sigma^k)^{\mathbb{N}}$, then the P-distribution of s implies that for all $\epsilon'' > 0$ and $k \in \mathbb{N}$ there is M_0 such that for all $\sigma \in \Sigma^k$ and $M \geq \max(M_0, 2/\epsilon'')$, we have

$$\left|\frac{\mathsf{occ}(\sigma, \langle s : k\rangle \restriction_M)}{M} - P(\sigma)\right| < \frac{\epsilon''}{2|\Sigma|^k}. \quad (26)$$

Take N_1 such that $\delta^*(s\restriction_{N_0}, q_0) = q^*$ for q^* some representative of $[q^*]$. Since finitely many summands do not alter the convergence of Cesaro limits, we may substitute $s' = T^{N_1}s$ for s and rewrite (20) as

$$\lim_N \sum_{n=1}^{N} \frac{\mathcal{X}_{(a',q')}(Z_n^{q^*}(s'))}{N} = k_{a',q'}.$$

Given $\epsilon, \epsilon', \epsilon'' > 0$ such that $\epsilon + \epsilon' + \epsilon'' < \eta$ for some $\eta > 0$, take some k satisfying (25) and take some M_0 satisfying (26) for this k. For any $M \geq \max(M_0, 2/\epsilon'')$ we consider $N = kM$ and notice that

$$\begin{aligned}
&\left| \sum_{n=1}^{N} \frac{\mathcal{X}_{(a',q')}(Z_n^{q^*}(s'))}{N} - k_{a',q'} \right| \\
&\qquad = \left| \sum_{j=0}^{M-1} M^{-1} \sum_{i=1}^{k} \frac{\mathcal{X}_{(a',q')}(\Phi^k_{Y_{jk+1}}(\langle s' : k \rangle_j)_i)}{k} - k_{a',q'} \right|,
\end{aligned}$$

where all Y_{jk+1} are guaranteed to be in $[q^*]$ by our substitution of s. This implies (using first (25) and then (26)) that

$$\begin{aligned}
&\left| \sum_{n=1}^{N} \frac{\mathcal{X}_{(a',q')}(Z_n^{q^*}(s'))}{N} - k_{a',q'} \right| \\
&\quad < \epsilon \sum_{\sigma \in \mathsf{B}^k_{q^*}(\epsilon)} \frac{\mathsf{occ}(\sigma, \langle s' : k \rangle \restriction_M)}{M} + 2 \sum_{\substack{\sigma \in \Sigma^k \\ \sigma \notin \mathsf{B}^k_{q^*}(\epsilon)}} \frac{\mathsf{occ}(\sigma, \langle s' : k \rangle \restriction_M)}{M} \\
&\quad < \epsilon + 2 \left| \sum_{\substack{\sigma \in \Sigma^k \\ \sigma \notin \mathsf{B}^k_{q^*}(\epsilon)}} P(\sigma) \right| + 2 \sum_{\substack{\sigma \in \Sigma^k \\ \sigma \notin \mathsf{B}^k_{q^*}(\epsilon)}} \left| \frac{\mathsf{occ}(\sigma, \langle s' : k \rangle \restriction_M)}{M} - P(\sigma) \right| \\
&\quad < \epsilon + 1 - P\Big(\bigcup_{\sigma \in \mathsf{B}^k_{q^*}(\epsilon)} [\sigma] \Big) + 2|\Sigma|^k \frac{\epsilon''}{2|\Sigma|^k} < \epsilon + \epsilon' + \epsilon'' < \eta.
\end{aligned}$$

For $N = kM + l$ (where $1 \leq l < k$) we have

$$\begin{aligned}
&\left| \sum_{n=1}^{N} \frac{\mathcal{X}_{(a',q')}(Z_n^{q^*}(s'))}{N} - k_{a',q'} \right| \\
&\qquad \leq \left| \sum_{n=1}^{Mk} \frac{\mathcal{X}_{(a',q')}(Z_n^{q^*}(s'))}{Mk} - k_{a',q'} \right| + \frac{l}{N} + |kM/N - 1| \\
&\qquad < \eta + \frac{2}{M} < \eta + \epsilon''.
\end{aligned}$$

This completes the proof. □

Proof of Theorem 3.13. Since our martingale is generated by a DFA with betting factors function b, states set Q and transition function δ, we know it satisfies

$$f(\sigma) = f(\emptyset) \prod_{\substack{a \in \Sigma \\ q \in Q}} b(a,q)^{\mathsf{occ}\left((a,q), \Phi^{|\sigma|}_{q_0}(\sigma)\right)}. \tag{27}$$

By the same reasoning used in the first part of the proof of Lemma 3.19 we know there is some N_1 such that $\delta^*(s\restriction_N, q_0) \in [q^*]$ for all $N \geq N_1$ and some ergodic class $[q^*]$, From (27) it is clear that success of f on s is equivalent to success of f' on $T^{N_1}(s)$, where f' is the martingale that is generated by a DFA having $Q' = [q^*] \cap [\tilde{q}]$ as its set of states, $q_1 = \delta^*(s\restriction_{N_1}, q_0)$ as its initial state (notice $[q_1] = [q^*]$), and the restriction of δ and b to Q' as its transition and betting factors functions, respectively. That is, we restrict our analysis to the case in which M starts and stays within a single ergodic class.

Since $s' = T^{N_1}(s) \in X_L$ we know s' will only visit tuples $(a, q) \in \mathsf{A}_{\mathsf{q}_1}$, that is, for all N, $\Phi^N_{q_1}(s'\restriction_N)$ is in $\mathsf{A}^*_{q_1}$. Then from Lemma 3.19 and (27) we have

$$\lim_N \frac{(f'(s'\restriction_N))^{1/N}}{\prod_{(a,q)\in\mathsf{A}_{\mathsf{q}_1}} b(a,q)^{k_{a,q}}} = 1.$$

If $r = \prod_{(a,q)\in\mathsf{A}_{\mathsf{q}_1}} b(a,q)^{k_{a,q}} < 1$, then $f'(s'\restriction_N) < (r+\epsilon')^N$ for $r + \epsilon' < 1$ for large enough N. Hence $\lim f'(s'\restriction_N) = 0$ and the martingale does not succeed on s'.

Let us now show that for fixed $q \in [q_1]$

$$U_q = \prod_{\substack{a\in\Sigma \\ (a,q)\in\mathsf{A}_{\mathsf{q}_1}}} b(a,q)^{k_{a,q}} \leq 1.$$

Indeed, by Observation 3.15, Theorem 3.17 and convexity of the logarithm we get

$$\begin{aligned}
\log U_q &= \sum_{\substack{a\in\Sigma \\ (a,q)\in\mathsf{A}_{\mathsf{q}_1}}} k_{a,q} \log b(a,q) \\
&= \sum_{(a',q')\in\mathsf{A}_{\mathsf{q}_1}} k_{a',q'} \sum_{\substack{a\in\Sigma \\ (a,q)\in\mathsf{A}_{\mathsf{q}_1}}} \widehat{P}^{q_1}_{(a',q'),(a,q)} \log b(a,q) \\
&\leq \sum_{(a',q')\in\mathsf{A}_{\mathsf{q}_1}} k_{a',q'} \log \left(\sum_{\substack{a\in\Sigma \\ (a,q)\in\mathsf{A}_{\mathsf{q}_1}}} \widehat{P}^{q_1}_{(a',q'),(a,q)} b(a,q) \right) \\
&= \sum_{\substack{(a',q')\in\mathsf{A}_{\mathsf{q}_1} \\ \delta(a',q')=q}} k_{a',q'} \log \left(\sum_{a\in\Sigma} P(a'a \mid a') b(a,q) \right) \\
&= \sum_{\substack{(a',q')\in\mathsf{A}_{\mathsf{q}_1} \\ \delta(a',q')=q}} k_{a',q'} \log(1) = 0, \qquad (28)
\end{aligned}$$

where the last line follows from the fact that $b(a,q)$ are the betting factors of a P-martingale.

It follows that $r = \prod_q U_q \leq 1$ and equality is achieved if and only if $U_q = 1$ for all $q \in [q_1]$. But, by strict convexity in (28), $\log U_q = 0$ if and only if $b(a, q)$ is the same for all a. Then $r = \prod_q U_q = 1$ if and only if the betting factors are constant at each state, implying $f'(\sigma a) = f'(\sigma)$ for all words σ and all a. Clearly, a constant martingale cannot succeed on any sequence, and the result follows. □

4 β-expansions and Pisot numbers

In this section we introduce some definitions and known results on the representation of reals in non-integer bases, and Pisot numbers. All of this material will be needed for our result of §5.

4.1 β-expansions

Let us now introduce a way of representing real numbers in a non-integer base β. Most of the presentation and the definitions are taken from [2]. Let $\lfloor x \rfloor$ and $\lceil x \rceil$ be the floor and ceiling of x, respectively, and let $\{x\}$ denote the integer and fractional part.

Let β be a real number greater than 1. Any real number x has a unique β-expansion $s_0^\beta, s_1^\beta, \ldots$ such that

$$x = s_0^\beta + \sum_{n>0} \frac{s_n^\beta}{\beta^n}, \tag{29}$$

where s_n^β are nonnegative integers, $0 \leq s_n^\beta < \beta$ for $n > 0$ and any of the following equivalent conditions are met:

1. $\forall n \geq 0 \sum_{i>n} (s_i^\beta / \beta^i) < 1/\beta^n$
2. s_n^β is defined inductively in the following way:

$$\begin{array}{rclrcl} s_0^\beta & = & \lfloor x \rfloor, & r_0 & = & \{x\} \\ s_{n+1}^\beta & = & \lfloor \beta r_n \rfloor, & r_{n+1} & = & \{\beta r_n\} \end{array}$$

This expansion coincides with the usual definition given for β an integer base. Notice that the β-expansion of the real number β need not be eventually periodic, in particular, it need not be finite, that is, eventually 0 (of course it is when β is an integer). It is easy to check from the definition of a β-expansion that if β had a finite expansion then it satisfies $\beta = a_0 + \frac{a_1}{\beta} + \cdots + \frac{a_s}{\beta^s}$ and the periodic sequence $a_0 a_1 \ldots (a_s - 1) a_0 a_1 \ldots (a_s - 1) \ldots$ would also satisfy (29) (but not the following two equivalent conditions, since the β-expansion is unique).

We will refer to such a periodic sequence as the *periodic* β-expansion of β and we we will write it $\hat{\beta}$ (notice that the periodic β-expansion may only

apply to the base β). We will also write $\mathfrak{s}(\beta)$ for the β-expansion of β in case it is not terminated by an infinite sequence of 0's and $\mathfrak{s}(\beta) = \hat{\beta}$ otherwise.

Example 4.1. The periodic 2-expansion of 2 is 1^∞, whereas its 2-expansion is 20^∞. Let ϕ be the golden number satisfying $\phi^2 = \phi + 1$. The periodic ϕ-expansion of ϕ is $101010\ldots$, whereas its ϕ-expansion is 110^∞.

Given a base β, let $\Sigma_\beta = \{0 \ldots \lceil \beta \rceil - 1]\}$ and let $p_\beta \colon [0,1) \to \Sigma_\beta^{\mathbb{N}}$ be the one-to-one mapping that sends each $x \in [0,1)$ to the fractional part of its β-expansion

$$s_1^\beta s_2^\beta \ldots s_n^\beta \ldots$$

Notice that $p_\beta(1) = \mathfrak{s}(\beta)$ (strictly speaking, p_β is defined on $[0,1)$ but it is trivially extended to $[0,1]$ by continuity).

If Σ is a finite set of digits, as in the definition of the mapping p_β, then the natural ordering of those digits induces a lexicographic order $\leq_{\text{lex}}$ on the full shift.

Theorem 4.2. [2, p. 273] If $\beta > 1$ is a real base then the image $p_\beta([0,1))$ is the set

$$\{s \in \Sigma_\beta^{\mathbb{N}} \colon (\forall n)\ T^n s <_{\text{lex}} \mathfrak{s}(\beta)\}.$$

Notice that the closure of the set above, that is,

$$\{s \in \Sigma_\beta^{\mathbb{N}} \colon (\forall n)\ T^n s \leq_{\text{lex}} \mathfrak{s}(\beta)\}.$$

is a subshift of $\Sigma_\beta^{\mathbb{N}}$. In fact, there is a nice converse to the above theorem.

Theorem 4.3. [2, p. 274] Suppose that for some alphabet $\Sigma = \{0, \ldots, k\}$ we have that (X, T) is a subshift such that

$$X = \{s \in \Sigma^{\mathbb{N}} \colon (\forall n)\ T^n s \leq_{\text{lex}} s^*\}$$

for some s^* in $\Sigma^{\mathbb{N}}$ which satisfies $(\forall n)\ T^n s^* \leq_{\text{lex}} s^*$. Then X is the closure of $p_\beta([0,1))$ for some real base β.

This allows us to define the following:

Definition 4.4. Given some real number $\beta > 1$, the *β-shift* is the subshift (X_β, T), where

$$X_\beta = \{s \in \Sigma_\beta^{\mathbb{N}} \colon (\forall n \in \mathbb{N})\ T^n s \leq_{\text{lex}} \mathfrak{s}(\beta)\}.$$

Example 4.5. The 2-shift is the full shift $\{0,1\}^{\mathbb{N}}$ (that is, the Cantor set with the shift operator). The ϕ-shift is the set of infinite sequences on $\{0,1\}$ such that no two 1's occur consecutively in them. In fact, this shift is Markov and it is the Markov shift which had the sofic shift of Example 2.6 as a factor.

Theorem 4.6. [2, Theorem 1] Let $\beta > 1$ be a real base. Then the β-shift is a Markov shift if and only if the β-expansion of β is finite, and it is sofic if and only if the β-expansion of β is periodic.

A result similar to Theorem 2.8 for β-shifts was also proved by Parry in [12]:

Theorem 4.7. Given a real base $\beta > 1$, there is a unique probability measure $\widehat{P}_\beta$ on $[0, 1)$ such that $P_\beta = \widehat{P}_\beta \circ p_\beta^{-1}$ is an invariant measure for the β-shift of maximal metric entropy on X_β. Moreover, $\widehat{P}_\beta$ has the closed expression

$$\widehat{P}_\beta([a, b]) = \int_a^b \sum_{n=1}^{\infty} \mathbb{1}_{[0,T_\beta^n(1))}(x) \frac{1}{\beta^n} dx, \tag{30}$$

and if (X_β, T) is a Markov shift with a grammar of wordlength $k - 1$, then P_β, called the *Parry measure*, is a k-step Markov measure.

Notice that the above expression implies that there are positive k and k' such that

$$k'\lambda(A) \leq \widehat{P}_\beta(A) \leq k\lambda(A) \tag{31}$$

for any Borel subset A of $[0, 1)$, and λ the Lebesgue measure.

4.2 Pisot numbers

While constructive considerations make us think of rational numbers as the closest relatives of integers, the analysis of real base expansions forces us to consider the "dynamic" properties of real numbers, and from a dynamical viewpoint non-integer rational numbers are quite distinct from integers. The following definition will introduce us to the closest analog of an integer from a dynamic point of view.

Definition 4.8. A real number β is a *Pisot number* if $\beta > 1$ and β is the root of a monic polynomial in integer coefficients, such that all its conjugate values (that is, all the other roots of its minimal polynomial) have absolute values strictly less than 1.

This purely algebraic condition is interesting for our purposes because of the next remarkable property. Let $\|x\|$ denote the distance from x to its closest integer

Theorem 4.9. [2, Lemma 1] A real number $\beta > 1$ is a Pisot number if and only if $\sum_{n\geq 0} \|\beta^n\|$ converges.

Pisot numbers are then "asymptotically integer" in a strong sense. Notice that all integers $n > 1$ are Pisot numbers, but no non-integer rational number is Pisot, since the only rationals which are roots of monic polynomials in $\mathbb{Z}$ are the integers. The following results relate Pisot numbers and β-expansions.

Theorem 4.10. [2, Theorem 5] If β is a Pisot number, $\mathfrak{s}(\beta)$ is eventually periodic and X_β is a sofic subshift.

Theorem 4.11. [2, Corollary 9] Let β be a Pisot number, x be a real number with β-expansion s and assume that s is P_β-distributed. Then $(x\beta^n)_{n\geq 0}$ is u.d. modulo one.

None of the above implications has a true converse.

5 Polynomial time randomness

In this section we will use the martingale constructed in §3 to show that if $x \in [0,1]$ is a real whose binary expansion is polynomial time random (i.e. no feasible martingale succeeds on it), then $(x\beta^n)_{n\in\mathbb{N}}$ is u.d. modulo one for any Pisot β.

The reasoning follows by contradiction: if x is such that $(x\beta^n)_{n\geq 0}$ is not u.d. modulo one for some Pisot base β, we know by Theorem 4.11 that there is some word σ in Σ_β^* whose average occurrences in the β-expansion of x do not converge to $P_\beta(\sigma)$. From Theorem 3.3 we then get a P_β-martingale on a DFA that succeeds on the β-expansion of x. Our task in this section is to show that such a martingale can be translated to a base 2 martingale that is computable in polynomial time. Base 2 suffices because of the (integer) base invariance of polynomial time randomness [6].

The rest of the section is organized as follows. In §5.1 we show a feasible method to approximate dyadic rationals with reals in base β. In §5.2 we introduce the *savings property* for P-martingales and show that any feasible P-martingale can be translated to one with the savings property, preserving the succeeding points. In §5.3 we derive some useful properties of the Parry measure P_β and introduce the measure μ_M over $[0,1]$ induced by any P_β-martingale M. In §5.4 we show that the cumulative distribution function of μ_M is polynomial time computable when restricted to β-adic inputs. Finally, in §5.5 we show the main result via an 'almost Lipschitz' property, as in [6].

5.1 Dyadic rationals to base β

We derive some feasibility properties of β-ary representation.

Proposition 5.1. If $\beta > 1$ is Pisot then the set $L(X_\beta) = \{\tau \in \Sigma_\beta^* \mid \tau 0^\infty \in X_\beta\}$ is decidable in linear time.

Proof. Immediate from Theorem 4.2, the fact that $\mathfrak{s}(\beta)$ is eventually periodic (Theorem 4.10) and the linear time complexity of lexicographic comparison. □

Fix a finite alphabet Σ. We say that a function $g\colon \Sigma^* \to \mathbb{R}$ is *computable* if there is a computable function $\widehat{g}\colon \Sigma^* \times \mathbb{N} \to \mathbb{Q}$ such that for all σ and i we

have $|\widehat{g}(\sigma, i) - g(\sigma)| \leq 2^{-i}$. We call $\widehat{g}$ a *computable approximation* of g. We say that g is $t(n)$-*computable* if there is a Turing machine which on input σ and i computes $\widehat{g}(\sigma, i)$ in time $O(t(i + |\sigma|))$. As usual, we say that g is *polynomial time computable* if it is $t(n)$-computable for some polynomial t.

Observation 5.2. If f, g are polynomial time computable then $f + g$ and fg are polynomial time computable.

For $\beta > 1$, let $\langle\cdot\rangle_\beta \colon \Sigma_\beta^* \to \mathbb{R}$ be the function $\langle\tau\rangle_\beta = \sum_{k=1}^{|\tau|} \tau(k-1)\beta^{-k}$. Observe that in case τ is a prefix of some sequence in X_β then $\langle\tau\rangle_\beta$ is the only real $x \in [0, 1)$ such that $p_\beta(x) = \tau 0^\infty$.

Proposition 5.3. If $\beta > 1$ is Pisot, then the function $\langle\cdot\rangle_\beta$ is polynomial time computable.

Proof. The number of summands in $\langle\tau\rangle_\beta$ is the length of τ, which is computable in linear time. For each summand, $\tau(k)$ is computable in linear time, and β is an algebraic number, which is computable in polynomial time [8, Corollary 4.3.1]. Since numbers computable in polynomial time form a field [8, Corollary 4.3.2], β^{-k} is computable in polynomial time. Then both $\tau(k)$ and β^{-k} are polynomial-time computable and Observation 5.2 applies. □

Given a real $r \in [0, 1)$ and $i \in \mathbb{N}$, a word $\tau \in L(X_\beta)$ is said to be an *approximation of* r *in base* β *with error* 2^{-i} if $|\langle\tau\rangle_\beta - r| \leq 2^{-i}$.

Proposition 5.4. If $\beta > 1$ is Pisot then the problem of finding an approximation in base β of a dyadic rational $\langle\sigma\rangle_2$ ($\sigma \in \{0, 1\}^*$) with error 2^{-i} is computable in time polynomial in $|\sigma| + i$.

Proof. Let $\langle\cdot,\cdot\rangle_\beta \colon \Sigma_\beta^* \times \mathbb{N} \to \mathbb{Q}$ be a polynomial time computable approximation of $\langle\cdot\rangle_\beta \colon \Sigma_\beta^* \to \mathbb{R}$ (which exists by Proposition 5.3). Consider Algorithm 1.

Notice that when the algorithm terminates, we have $|\langle\tau, i+2\rangle_\beta - r| \leq 2^{-i-1}$; since $|\langle\tau, i+2\rangle_\beta - \langle\tau\rangle_\beta| \leq 2^{-i-2} < 2^{-i-1}$, we have $|\langle\tau\rangle_\beta - r| \leq 2^{-i}$. Observe also that by construction, τ is always a prefix of some sequence in X_β. Hence the value of τ by the time the algorithm terminates satisfies the postcondition. After each execution of the loop body, either

1. $|\langle\tau, i+2\rangle_\beta - r| \leq 2^{-i-1}$ (in which case it will immediately terminate), or
2. $\tau \prec p_\beta(r)$.

Let $I_\tau = \{x \in [0, 1) \mid \tau \prec p_\beta(x)\}$. If 1 does not hold then, by construction, $r \in I_\tau$, and it is clear that in case $\lambda I_\tau \leq 2^{-i-2}$ then it terminates (since $|\langle\tau, i+2\rangle_\beta - \langle\tau\rangle_\beta| \leq 2^{-i-2}$ and $|\langle\tau\rangle_\beta - r| \leq \lambda I_\tau \leq 2^{-i-2}$, and so

Algorithm 1: Approximation of a dyadic rational in base β

input : $\sigma \in \{0,1\}^*$, $i \in \mathbb{N}$
output: τ, a prefix of some sequence in X_β, such that $|\langle\tau\rangle_\beta - \langle\sigma\rangle_2| \leq 2^{-i}$

let $r = \langle\sigma\rangle_2$ and $\tau = \emptyset$
while $|\langle\tau, i+2\rangle_\beta - r| > 2^{-i-1}$ **do**
  let $S = \{a \in \Sigma_\beta \mid \tau a 0^\infty \in X_\beta\}$
  let $b \in S$ be the greatest such that $\langle\tau b, i+2\rangle_\beta \leq r$
  if $b = \max S$ **then**
    $\tau = \tau b$
  else
    $b' = b + 1$
    if $\langle\tau b', i+2\rangle_\beta - 2^{-i-1} < r$ **then**
      $\tau = \tau b'$
    else
      $\tau = \tau b$

$|\langle\tau, i+2\rangle_\beta - r| \leq 2^{-i-1}$). At each iteration the string τ is extended in one symbol. We will later see (Corollary 5.15) that $\lambda I_\tau \leq \beta^{-|\tau|}$, if $\beta^{-|\tau|} \leq 2^{-i-2}$ then the algorithm terminates, and so $|\tau|$ is $O(i)$. By Proposition 5.1 and Proposition 5.3, the execution of a single iteration is polynomial in $|\sigma|+i+|\tau|$. Since both the number of iterations and $|\tau|$ is $O(i)$, the execution of Algorithm 1 on in input σ, i is also polynomial in $|\sigma| + i$. □

5.2 The savings property

We say that a P-martingale M on $L \subseteq \Sigma^*$ has the *savings property* if there is $c > 0$ such that for all $\tau, \sigma \in L$, if $\tau \succeq \sigma$ then $M(\sigma) - M(\tau) \leq c$.

Proposition 5.5. Let $L \subseteq \Sigma^*$ be a nonempty, factorial and prolongable language, let P be an L-supported probability measure on $\Sigma^{\mathbb{N}}$ such that there is $a > 0$ such that $1 - P(\sigma b|\sigma) \leq a \cdot P(\sigma b|\sigma)$ for all $\sigma b \in L$, and let M be a P-martingale on L with the savings property via c. Then $M(\sigma) \leq c \cdot d \cdot |\sigma| + M(\emptyset)$ for all $\sigma \in L$.

Proof. As in [6, Proposition], proof is by induction on the length of σ. For the inductive step, we only need notice that $P(\sigma b|\sigma)$ is well defined and

positive. Then

$$\begin{aligned} M(\sigma b) &= \frac{M(\sigma) - \sum_{d\in\Sigma, d\neq b} P(\sigma d|\sigma) M(\sigma d)}{P(\sigma b|\sigma)} \\ &\leq \frac{M(\sigma) - \sum_{d\in\Sigma, d\neq b} P(\sigma d|\sigma)(M(\sigma) - c)}{P(\sigma b|\sigma)} \\ &\qquad\qquad\qquad\qquad\qquad\qquad (M \text{ has the savings property}) \\ &= M(\sigma) + c \cdot \frac{1 - P(\sigma b|\sigma)}{P(\sigma b|\sigma)} \leq M(\sigma) + c \cdot a \\ &\leq c \cdot a \cdot |\sigma| + M(\emptyset) + c \cdot a = c \cdot a \cdot |\sigma b| + M(\emptyset). \\ &\qquad\qquad\qquad\qquad\qquad\qquad (\text{inductive hypothesis}) \end{aligned}$$

This concludes the proof. □

Lemma 5.6 (Polynomial time bounded savings property)**.** For each polynomial time computable P-martingale N there is a polynomial time computable P-martingale M which has the savings property and succeeds on all the sequences that N succeeds on.

Proof. The proof of [6, Lemma 6] basically works in this case. The only difference is that here N is real-valued instead of rational-valued. This fact is irrelevant for the polynomial time bound. One can verify that the same definition of M as in [6, Lemma 6] yields a polynomial time P-martingale. □

5.3 The measure induced by P_β-martingales

Recall that p_β is the one-to-one mapping that sends each real in $[0, 1)$ to its unique β-expansion, and that $\widehat{P}_\beta$ is the Parry measure induced on the unit interval, i.e. $\widehat{P}_\beta = P_\beta \circ p_\beta$. Let $T_\beta \colon [0, 1] \to [0, 1)$ be the map $T_\beta(x) = \{\beta x\}$.

We derive some useful properties of the Parry measure.

Theorem 5.7. [12] Let $\beta > 1$ be a real base, then the Parry measure P_β is $L(X_\beta)$-supported.

We will use

$$\xi = \lambda \circ p_\beta^{-1}$$

for the push-forward of the Lebesgue measure on the β-shift. Of course, inequality (31) translates to

$$k' \cdot \xi(\sigma) \leq P_\beta(\sigma) \leq k \cdot \xi(\sigma) \tag{32}$$

for any $\sigma \in L(X_\beta)$. Let us say that $x \in [0, 1]$ is β-adic if x has a finite β-expansion. Clearly, β-adic numbers correspond bijectively to words in $L(X_\beta)$ not ending in 0. For instance, if β is 2.5 we have that both 2/5 and 24/25

are β-adic numbers, since their fractional β-expansions are $p_\beta(2/5) = 10^\infty$ and $p_\beta(24/25) = 210^\infty$.

We will write I_σ for the interval of real numbers in $(0,1)$ whose fractional β-expansion begins with σ. Observe that if $\sigma \notin L(X_\beta)$ then $I_\sigma = \emptyset$.

Since we will be working with the Parry measure and with β-expansions, and since the Parry measure has a closed expression in terms of Lebesgue measure, it will be helpful to know what kind of values $\lambda(I_\sigma)$ can take, given that in the non-integer case it is no longer true that $\lambda(I_\sigma) = \beta^{-|\sigma|}$.

For this purpose we introduce some new notation. For $\sigma \in L(X_\beta)$, write

$$\begin{aligned}
\mathsf{Suc}_1(\sigma) &= \{b \in \Sigma_\beta \mid \ \sigma b \in L(X_\beta)\},\\
\overline{\sigma}^+ &= \max \mathsf{Suc}_1(\sigma),\\
\mathrm{next}(\sigma) &= \min_{\leq_{\mathrm{lex}}}\{\tau \in L(X_\beta)\colon \sigma <_{\mathrm{lex}} \tau\},\\
\mathcal{L} &= \{\sigma b \in L(X_\beta)\colon b \neq \overline{\sigma}^+\}.
\end{aligned}$$

Notice that, because of Theorem 4.2, $\mathsf{Suc}_1(\sigma)$ has the form $\{1,\dots,r\}$ for some $r = \overline{\sigma}^+ \leq \lceil\beta\rceil - 1$. Also, given any $b \in \Sigma$ we have that $\sigma b \in \mathcal{L}$ if and only if some suffix of σ is a prefix of $\mathfrak{s}(\beta)$.

Let us make a remark concerning $\overline{s\restriction_i}$ for some β-expansion s. When β is an integer and $x \in (0,1)$, if $I^n_\beta(x) = [a,b)$ denotes the β-adic half-open interval of measure β^{-n} that x lies in, then the sequence $(I^n_\beta(x))_{n\in\mathbb{N}}$ cannot eventually consist of the rightmost β-adic subinterval of the previous β-adic interval. In terms of its β-expansion, it cannot eventually consist of an infinite sequence of $\beta-1$, since the rules for the construction of β-expansions mandate that $\dots a(\beta-1)^\infty$ $(a < \beta-1)$ be written $\dots(a+1)0^\infty$. The same observation is true for non-integer bases β, when the symbol identifying the rightmost β-adic subinterval of a β-adic interval is not necessarily $\lfloor\beta\rfloor - 1$. In this case, $\overline{\sigma}^+$ is used to identify the rightmost β-adic subinterval of I_σ, i.e. $I_{\sigma\overline{\sigma}^+}$, and our observation takes the following form.

Lemma 5.8. Let s be the fractional β-expansion of some real $x \in [0,1)$ (that is, $s \in p_\beta([0,1))$). Then for any natural number n, there is $i > n$ such that $s_{i+1} \neq \overline{s\restriction_i}^+$.

Observation 5.9. $\xi(\sigma) = \langle \mathrm{next}(\sigma)\rangle_\beta - \langle\sigma\rangle_\beta$.

Lemma 5.10. Let $\sigma = \sigma' b \in \mathcal{L}$. Then $\xi(\sigma) = \beta^{-|\sigma|}$.

Proof. Since $b \neq \overline{\sigma'}^+$ we have that $\sigma'(b+1) \in L(X_\beta)$, so that $\mathrm{next}(\sigma) = \sigma'(b+1)$ and by Observation 5.9, $\lambda(I_\sigma) = \xi(\sigma) = \langle\sigma'(b+1)\rangle_\beta - \langle\sigma' b\rangle_\beta = \beta^{-(|\sigma'|+1)}$ □

Observation 5.11. Let $\tau c \in \mathcal{L}$ and α be some prefix of $\mathfrak{s}(\beta)$. Then $\mathrm{next}(\tau c\alpha) = \tau(c+1)$

Define $N_\sigma = \min\{n\colon \sigma_{n+1}\dots\sigma_{|\sigma|} = \mathfrak{s}(\beta)_1\dots\mathfrak{s}(\beta)_{|\sigma|-n}\} \cup \{|\sigma|\}$.

Observation 5.12. Let $\sigma \in L(X_\beta)$, $N_\sigma > 1$ and $\tau c = \sigma\restriction_{N_\sigma}$, then $c \neq \overline{\tau}^+$.

The following lemma extends Lemma 5.10 when β is Pisot, in the sense that it completes the characterization of the values that $\xi(\sigma)$ may take.

Lemma 5.13. Let β be a Pisot real, let $\sigma \in L(X_\beta) \setminus \mathcal{L}$, and suppose $\mathfrak{s}(\beta) = r_1\dots r_m(a_1\dots a_n)^\infty$ and $\phi_j = \langle a_1\dots a_j\rangle_\beta$ for $j \leq n$. Let $\tau c = \sigma\restriction_{N_\sigma}$. Then, $\xi(\sigma) = \xi(\tau c)\xi(r_1\dots r_l)$ in case $\sigma = \tau c r_1\dots r_l$, $l \leq m$), or $\xi(\sigma) = \xi(\tau c)\xi(r_1\dots r_m a_1\dots a_k)\beta^{-ln}$ in case $\sigma = \tau c r_1\dots r_l(a_1\dots a_n)^l a_1\dots a_k$, $k \leq n$, $0 \leq l$.

Proof. Since $\sigma \notin \mathcal{L}$ some suffix of σ is a prefix of $\mathfrak{s}(\beta)$, which means either $\sigma = \tau c r_1\dots r_l$, for some $l \leq m$, or $\sigma = \tau c r_1\dots r_m(a_1\dots a_n)^l a_1\dots a_k$, for some $l \geq 0$ and $k \leq n$. Write $\psi_i = \langle r_1\dots r_i\rangle_\beta$. One then has

$$1 = \langle \mathfrak{s}(\beta)\rangle_\beta = \psi_m + \frac{1}{\beta^m}\phi_n\frac{1}{1-\beta^{-n}}. \tag{33}$$

In case $\sigma = \tau c r_1\dots r_l$, we have $\langle\sigma\rangle_\beta = \langle\tau c\rangle_\beta + \beta^{-(|\tau|+1)}\psi_l$, and by Observations 5.9 and 5.11 we have $\xi(\sigma) = \langle\tau(c+1)\rangle_\beta - \langle\sigma\rangle_\beta$, so that

$$\frac{\xi(\sigma)}{\xi(r_1\dots r_l)} = \frac{\langle\tau(c+1)\rangle_\beta - \langle\tau c\rangle_\beta - \beta^{-(|\tau|+1)}\psi_l}{1-\psi_l} = \frac{\xi(\tau c) - \xi(\tau c)\psi_l}{1-\psi_l} = \xi(\tau c),$$

where we have used that $\xi(\tau c) = \beta^{-(|\tau|+1)}$ (which follows from Observation 5.12).

For the case when $\sigma = \tau c r_1\dots r_m(a_1\dots a_n)^l a_1\dots a_k$, we have

$$\begin{aligned}\langle\sigma\rangle_\beta &= \langle\tau c\rangle_\beta + \beta^{-(|\tau|+1)}[\psi_m + \beta^{-m}(\phi_n(1+\beta^{-1}+\dots+\beta^{-(l-1)n}) + \beta^{-ln}\phi_k)] \\ &= \langle\tau c\rangle_\beta + \beta^{-(|\tau|+1)}\left[\psi_m + \beta^{-m}\left(\phi_n\frac{(\beta^{-ln}-1)}{\beta^{-n}-1} + \beta^{-ln}\phi_k\right)\right],\end{aligned}$$

and from (33) and Observations 5.9 and 5.11

$$\xi(r_1\dots r_m a_1\dots a_k) = 1 - (\psi_m + \beta^{-m}\phi_k) = \beta^{-m}\left[\frac{\phi_n}{1-\beta^{-n}} - \phi_k\right]$$

so that (using Observations 5.9 and 5.11 again)

$$\begin{aligned}\xi(\sigma) &= \langle\tau(c+1)\rangle_\beta - \langle\sigma\rangle_\beta \\ &= \xi(\tau c) - \beta^{-(|\tau|+1)}\left[\psi_m + \beta^{-m}\left(\phi_n\frac{(\beta^{-ln}-1)}{\beta^{-n}-1} + \beta^{-ln}\phi_k\right)\right] \\ &= \xi(\tau c)\left[1 - (\psi_m + \beta^{-m}\phi_k) - \beta^{-m}(\beta^{-ln}-1)\left(\phi_k - \frac{\phi_n}{1-\beta^{-n}}\right)\right] \\ &= \xi(\tau c)\xi(r_1\dots r_m a_1\dots a_k)\beta^{-ln}.\end{aligned}$$

This concludes the proof. □

Corollary 5.14. There exist positive constants d and d' such that $d \leq \xi(\sigma b \mid \sigma) \leq d'$ for any σb such that $\xi(\sigma b) > 0$.

Proof. It suffices to see that $\xi(\sigma b \mid \sigma)$ takes only finitely many values. First of all, in the case when $\sigma b, \sigma \in \mathcal{L}$ then $\xi(\sigma b \mid \sigma) = \beta^{-1}$, by Lemma 5.10. When $\sigma \in \mathcal{L}$ but $\sigma b \notin \mathcal{L}$ then $\xi(\sigma b \mid \sigma) = \xi(b)$, by Lemma 5.13. When $\sigma b \in \mathcal{L}$ but $\sigma \notin \mathcal{L}$, write $\tau c = \sigma \restriction_{N_\sigma}$. As remarked above, either $\sigma = \tau c r_1 \dots r_l$ or $\sigma = \tau c r_1 \dots r_l (a_1 \dots a_n)^l a_1 \dots a_k$. Then, by Lemma 5.13 either

$$\xi(\sigma b \mid \sigma) = \frac{\beta^{-(|\tau c|+l+1)}}{\xi(\tau c)\xi(r_1 \dots r_l)} = \frac{\beta^{-(l+1)}}{\xi(r_1 \dots r_l)}$$

(which can take only finitely many values, since $l \leq m$ and m is fixed) or

$$\xi(\sigma b \mid \sigma) = \frac{\beta^{-(|\tau c|+m+ln+k+1)}}{\xi(\tau c)\xi(r_1 \dots r_m a_1 \dots a_k)\beta^{-ln}} = \frac{\beta^{-(k+1)}}{\xi(r_1 \dots r_m a_1 \dots a_k)}$$

(which can take only finitely many values, since $k \leq n$, and n is fixed).

When neither σb nor σ are in $\mathcal{L}$, Lemma 5.13 means the conditional probability $\xi(\sigma b \mid \sigma)$ may take the following values.

- $\xi(r_1 \dots r_{l+1})/\xi(r_1 \dots r_l)$, if $\sigma = (\sigma \restriction_{N_\sigma}) r_1 \dots r_l$ for some $l \leq m-1$
- $\xi(r_1 \dots r_m a_1)/\xi(r_1 \dots r_m)$, if $\sigma = (\sigma \restriction_{N_\sigma}) r_1 \dots r_m$
- $\xi(r_1 \dots r_m a_1 \dots a_{k+1})/\xi(r_1 \dots r_m a_1 \dots a_k)$, if

  $$\sigma = (\sigma \restriction_{N_\sigma}) r_1 \dots r_m (a_1 \dots a_n)^l a_1 \dots a_k$$

  for $0 \leq l$, $k \leq n-1$
- $\xi(r_1 \dots r_m a_1)\beta^{-1}/\xi(r_1 \dots r_m a_1 \dots a_n)$, if

  $$\sigma = (\sigma \restriction_{N_\sigma}) r_1 \dots r_m (a_1 \dots a_n)^l a_1 \dots a_n$$

  for $0 \leq l$

Since these expressions may only take finitely many values (for fixed $r_1, \dots, r_m$ and fixed $a_1, \dots, a_n$), the proof is finished. □

Corollary 5.15. If β is Pisot, then $\xi(\sigma) \leq \beta^{-|\sigma|}$.

Proof. It is enough to note that $\xi(\mathfrak{s}(\beta) \restriction_k) \leq \beta^{-k}$ (this is true for any β, Pisot or not) and then use Lemma 5.13. □

The following is a straightforward consequence of Proposition 5.5 and Corollary 5.14:

Corollary 5.16. If $\beta > 1$ is Pisot and M is a P_β-martingale on $L(X_\beta)$ with the savings property, then there is c such that $M(\sigma) \leq c \cdot |\sigma| + M(\emptyset)$ for all $\sigma \in L(X_\beta)$.

Let $\beta > 1$ be Pisot. Each P_β-martingale M on $L(X_\beta)$ induces a measure μ_M on the algebra of word cylinders defined by $\mu_M([\sigma]) = M(\sigma) \cdot P_\beta(\sigma)$, for $\sigma \in L(X_\beta)$. Via Carathéodory's extension theorem this measure can be extended to a Borel measure on $\Sigma^{\mathbb{N}}$, and if μ_M is atomless (i.e. no point has positive measure), we can also think of it as a Borel measure on $[0, 1]$, which is given by $\mu_M(I_\sigma) = M(\sigma) \cdot P_\beta(\sigma)$,

We say that a martingale M is *atomless* if μ_M is atomless.

Observation 5.17. If M is a P_β-martingale with the savings property then it is atomless.

Proof. Indeed, by (32) and Corollary 5.15, there is a constant k such that for any $\sigma \in L(X_\beta)$, $P_\beta(\sigma) \leq k \cdot \beta^{-|\sigma|}$. By Corollary 5.16, there is a constant c such that for any $\sigma \in L(X_\beta)$ of length n we have $\mu_M(I_\sigma) \leq k \cdot \beta^{-n} \cdot (d \cdot n + M(\emptyset))$, and this goes to 0 as n goes to infinity. Hence μ_M is atomless. □

The cumulative distribution function associated with μ_M will be written $\mathsf{cdf}_M(x)$ for $x \in [0, 1]$. We now want to prove an analogue of the left-to-right implication of Theorem 3.6 from [3] (which is used as Theorem 3 in [6]).

Theorem 5.18. Let M be a P_β-martingale with the savings property that succeeds on the β-expansion of $z \in (0, 1)$, a non-β-adic real. Then

$$\liminf_{h \to 0} \frac{\mathsf{cdf}_M(z + h) - \mathsf{cdf}_M(z)}{h} = \infty.$$

Proof. The proof is essentially the same as in [3]. Let $g = \mathsf{cdf}_M$ and let $r > 0$. We will show that there is some $\epsilon > 0$ such that if $|h| < \epsilon$ then $|g(z+h) - g(z)| > rk'|h|$, where k' is the constant such that $k'\lambda(A) \leq \widehat{P}_\beta(A)$ from (31).

Let $(z_i^\beta)_{i \geq 1}$ be the fractional β-expansion of z. Since M succeeds on $(z_{\beta,i})_{i \geq 1}$ and has the savings property, there is some i such that, if $\rho = z^\beta \restriction_i$, then $M(\rho\tau) > r$ for any τ such that $\rho\tau \in L(X_\beta)$. Since z is not β-adic there is some $j > i$ such that $z_j^\beta \neq 0$, and by Lemma 5.8 there is some $k > j$ such that $z_{k+1}^\beta \neq \overline{z^\beta \restriction_k}^+$. Let $\epsilon = \beta^{-k-1}$. If $0 < |h| < \epsilon$ then the β-expansion of $z + h$ extends ρ. If $h > 0$ this is because $z + h < z + \beta^{-k-1}$ and β^{k+1} has the same β-expansion as z, except that z_{k+1}^β is replaced with $1 + z_{k+1}^\beta$, which at worst is $\overline{z^\beta \restriction_k}^+$. Similarly, if $h < 0$, then $z + h > z - \beta^{-k-1} > z - \beta^{-j}$ and $z - \beta^{-j}$ has the same β-expansion as z, except that z_j^β is replaced with $z_j^\beta - 1$, which at worst is 0. This means that, if $W \subseteq L(X_\beta)$ is a prefix-free set of

strings such that $\bigcup_{\sigma\in W} I_\sigma = (z, z+h)$ in case $h > 0$, or $\bigcup_{\sigma\in W} I_\sigma = (z+h, z)$ in case $h < 0$, then all strings in W extend ρ. Hence,

$$|g(z+h)-g(z)| = \sum_{\sigma\in W} M(\sigma)P_\beta(\sigma) > rk' \sum_{\sigma\in W} \xi(\sigma) = rk' \sum_{\sigma\in W} \lambda(I_\sigma) = rk'|h|.$$

□

In [3] it is shown that if f is a nondecreasing function with domain containing $[0,1]\cap\mathbb{Q}$ then $\mathsf{mart}_f\colon \{0,1\}^* \to \mathbb{R}$ defined by

$$\mathsf{mart}_f(\tau) = \frac{f(\langle\tau\rangle_2 + 2^{-|\tau|}) - f(\langle\tau\rangle_2)}{2^{-|\tau|}}$$

is a classical martingale. It is also observed in [3, Fact 3.5] that if $f(0) = 0$ then $\mathsf{cdf}_{\mathsf{mart}_f} = f$. In the next lemma we use these facts for $f = \mathsf{cdf}_M$, the cumulative distribution function of our P_β-martingale.

Lemma 5.19. Let $\beta > 1$ be Pisot. Suppose M is a P_β-martingale with the savings property. Let $N\colon \{0,1\}^* \to \mathbb{R}_{\geq 0}$ be the following (classical) martingale:

$$N(\tau) = \mathsf{mart}_{\mathsf{cdf}_M}(\tau) = \frac{\mathsf{cdf}_M(\langle\tau\rangle_2 + 2^{-|\tau|}) - \mathsf{cdf}_M(\langle\tau\rangle_2)}{2^{-|\tau|}}.$$

Suppose $s \in X_\beta$ and that there exists $x \in [0,1]$ neither β-adic, nor a dyadic rational such that $p_\beta(x) = s$. If M succeeds on s then N succeeds on the fractional binary expansion of x.

Proof. Same proof as in [6, Lemma 11], with Theorem 5.18 substituting for [6, Theorem 10] □

5.4 μ_M and cdf_M are polynomial time computable

As in [6], we show an 'almost Lipschitz' condition for cdf_M:

Proposition 5.20. Let $\beta > 1$ be Pisot and let M be a P_β-martingale on $L(X_\beta)$ with the savings property. Then there are constants $k, \epsilon > 0$ such that for every $x, y \in [0,1]$, if $y - x \leq \epsilon$ then

$$\mathsf{cdf}_M(y) - \mathsf{cdf}_M(x) \leq -k \cdot (y-x) \cdot \log(y-x).$$

Proof. We actually show that there are constants c and d such that

$$\mathsf{cdf}_M(y) - \mathsf{cdf}_M(x) \leq d \cdot (y-x) \cdot (c \cdot (1 - \log_\beta(y-x)) + M(\emptyset)) \qquad (34)$$

for $0 \leq x < y \leq 1$. Let $n \in \mathbb{N}$ be the least integer such that $\beta^{-n} < y - x$, and let

$$\Theta = \{\sigma \in L(X_\beta) \mid |\sigma| = n, I_\sigma \cap [x,y] \neq \emptyset\}.$$

So we may write $\Theta = \{\sigma_1, \ldots, \sigma_m\}$, with $\sigma_i <_{\text{lex}} \sigma_{i+1}$ for all $i < m$. Let p be the left end-point of I_{σ_1} and let q the right end-point of I_{σ_m}. Clearly both p and q are β-adic and $[x, y] \subseteq \bigcup_{\sigma\in\Theta} I_\sigma = [p, q]$.

We have

$$\begin{aligned}
\mathsf{cdf}_M(y) - \mathsf{cdf}_M(x) \leq \mathsf{cdf}_M(q) - \mathsf{cdf}_M(p) = \mu_M[p, q] &= \sum_{\sigma\in\Theta} P_\beta(\sigma) \cdot M(\sigma) \\
&\leq (c \cdot n + M(\emptyset)) \sum_{\sigma\in\Theta} P_\beta(\sigma) \qquad \text{(by Corollary 5.16)} \\
&\leq \big((c \cdot (1 - \log_\beta(y - x))\big) + M(\emptyset)) \sum_{\sigma\in\Theta} P_\beta(\sigma). \\
&\qquad\qquad (\beta^{-(n-1)} \geq y - x)
\end{aligned}$$

Since for each $\sigma \in \Theta$ we have $P_\beta(\sigma) = \widehat{P}_\beta(I_\sigma)$, from (31) we conclude $P_\beta(\sigma) \leq k \cdot \lambda(I_\sigma)$, and so $\sum_{\sigma\in\Theta} P_\beta(\sigma) \leq k \cdot \lambda([x, y]) + P_\beta(\sigma_1) + P_\beta(\sigma_m)$. From (32) and Corollary 5.15 we know that $P_\beta(\sigma_1)$ and $P_\beta(\sigma_m)$ are at most $\beta^{-n} < y - x$). Hence we conclude (34) for $d = k + 2$. □

Proposition 5.21. If $\beta > 1$ is Pisot then P_β is polynomial time computable.

Proof. Recall from (30) the closed expression for $\widehat{P}_\beta$, the measure that P_β induces on the unit interval. As before, since β is Pisot its periodic β-expansion can be written $\mathfrak{s}(\beta) = r_1 \ldots r_m (a_1 \ldots a_n)^\infty$.

Define

$$B_k = \{\tau \in \Sigma_\beta^* \colon \tau \leq_{\text{lex}} T^k(\mathfrak{s}(\beta))\}.$$

Notice that B_k is computable in linear time. Since $x \leq T_\beta^n(1)$ iff $p_\beta(x) \leq_{\text{lex}} T^n(\mathfrak{s}(\beta))$ we have, for any $\sigma \in L(X_\beta)$,

$$\widehat{P}_\beta(I_\sigma) = P_\beta(\sigma) = \sum_{j=0}^{m-1} \frac{\xi(I_\sigma)}{\beta^j} \mathbb{1}_{B_j}(\sigma) + \frac{1}{\beta^{m-1}} \sum_{j=1}^{n} \frac{\xi(I_\sigma)}{\beta^j - 1} \mathbb{1}_{B_{j+m-1}}(\sigma),$$

which is polynomial time computable because it consists of fixed-length sums of products of polynomial time computable functions, since $\xi(\cdot)$, β and the set B_k are polynomial time computable. □

Proposition 5.22. Let $\beta > 1$ be Pisot, and let M be a polynomial time computable P_β-martingale with the savings property. Then both $\mu_M \colon L(X_\beta) \to \mathbb{R}$ and $f \colon L(X_\beta) \to \mathbb{R}$ given by $f(\sigma) = \mathsf{cdf}_M(\langle\sigma\rangle_\beta)$ are computable in polynomial time.

Proof. We have $\mu_M(\sigma) = M(\sigma)P_\beta(\sigma)$. By Proposition 5.21 P_β is computable in polynomial time and M is also computable in polynomial time by hypothesis. So that by Observation 5.2 their product is computable in polynomial time.

For f we have

$$f(\sigma) = \mu_M([0, \langle\sigma\rangle_\beta]) = \sum_{i=0}^{|\sigma|-1} \sum_{b \in \mathsf{Suc}_1(\sigma\restriction_i)} \mu_M((\sigma\restriction_i)b).$$

By Proposition 5.1 membership in $\mathsf{Suc}_1(\sigma)$ is checked in linear time and we have a sum of at most $|\sigma| \cdot (1 + \lfloor\beta\rfloor)$ many terms, each of which is computable in polynomial time. □

5.5 Polynomial time randomness implies normality to Pisot bases

Lemma 5.23. Let β be a Pisot number and M be a P_β-supermartingale that is computable in polynomial time and succeeds on $s \in X_\beta$. Then there is a P_β-martingale $\widehat{M}$ computable in polynomial time that succeeds on s.

Proof. Same as in [6, Lemma 4]. Define

$$d(\sigma) = M(\sigma) - P(\sigma)^{-1} \sum_{a \in \Sigma_\beta} P(\sigma a) M(\sigma a)$$

for any $\sigma \in L(X_\beta)$. Notice that $P(\sigma a)M(\sigma a)$ is computable in polynomial time by the same argument used to show in Proposition 5.22 that μ_M is computable in polynomial time. Then $\widehat{M}(\sigma) = M(\sigma) + \sum_{\tau \prec \rho} d(\sigma)$ is a P_β-martingale computable in polynomial time. □

Lemma 5.24. Let $\beta > 1$ be Pisot. If $s \in X_\beta$ is not P_β-distributed then there is a polynomial time computable P_β-martingale which succeeds on s.

Proof. We know from Theorem 4.10 that X_β is a sofic subshift. Moreover, by Theorem 4.2 it is clearly irreducible (if $\sigma, \tau \in L(X_\beta)$ then $\sigma 0 \tau \in L(X_\beta)$). Let $(G, \mathcal{L})$ be its minimal, irreducible, right-resolving presentation, whose existence is guaranteed by Theorem 3.10) and let $\mathcal{L}^* \colon X_G \to X_\beta$ be the natural factor map induced by $\mathcal{L}$. Of course, X_G is an irreducible, 1-step Markov shift and it is shown in [9, Example 8.1.6] that $\mathcal{L}^*$ is finite-to-one[2], that is, for any $s \in X_\beta$, $\mathcal{L}^{*-1}(s)$ is a finite set. It is also shown in [9, Corollary 8.1.20] that finite-to-one factor maps between irreducible sofic shifts preserve topological entropy, where the topological entropy $h(X)$ of a subshift (X, T) is defined as[3]

$$h(X) = \sup_{T\text{-invariant } \mu} h_\mu(X).$$

[2] I thank Mike Boyle and Brian Marcus for pointing out this argument to me.

[3] The topological entropy of a dynamical system has a different, intrinsic definition not depending on measures, and our "definition" is actually a theorem known as the *variational principle*.

Thus, $h(X_G) = h(X_\beta)$. But from Theorem 2.8 we know that there is a unique invariant P such that $h(X_G) = h_P(X_G)$ where P is an irreducible, 1-step Markov measure on X_G. Furthermore, from Theorem 4.7 we know that $h(X_\beta) = h_{P_\beta}(X_\beta)$. Now, it is shown in [5, Theorem 1.1] that for any factor map between subshifts $\pi\colon X \to Y$ and any invariant measure ν on Y, there is an invariant measure μ on X such that $\nu = \mu \circ \pi^{-1}$. Hence, there is some invariant μ on X_G such that $P_\beta = \mu \circ \mathcal{L}^{*-1}$ and since factor maps cannot increase metric entropy ([19, Theorem 4.11]) we have that $h_\mu(X_G) \geq h_{P_\beta}(X_\beta) = h(X_\beta) = h(X_G) = h_P(X_G)$. But since P is the unique invariant measure of maximal metric entropy on X_G, it follows that $P = \mu$, and therefore P_β is the push-forward of P. Thus, we are under the conditions of Theorem 3.12 and we have, according to item 1, a P_β-martingale generated by a DFA which succeeds on s, and, according to item 2, a P_β-supermartingale generated by a DFA which succeeds on s. In any case, we have a P_β-supermartingale generated by a DFA which succeeds on s and whose two betting factors other than 1 are either rational or have the form $(1 - \delta p^*/(1 - p^*))$, where p^* is the conditional P_β probability on some fixed words. Since P_β is polynomial time computable, then all betting factors are polynomial time computable.

Thus, our P_β-supermartingale is of the form $L(\sigma) = p^{m_1(\sigma)} r^{m_2(\sigma)}$, where r is polynomial time computable, p is some fixed rational and $m_1(\sigma)$ and $m_2(\sigma)$ are non negative integers smaller than $|\sigma|$ and computable in time linear in $|\sigma|$, since a DFA reads its entry in linear time.

For rational p it is clear that $p^{m_1(\sigma)}$ is polynomial time computable. Now, when r is not rational, r is strictly smaller than 1, and is also computable in polynomial time. So, given n, we can compute a rational r_n in time $O(q(n))$ (q some polynomial) such that $|r_n - r| < 2^{-n}$. Then, if $\epsilon_n = r - r_n$, we have, for $m = m_2(\sigma)$,

$$|r^m - r_n^m| \leq |r^m| + |(r + \epsilon_n)^m| \leq 2 \sum_{i=1}^{m} \binom{m}{i} |\epsilon_n|^i r^{m-i} \leq 2^m |\epsilon_n| = 2^{m-n}.$$

Thus, given $|\sigma|$ and k, we can compute a rational r_n in time $O(q(n))$ for $n = m_2(\sigma) + k + 1 \leq |\sigma| + k + 1$, and we can compute $m_2(\sigma)$ in time $O(q'(|\sigma|)$ for some polynomial q'. Therefore, since exponentiation by squaring has (strictly less than) polynomial time complexity the number $r(\sigma, k) = r_n^{m_2(\sigma)}$ can be computed in $O(q''(|\sigma| + k))$ time for some polynomial q'', and satisfies $|r^{m_2} - r(\sigma, k)| < 2^{-k}$. Hence, $r^{m_2(\sigma)}$ is computable in polynomial time and so is the P_β-supermartingale L. By Lemma 5.23 there is a polynomially time computable P_β-martingale that succeeds on s. □

Theorem 5.25. Let $\beta > 1$ be Pisot and $x \in [0, 1]$ be a number such that a P_β-martingale computable in polynomial time succeeds on the β-expansion of x. Then there is a polynomial time binary martingale that succeeds on the binary expansion of x.

The proof is the same as that of [6, Theorem 14] with [6, Lemma 15] replaced with the following:

Lemma 5.26. Let $\beta > 1$ be Pisot. For any polynomial time computable P_β-martingale $M\colon L(X_\beta) \to \mathbb{R}_{\geq 0}$ with the savings property there is a classic martingale $N\colon \{0,1\}^* \to \mathbb{R}_{\geq 0}$ such that N is polynomial time computable, and whenever M succeeds on $s \in X_\beta$, and $x \in [0,1]$ is such that $p_\beta(x) = s$ then N succeeds on the fractional binary expansion of x.

Proof. By Proposition 5.22, there is a polynomial time computable function $\widehat{\mathsf{cdf}_M}\colon \Sigma_\beta^* \times \mathbb{N} \to \mathbb{Q}$ such that $|\widehat{\mathsf{cdf}_M}(\tau, i) - \mathsf{cdf}_M(\langle\tau\rangle_\beta)| \leq 2^{-i}$.

Define the classical martingale $N\colon \{0,1\}^* \to \mathbb{R}_{\geq 0}$ as $N(\tau) = (\mathsf{cdf}_M(p_2) - \mathsf{cdf}_M(p_1))/2^{-|\tau|}$, where $p_1 = \langle\tau\rangle_2$ and $p_2 = \langle\tau\rangle_2 + 2^{-|\tau|}$. N has a polynomial time computable approximation $\widehat{N}\colon \{0,1\}^* \times \mathbb{N} \to \mathbb{Q}$, defined by

$$\widehat{N}(\tau, i) = \frac{\widehat{\mathsf{cdf}_M}(\tau_2, i+2) - \widehat{\mathsf{cdf}_M}(\tau_1, i+2)}{2^{|\tau|}},$$

where for $j = 1, 2$ the string $\tau_j \in \Sigma_\beta^*$ is an approximation of p_j with error 2^{-2v-1}, for $v = i + 2 + k$. By Proposition 5.4 and the definition of N, we conclude that N is polynomial time computable. The proof that $|N(\tau) - \widehat{N}(\tau, i)| \leq 2^{-i}$ is the same as that of [6, Fact 16 in the proof of Lemma 15], using Proposition 5.20 instead of [6, Proposition 12] for the Lipschitz condition. □

We finally arrive to the main theorem of his section:

Theorem 5.27. Let $\beta > 1$ be Pisot. If the fractional binary expansion of $x \in [0,1]$ is polynomial time random then it is normal to base β.

Proof. We proceed by contradiction. Suppose that $Y \in \{0,1\}^{\mathbb{N}}$, the fractional binary expansion of x, is not polynomial time random to base β. By Lemma 5.24, there is a polynomial time computable P_β-martingale M which succeeds on $s = p_\beta(x)$, and by Lemma 5.6 there is a polynomial time computable P_β-martingale $\widetilde{M}$ with the savings property that succeeds on all the sequences M succeeds on, in particular on s. By Theorem 5.25, Y is not polynomial time random. □

Acknowledgements. This work was partially supported by grant ANPCyT-PICT-2011-0365 and UBACyT 20020110100025.

References

[1] Verónica Becher, Pablo Heiber, and Theodore A. Slaman, *A polynomial-time algorithm for computing absolutely normal numbers*, Information and Computation **232** (2013), 1–9.

[2] Anne Bertrand-Mathis, *Développement en base θ, répartition modulo un de la suite $(x\theta^n), n \geq 0$, langages codés et θ-shift*, Bull. Soc. math. France **114** (1986), 271–323.

[3] Vasco Brattka, Joseph S. Miller, and André Nies, *Randomness and differentiability*, Trans. Amer. Math. Soc. To appear.

[4] Gavin Brown, William Moran, and Charles E. M. Pearce, *A decomposition theorem for numbers in which the summands have prescribed normality properties*, J. Number Theory **24** (1986), no. 3, 259–271. MR866972 (88c:11044)

[5] Ethan M. Coven and Michael E. Paul, *Endomorphisms of irreducible subshifts of finite type*, Math. Syst. Th. **8** (1974), 167–175.

[6] Santiago Figueira and André Nies, *Feasible analysis, randomness and base invariance*, Theory of Computing Systems (2013).

[7] Bruce Kitchens, *Symbolic dynamics. one-sided, two-sided and countable state Markov shifts*, Springer, Berlin, 1998.

[8] Ker-I Ko and Harvey Friedman, *Computational complexity of real functions*, Theor. Comput. Sci. **20** (1982), 323–352.

[9] Douglas Lind and Brian Marcus, *An introduction to symbolic dynamics and coding*, Cambridge University Press, Cambridge, 1995.

[10] J. Lutz and E. Mayordomo, *Construction of an absolutely normal real number in polynomial time*, 2012. Manuscript.

[11] James R. Norris, *Markov chains*, Cambridge University Press, Cambridge, 1997.

[12] William. Parry, *On the β-expansions of real numbers*, Acta Mathematica Hungarica **11** (1960), 401–416.

[13] William Parry, *Intrinsic Markov chains*, Trans. Amer. Math. Soc. **112** (1964), 55–66.

[14] A. D. Pollington, *The Hausdorff dimension of a set of normal numbers.*, Pacific Journal of Mathematics **95** (1981), 193–204.

[15] Jan Reimann, *Randomness: Beyond Lebesgue measure*, Logic colloquium 2006, 2006.

[16] Claus-Peter Schnorr, *Zufälligkeit und Wahrscheinlichkeit*, Lecture Notes in Mathematics **218** (1971).

[17] Claus-Peter Schnorr and H. Stimm, *Endliche Automaten und Zufallsfolgen*, Acta Inf. **1** (1972), 345–359.

[18] Claude E. Shannon, *A mathematical theory of communication*, Bell System Tech. J. **27** (1948), 379–423.

[19] Peter Walters, *An introduction to ergodic theory*, Springer-Verlag, New York, 1982. Graduate Texts in Mathematics.