Density of Rational Languages Under Shift Invariant Measures

Valérie Berthé
Université Paris Cité, IRIF, CNRS, France

Herman Goulet-Ouellet
Faculty of Information Technology, Czech Technical University in Prague, Czech Republic

Dominique Perrin
LIGM, Université Gustave Eiffel, Marne-la-Vallée, France

Abstract

We study density of rational languages under shift invariant probability measures on spaces of two-sided infinite words, which generalizes the classical notion of density studied in formal languages and automata theory. The density for a language is defined as the limit in average (if it exists) of the probability that a word of a given length belongs to the language. We establish the existence of densities for all rational languages under all shift invariant measures. We also give explicit formulas under certain conditions, in particular when the language is aperiodic. Our approach combines tools and ideas from semigroup theory and ergodic theory.

2012 ACM Subject Classification Theory of computation → Regular languages

Keywords and phrases Automata theory, Symbolic dynamics, Semigroup theory, Ergodic theory

Digital Object Identifier 10.4230/LIPIcs.ICALP.2025.143

Category Track B: Automata, Logic, Semantics, and Theory of Programming

Related Version *Full Version*: `https://arxiv.org/abs/2504.16708`

Funding *Valérie Berthé*: The first author was supported by the Agence Nationale de la Recherche through the projects "SymDynAr" (ANR-23-CE40-0024) and "IZES" (ANR-22-CE40-0011), as well as by the 2024 ERC Synergy Project DynAMiCs (101167561).
Herman Goulet-Ouellet: The second author was supported by the CTU Global Postdoc Fellowship program.
Dominique Perrin: The third author was supported by the Agence Nationale de la Recherche through the project "IZES" (ANR-22-CE40-0011).

Introduction

The natural density for a language L on a finite alphabet A is defined as the limit in average (if it exists) of the probability that a word of length n belongs to L, where letters are drawn independently with equal probabilities. This notion can be traced back to the work of Berstel [5], who took inspiration from Schützenberger [37]. It has been widely studied in the context of automata theory, logic and the theory of codes [7, 36, 26, 19, 20, 10, 39, 22, 23], and also appears in ergodic theory, for instance in the work of Veech [40, 41]. Densities have also been studied in the more general case where words are drawn with a Bernoulli measure, i.e. letters are drawn independently with possibly different probabilities [7].

This paper is related with recent work on the density of group languages [9] using a more general notion of density motivated by symbolic dynamics. Let μ be a probability measure on the space $A^{\mathbb{Z}}$ of two-sided infinite words on the alphabet A. Then the density of a language L with respect to μ is the limit

$$\delta_\mu(L) = \lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} \mu(\{x \in A^{\mathbb{Z}} \mid x_0 \dots x_{i-1} \in L\}), \tag{1}$$

© Valérie Berthé, Herman Goulet-Ouellet, and Dominique Perrin;
licensed under Creative Commons License CC-BY 4.0
52nd International Colloquium on Automata, Languages, and Programming (ICALP 2025).
Editors: Keren Censor-Hillel, Fabrizio Grandoni, Joël Ouaknine, and Gabriele Puppis
Article No. 143; pp. 143:1–143:20
Leibniz International Proceedings in Informatics
LIPICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

if it exists. The classical definition discussed above corresponds to the case where μ is a Bernoulli measure. Our main result states that the density of a rational language L under a shift invariant probability measure μ always exists (Theorem 4.1). A shift invariant measure is a probability measure that behaves well with respect to left extensions of words (see Equation (5)). The proof closely combines dynamical and algebraic methods.

On the dynamical side, the proof relies mainly on the construction of a skew product between the shift space (made of two-sided infinite words) that supports the measure μ and an $\mathcal{R}$-class of the finite monoid M defining the rational language L (the transition monoid). The skew product we use is closely related to the notion of wreath product used in semigroup and automata theory [16, Chapter 1, Section 10]. We construct a natural measure on the skew product, called the weighted counting measure, which leads to a closed formula for the density under the condition that this measure is ergodic (Theorem 4.8). This generalizes known results for the case of Bernoulli measures. We also consider in Theorem 3.6 aperiodic languages (also called star-free) for which densities are proved to exist in a strong sense.

On the algebraic side, this construction relies crucially on the theory of Green's relations and on the key notion of the $\mathcal{J}$-class associated with a shift. Let the monoid M be the image of A^* by a morphism φ. The $\mathcal{J}$-class of M associated with X is the set of generators of the least ideal of M which meets the image by φ of the language of the shift X. Therefore, it can be called the minimal $\mathcal{J}$-class of the monoid M with respect to the shift X. The use of such a $\mathcal{J}$-class is a useful tool in automata theory, and appears in several different contexts. See [13] for a survey on the use of this idea, originating in the proof by Schützenberger of the characterization of aperiodic languages and [29] for an application close to the present paper.

Let us give a brief outline of the paper. In Section 1 we present preliminaries on symbolic dynamics. In Section 2 we show how to calculate the densities of ideals of A^*. In Section 3, we study the $\mathcal{J}$-class associated with a shift space in a finite monoid and derive an explicit formula for the density of aperiodic languages (Theorem 3.6). We also discuss connections with the notion of degree of an automaton and decidability results concerning the $\mathcal{J}$-class. In Section 4, we introduce the skew products and weighted counting measures, give the proof of our main result (Theorem 4.1), along with an explicit expression for the density when the weighted counting measure is ergodic (Theorem 4.8). In Section 5, we consider algebraic properties of the density (see Corollary 5.1). Lastly, Section 6 contains a discussion on the case of sofic measures, motivated by Corollary 5.1.

1 Symbolic dynamics

This section covers some necessary material from symbolic dynamics. We refer to [14] for a complete exposition on the topic, including proofs, and also to [17, 34] as alternative sources. For more specialized texts on topological dynamics and ergodic theory, we refer to [31, 42].

1.1 Topological dynamical systems

We first recall some terminology from topological dynamics. A *topological dynamical system* is a pair (X, T) of a compact metric space X and a continuous transformation $T\colon X \to X$. A subset $Y \subseteq X$ is called *invariant* if $T^{-1}(Y) \supseteq Y$. A nonempty topological dynamical system (X, T) is *minimal* if the only closed invariant subsets of X are X and $\varnothing$. Equivalently, all $x \in X$ have dense forward orbits $\{T^n(x) \mid n \geq 0\}$. As a weaker notion, (X, T) is *transitive* if there exists $x \in X$ with dense forward orbit.

Let μ be a Borel probability measure on X. The *support* of μ is the smallest closed set of measure 1. We say that μ is *invariant* if $\mu(T^{-1}U) = \mu(U)$ for every Borel set $U \subseteq X$. Note that the support of an invariant measure is a closed invariant subset.

The measure μ is called *ergodic* if it is invariant and every Borel set such that $T^{-1}(U) = U$ has measure 0 or 1. By Birkhoff's ergodic theorem, ergodicity can be interpreted as a form of asymptotic independence. More precisely, an invariant measure μ is ergodic if and only if

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} \mu(U \cap T^{-i}V) = \mu(U)\mu(V) \tag{2}$$

for every pair U, V of Borel sets (see [31]). In particular, note that the support of an ergodic measure is a transitive system. As a stronger property, the measure μ is *mixing* if

$$\lim_{n\to\infty} \mu(U \cap T^{-n}V) = \mu(U)\mu(V) \tag{3}$$

for every pair U, V of Borel sets. Every mixing measure is ergodic but the converse is false.

Let $\mathcal{M} = \mathcal{M}(X, T)$ be the set of invariant probability measures on (X, T) and $\mathcal{E} = \mathcal{E}(X, T)$ be the subset of those measures which are ergodic. We can view $\mathcal{M}$ as a subset of the dual $\mathcal{C}(X, \mathbb{R})^*$ of the space of continuous maps $X \to \mathbb{R}$. If we equip $\mathcal{C}(X, \mathbb{R})^*$ with the weak-$*$ topology, then $\mathcal{M}$ is a compact convex subset and $\mathcal{E}$ is its set of extreme points. In particular if the system (X, T) has only one invariant measure, it must be ergodic; and we call (X, T) *uniquely ergodic*. Standard results from functional analysis imply that any invariant measure can be decomposed in terms of ergodic measures. More precisely, for every invariant measure $\mu \in \mathcal{M}$, there exists a Borel probability measure τ on $\mathcal{E}$ such that

$$\mu = \int_{\mathcal{E}} \nu \, d\tau(\nu). \tag{4}$$

More details can be found in e.g. [32, Chapter 12].

1.2 Shift spaces

We now turn to symbolic dynamics. Let A be a finite alphabet. We denote by $A^{\mathbb{Z}}$ the set of two-sided infinite sequences over A equipped with the product topology, where A has the discrete topology. We denote by S the shift transformation on $A^{\mathbb{Z}}$, defined by $y = S(x)$ if $y_n = x_{n+1}$ for all $n \in \mathbb{Z}$. A *shift space* X on the alphabet A is, by definition, a closed subset of $A^{\mathbb{Z}}$ which is invariant under the shift transformation. Note that (X, S) is a topological dynamical system.

Let A^* be the free monoid on A, ε be the empty word, and $A^+ = A^* \setminus \{\varepsilon\}$. We denote by $\mathcal{L}(X)$ the set of finite words w which appear in the elements $x \in X$ and by $\mathcal{L}_n(X)$ the set of words of length n in $\mathcal{L}(X)$. For a shift space, transitivity and minimality can be characterized in terms of the language $\mathcal{L}(X)$. A shift space X is transitive if and only if for every $u, v \in \mathcal{L}(X)$, there is some $w \in \mathcal{L}(X)$ such that $uwv \in \mathcal{L}(X)$. Moreover X is minimal if and only if for every $n \geq 0$, there is an $N \geq 0$ such that every word of $\mathcal{L}_n(X)$ appears in every word of $\mathcal{L}_N(X)$. Shift spaces which are transitive are also called *irreducible*.

For words u, v of length m, n respectively, we define the right and two-sided cylinders by

$$[v]_X = \{x \in X \mid x_0 \dots x_{n-1} = v\}, \qquad [u \cdot v]_X = \{x \in X \mid x_{-m} \dots x_{n-1} = uv\}.$$

The two-sided cylinders form a clopen basis for the topology of X. For $L, K \subseteq A^*$, we define

$$[L]_X = \bigcup_{w \in L} [w]_X, \qquad [L \cdot K]_X = \bigcup_{u \in L, v \in K} [u \cdot v]_X.$$

Let μ be a Borel probability measure on a shift space X. By a slight abuse of notation, we also denote by μ the map which assigns to a word $w \in \mathcal{L}(X)$ the number $\mu([w]_X)$. With this notation, we have $\mu(\varepsilon) = 1$ and

$$\sum_{a \in A} \mu(wa) = \mu(w).$$

We extend this to subsets $L \subseteq \mathcal{L}(X)$ by letting $\mu(L) = \sum_{w \in L} \mu(w)$. When the cylinders $[u]_X$ for $u \in L$ are disjoint, we have $\mu(L) = \mu([L]_X)$. Moreover, if μ is an invariant measure, then

$$\mu(w) = \mu([w \cdot \varepsilon]_X) = \sum_{a \in A} \mu(aw). \tag{5}$$

When the cylinders $[u \cdot \varepsilon]_X$ for $u \in L$ are disjoint, we have $\mu(L) = \mu([L \cdot \varepsilon]_X)$.

Observe that the support of an invariant measure on $A^{\mathbb{Z}}$ is a shift space X of measure 1, and when the measure is ergodic, X is irreducible.

A *Bernoulli measure* is the simplest case of an ergodic measure on $A^{\mathbb{Z}}$. The values $\mu(w)$ for $w \in A^*$ are given by a morphism $\mu\colon A^* \to [0,1]$ such that $\sum_{a \in A} \mu(a) = 1$. It corresponds, in classical terms of probability theory, to a sequence $(\zeta_n)_{n \in \mathbb{N}}$ of independent identically distributed random variables, where $\zeta_n\colon x \mapsto x_n$. A Bernoulli measure is in fact mixing. Further mixing examples can be found among sofic measures, also called rational measures or hidden Markov measures [20, 11]. See Section 6 for a discussion on this topic.

Another important source of examples is the following. A *substitution* is a monoid morphism $\sigma\colon A^* \to A^*$. The substitution σ is *primitive* if there is an $n \geq 0$ such that every $b \in A$ occurs in every $\sigma^n(a)$. The shift $X(\sigma)$, called a *substitution shift*, is the set of all $x \in A^{\mathbb{Z}}$ such that all words in $\mathcal{L}(x)$ appear in some $\sigma^n(a)$, $n \geq 0$, $a \in A$. If σ is primitive then $X(\sigma)$ is minimal, and uniquely ergodic [27].

▶ **Example 1.1.** The Fibonacci substitution is the morphism $\sigma\colon \{a,b\}^* \to \{a,b\}^*$ defined by $\sigma\colon a \mapsto ab, b \mapsto a$. It is primitive and the substitution shift $X = X(\sigma)$ is called the *Fibonacci shift*. A description of its unique ergodic measure can be found e.g. in [14, Example 3.8.19].

The Fibonacci shift also belongs to the family of *Sturmian shifts*, whose definition may be recalled in [25]. More precisely, the Fibonacci shift is Sturmian of slope $(3-\sqrt{5})/2$. The following is an example of an automatic sequence (see [1]).

▶ **Example 1.2.** The primitive substitution $\sigma\colon a \mapsto ab, b \mapsto ba$ is called the *Thue–Morse substitution*. The shift $X(\sigma)$ is called the *Thue–Morse shift*. Its unique ergodic measure is described in [14, Example 3.8.20].

1.3 Density of a language

Let L be a language on A and let μ be an ergodic measure on $A^{\mathbb{Z}}$ with support X. We define the *density* of L relative to μ as

$$\delta_\mu(L) = \lim_{n \to \infty} \frac{1}{n} \sum_{i=0}^{n-1} \mu(L \cap A^i),$$

whenever the limit exists. In other words $\delta_\mu(L)$ is the Cesàro limit of $(\mu(L \cap A^n))_{n=0}^{\infty}$. When the classical limit of $(\mu(L \cap A^n))_{n=0}^{\infty}$ exists, we say that L has a density *in the strong sense*. It is known that when μ is a Bernoulli measure, the density of any rational language exists, and if μ has rational values on letters, all densities are rational numbers [5, Theorem 2.1].

Let us note some basic properties of the density. If L has a density, then $0 \leq \delta_\mu(L) \leq 1$, as $0 \leq \mu(L \cap A^i) \leq 1$ for every $i \geq 0$. Moreover, since $\mu(w) = 0$ when $w \notin \mathcal{L}(X)$, we have $\delta_\mu(L) = \delta_\mu(L \cap \mathcal{L}(X))$. Thus if L, K satisfy $L \cap \mathcal{L}(X) = K \cap \mathcal{L}(X)$ then $\delta_\mu(L) = \delta_\mu(K)$. The density is also finitely additive: if L, K have densities and $L \cap K = \varnothing$, then $L \cup K$ has density $\delta_\mu(L \cup K) = \delta_\mu(L) + \delta_\mu(K)$. Additionally, we have $\delta_\mu(A^* \setminus L) = 1 - \delta_\mu(L)$, and if $L \cup L' = A^*$ then $L \cap L'$ has density $\delta_\mu(L \cap L') = \delta_\mu(L) + \delta_\mu(L') - 1$. However, the density of an intersection $L \cap K$ might not exist, even when the density of both L and K exist.

▶ **Example 1.3.** On a fixed finite alphabet A, consider the two languages:

$$L = \{w \in A^* : |w| \equiv 1 \mod 2\}, \qquad K = \{w \in A^+ : |w| \equiv \lfloor \log_2 |w| \rfloor \mod 2\}.$$

It is clear that $\delta_\mu(L) = \delta_\mu(K) = 1/2$ (no matter the measure μ). However, it can be shown that the sequence of sums $s_n = \frac{1}{n}\sum_{i=1}^{n} \mu(L \cap K \cap A^i)$ has subsequences $(s_{2^{2n}})_{n\in\mathbb{N}}$ and $(s_{2^{2n+1}})_{n\in\mathbb{N}}$ converging to $1/3$ and $1/6$ respectively, thus $\delta_\mu(L \cap K)$ does not exist.

In the above example, the language K is not rational. When both languages are rational our main result (Theorem 4.1) implies that the density of the intersection also exists.

2 Density of ideals

In this section, we show how to calculate densities for languages that are ideals of A^* (and thus not necessarily rational). We start with *right ideals*, that is, languages L such that $LA^* = L$. The set $D = L \setminus LA^+$ is the minimal generating set of L as a right ideal, which means that $L = DA^*$ and D is contained in every other set with this property. Moreover the set D is a *prefix code*, meaning that no element of D is a proper prefix of another element of D. For example, for $a \in A$, the language $L = aA^*$ is the set of words that have the letter a as a prefix. Then one has $D = \{a\}$.

▶ **Proposition 2.1.** *Let μ be a probability measure on $A^{\mathbb{Z}}$ and L be a right ideal of A^*. Then L has a density in the strong sense and $\delta_\mu(L) = \mu(D)$ where $D = L \setminus LA^+$.*

Proof. Let X be the support of μ. Since D is a prefix code,

$$\mu(DA^* \cap A^n) = \sum_{u\in D} \mu(uA^* \cap A^n) = \sum_{u\in D, |u|\leq n} \mu(uA^{n-|u|}) = \sum_{u\in D, |u|\leq n} \mu(u)$$

where the last equality uses the fact that $\mu(uA^i) = \mu(u)$ for all i. Clearly this tends to $\mu(D)$ when $n \to \infty$. ◀

Next we establish a similar result for *left ideals* of A^*, that is, languages L such that $A^*L = L$. Similar to the right-sided case, the set $G = L \setminus A^+L$ is the minimal generating set of L as a left ideal and also a *suffix code*, i.e., no element of G is proper suffix of another.

▶ **Proposition 2.2.** *Let μ be an invariant probability measure of $A^{\mathbb{Z}}$ and L be a left ideal of A^*. Then L has a density in the strong sense and $\delta_\mu(L) = \mu(G)$ where $G = L \setminus A^+L$.*

Proof. Let X be the support of μ. Since G is a suffix code,

$$\mu(A^*G \cap A^n) = \sum_{u\in G} \mu(A^*u \cap A^n) = \sum_{u\in G, |u|\leq n} \mu(A^{n-|u|}u) = \sum_{u\in G, |u|\leq n} \mu(u),$$

where the last equality uses the fact that since μ is invariant, $\mu(A^iu) = \mu(u)$ for all i. By invariance of μ again, this tends to $\mu(G)$ when $n \to \infty$. ◀

The density of a left ideal may not exist if the measure is not invariant, as shown next.

▶ **Example 2.3.** Let $x \in A^{\mathbb{Z}}$ and let μ be the Dirac measure of x, that is, the probability measure μ on $A^{\mathbb{Z}}$ such that $\mu(U) = 1$ if $x \in U$ and 0 otherwise. Then, for $a \in A$, the density of A^*a is the frequency of a in the sequence $x^+ = x_0x_1 \cdots$. If the frequency of a does not exist in x^+, the language A^*a does not have a density. For example, if $x^+ = aba^2b^2 \cdots a^{2^n}b^{2^n} \cdots$, then the frequency of a does not exist. Indeed, the frequency of a is $1/2$ in $aba^2b^2 \cdots a^{2^n}b^{2^n}$, while it is close to $2/3$ in $aba^2b^2 \cdots a^{2^n}b^{2^n}a^{2^{n+1}}$.

Next we consider the case of a *quasi-ideal*, that is, the intersection of a left ideal and a right ideal. In this case we need the stronger assumption that μ is ergodic.

▶ **Proposition 2.4.** *Let μ be an ergodic measure on $A^{\mathbb{Z}}$, L be a right ideal of A^*, and K be a left ideal of A^*. Then $L \cap K$ has a density and $\delta_\mu(L \cap K) = \mu(D)\mu(G)$ where $D = L \setminus LA^+$ and $G = K \setminus A^+K$. Moreover, if μ is mixing, then $L \cap K$ has a density in the strong sense.*

Proof. Let X be the support of μ. Fix $u \in D$ and $v \in G$ and let $m = \max(|u|, |v|)$. Then $\mu(uA^* \cap A^*v \cap A^n) = 0$ if $n < m$ and otherwise

$$\mu(uA^* \cap A^*v \cap A^n) = \mu([u]_X \cap S^{|v|-n}[v]_X)$$

and thus using Equation (2),

$$\begin{aligned}\delta_\mu(uA^* \cap A^*v) &= \lim_{n\to\infty} \frac{1}{n}\sum_{i=0}^{n-1} \mu(uA^* \cap A^*v \cap A^i) = \lim_{n\to\infty} \frac{1}{n}\sum_{i=m}^{n-1} \mu([u]_X \cap S^{|v|-i}[v]_X) \\ &= \lim_{n\to\infty} \frac{1}{n}\sum_{i=0}^{n-1} \mu([u]_X \cap S^{-i}[v]_X) = \mu([u]_X)\mu([v]_X) = \mu(u)\mu(v).\end{aligned}$$

This shows that

$$\delta_\mu(DA^* \cap A^*G) = \sum_{u,v} \delta_\mu(uA^* \cap A^*v) = \sum_{u,v} \mu(u)\mu(v) = \mu(D)\mu(G).$$

If μ is mixing, then with similar arguments

$$\lim_{n\to\infty} \mu(uA^* \cap A^*v \cap A^n) = \lim_{n\to\infty} \mu([u]_X \cap S^{-n}[v]_X) = \mu(u)\mu(v),$$

which shows that the density exists in the strong sense. ◀

▶ **Example 2.5.** If $X = \{(ab)^\infty, (ba)^\infty\}$, then the language $L = aA^* \cap A^*b$ has density $1/2$ with respect to the unique invariant measure μ on X. However μ is not mixing and the density does not exist in the strong sense as $\mu(aA^* \cap A^*b \cap A^n)$ alternates between 0 and 1.

Next we turn to *two-sided ideals*, that is, languages L such that $A^*LA^* = L$.

▶ **Proposition 2.6.** *Let μ be an ergodic measure on $A^{\mathbb{Z}}$ with support X. Every two-sided ideal L that intersects $\mathcal{L}(X)$ has density 1 in the strong sense.*

Proof. Let $D = L \setminus LA^+$ and $G = L \setminus A^+L$. Then $L = DA^* = A^*G$ and by the formulas for right and left ideals, the density exists in the strong sense and $\delta_\mu(L) = \mu(D) = \mu(G)$. But, using the formula for quasi-ideals, we also have $\delta_\mu(L) = \mu(D)\mu(G) = \delta_\mu(L)^2$, so $\delta_\mu(L) = 0$ or 1. The fact that $L \cap \mathcal{L}(X) \neq \varnothing$ implies $\delta_\mu(L) > 0$. ◀

Let X be a shift space. We derive from Proposition 2.6 the following property for X-thin codes, where the languages L such that there exists $w \in \mathcal{L}(X)$ satisfying $A^*wA^* \cap L = \varnothing$ are called *X-thin* by the terminology of [6]. A suffix code $C \subseteq \mathcal{L}(X)$ is *X-maximal* if it is not properly included in a suffix code $D \subseteq \mathcal{L}(X)$. A dual notion holds for prefix codes. For example, every $\mathcal{L}_n(X)$, for $n \geq 1$, is an X-maximal suffix code.

▶ **Proposition 2.7.** *Let μ be an ergodic measure with support X and let $C \subseteq \mathcal{L}(X)$ be an X-maximal prefix or suffix code. If C is X-thin, then $\mu(C) = 1$.*

Proof. We assume that C is an X-maximal suffix code. Let $w \in \mathcal{L}(X)$ be such that $A^*wA^* \cap C = \varnothing$. We claim that $A^*wA^* \cap \mathcal{L}(X) \subseteq A^*C$. Indeed, let u, v be such that $uwv \in \mathcal{L}(X)$. Since C is X-maximal, either uwv has a suffix in C or is a suffix of some $c \in C$. The second case being impossible, we conclude that $uwv \in A^*C$, which proves the claim. Therefore, we have $\delta_\mu(A^*wA^*) \leq \delta_\mu(A^*C)$. Then by Propositions 2.6 and 2.2, we have $\mu(C) = \delta_\mu(A^*C) \geq \delta_\mu(A^*wA^*) = 1$. The proof for the dual statement for prefix codes works analogously. ◀

In particular, a finite prefix or prefix code is X-maximal if and only if $\mu(C) = 1$ for an ergodic measure with support X.

3 J-class of a shift space

This section is devoted to $\mathcal{J}$-classes of finite monoids associated with shift spaces. The reader who is not familiar with semigroup theory and Green's relations might want to consult textbooks such as [7, 24, 18, 21, 15, 16]. The notion of a $\mathcal{J}$-class discussed here also appeared in [28, 29], and is tangentially related with the work of Almeida on free profinite semigroups [2]. See also the survey [13] for more on the connections between Green's relations and automata theory.

3.1 Definition and first properties

Let X be a shift space on A and let $\varphi\colon A^* \to M$ be a morphism onto a finite monoid M. We first introduce two subsets of M naturally associated with X.

▶ **Definition 3.1.** *Let $K_X(M)$ be the intersection of all two-sided ideals I of M such that $I \cap \varphi(\mathcal{L}(X)) \neq \varnothing$, called the X-*minimal ideal of M*. Let $J_X(M)$ be the set of elements of M which generate $K_X(M)$ as a two-sided ideal, i.e.,*

$$J_X(M) = \{m \in M \mid MmM = K_X(M)\}.$$

It follows from the definition that $J_X(M)$ is a $\mathcal{J}$-class. Consider the quasi-order defined on M by $m \leq_{\mathcal{J}} n \iff MmM \subseteq MnM$. Note that $m \mathrel{\mathcal{J}} n$ if and only if $m \leq_{\mathcal{J}} n$ and $n \leq_{\mathcal{J}} m$. The next proposition establishes some basic properties of $J_X(M)$ when X is irreducible.

▶ **Proposition 3.2.** *Let X be an irreducible shift space on A and let $\varphi\colon A^* \to M$ be a morphism onto a finite monoid M. Then*

1. *$K_X(M)$ is an ideal of M which meets $\varphi(\mathcal{L}(X))$;*
2. *$J_X(M) = \{m \in K_X(M) \mid MmM \cap \varphi(\mathcal{L}(X)) \neq \varnothing\}$;*
3. *$J_X(M) \cap \varphi(\mathcal{L}(X))$ is the non-empty set of $\leq_{\mathcal{J}}$-minimal elements of $\varphi(\mathcal{L}(X))$;*
4. *Either $J_X(M)$ is the minimal ideal $K(M)$ of M, or $J_X(M) \cup \{0\}$ is the unique 0-minimal ideal in the quotient of M by the largest ideal of M which does not meet $\varphi(\mathcal{L}(X))$.*

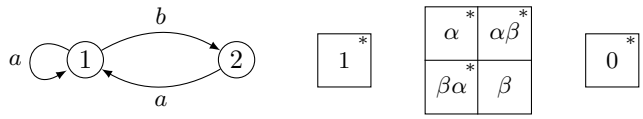


Figure 1 An automaton $\mathcal{A}$ and the eggbox picture of its transition monoid M, where α represents the transformation of the states induced by a, and β the one induced by b.

Proof. 1. Let I, J be two ideals which meet $\varphi(\mathcal{L}(X))$. Let $u, v \in \mathcal{L}(X)$ be such that $\varphi(u) \in I$ and $\varphi(v) \in J$. Since X is irreducible, there is a word w such that $uwv \in \mathcal{L}(X)$. Then $\varphi(uwv) \in I \cap J$. This proves that $K_X(M)$ is an ideal which meets $\varphi(\mathcal{L}(X))$.

2. Let $m \in J_X(M)$ and let $n \in K_X(M) \cap \varphi(\mathcal{L}(X))$. Then n is in MmM. This proves the inclusion from left to right. Conversely, let $m \in K_X(M)$ be such that $MmM \cap \varphi(\mathcal{L}(X)) \neq \varnothing$. Then, the ideal MmM generated by m is contained in $K_X(M)$ and it meets $\varphi(\mathcal{L}(X))$. This implies $MmM = K_X(M)$ and therefore m is in $J_X(M)$.

3. Let $\varphi(u)$ be $\leq_{\mathcal{J}}$-minimal in $\varphi(\mathcal{L}(X))$. Fix $v \in \mathcal{L}(X)$. By irreducibility there is $w \in \mathcal{L}(X)$ such that $uwv \in \mathcal{L}(X)$. It follows that $\varphi(uwv) \leq_{\mathcal{J}} \varphi(u)$, which by minimality implies $\varphi(u) \leq_{\mathcal{J}} \varphi(uwv) \leq_{\mathcal{J}} \varphi(v)$. This shows that $\varphi(u)$ generates an ideal contained in every ideal which meets $\mathcal{L}(X)$, thus $\varphi(u) \in J_X(M)$.

4. First assume that $K(M) \cap \varphi(\mathcal{L}(X)) \neq \varnothing$. Since the elements of $K(M)$ are $\mathcal{J}$-minimal in all of M, it follows from part 3 that $J_X(M) = K(M)$.

Assume next that $K(M) \cap \varphi(\mathcal{L}(X)) = \varnothing$. Let $I = \{s \in M \mid MsM \cap \varphi(\mathcal{L}(X)) = \varnothing\}$. It is clear that I is the largest ideal of M which does not meet $\varphi(\mathcal{L}(X))$. It is non-empty since it contains the minimal ideal $K(M)$ by assumption. Let us show that M/I is a prime monoid (see [7, Section 1.12]). Take $s, t \in M/I$ which are $\neq 0$. Then there exist x_1, x_2, y_1, y_2 such that $x_1 s x_2 = \varphi(u)$ and $y_1 t y_2 = \varphi(v)$, where $u, v \in \mathcal{L}(X)$. By irreducibility there is w such that $uwv \in \mathcal{L}(X)$, hence $x_1 s x_2 \varphi(w) y_1 t y_2 = \varphi(uvw)$, and in particular $s x_2 \varphi(w) y_1 t \neq 0$. This shows that M/I is prime, and thus it admits a unique 0-minimal ideal, by [7, Proposition 1.12.9]. Finally notice that any non-zero element s of the quotient M/I must satisfy $xsy = \varphi(u)$ for some $x, y \in M$, thus it is $\mathcal{J}$-above some element of $J_X(M)$. This implies that $J_X(M) \cup \{0\}$ is the 0-minimal ideal of M/I. ◀

▶ **Example 3.3.** Let $\mathcal{A}$ be the automaton depicted, alongside its transition monoid, in Figure 1 and X be the Fibonacci shift (Example 1.1). Let $\varphi\colon A^* \to M$ be the transition morphism of $\mathcal{A}$. Then $\varphi^{-1}(0) = A^* bb A^*$, and since $\mathcal{L}(X)$ does not contain bb, it follows that $J_X(M)$ is the $\mathcal{J}$-class of $\alpha = \varphi(a)$. In this example, the automaton $\mathcal{A}$ is the minimal automaton of the X-maximal prefix code $C = \{a, ba\}$.

The following result links the $\mathcal{J}$-class $J_X(M)$ with the densities of the languages recognized by M. Roughly speaking, it shows that the $\mathcal{J}$-class $J_X(M)$ concentrates the density.

▶ **Proposition 3.4.** *Let μ be an ergodic measure on $A^{\mathbb{Z}}$ with support X. Let $\varphi\colon A^* \to M$ be a morphism onto a finite monoid M. Then the density of $L = \varphi^{-1}(J_X(M))$ exists in the strong sense and $\delta_\mu(L) = 1$.*

Proof. It follows from Proposition 3.2 that $K_X(M) \cap \varphi(\mathcal{L}(X)) = J_X(M) \cap \varphi(\mathcal{L}(X))$, hence $\delta_\mu(\varphi^{-1}(J_X(M))) = \delta_\mu(\varphi^{-1}(K_X(M)))$. But the density of $K_X(M)$ exists in the strong sense and is 1 by Proposition 2.6. ◀

Let us highlight a simple consequence of this.

▶ **Corollary 3.5.** *Let μ be an ergodic measure with support a shift space X. Let $\varphi\colon A^* \to M$ be a morphism onto a finite monoid M. For every $m \notin J_X(M)$, the density of $\varphi^{-1}(m)$ exists in the strong sense and is 0.*

Proof. Indeed, let $L = \varphi^{-1}(m)$, $L' = \varphi^{-1}(J_X(M))$. Then as $L \subseteq A^* \setminus L'$ and the density of L' has density 1 in the strong sense,

$$0 = 1 - \lim_{n\to\infty} \mu(L' \cap A^n) = \lim_{n\to\infty} \mu(A^n \setminus L') \geq \lim_{n\to\infty} \mu(L \cap A^n) = \delta_\mu(L). \qquad \blacktriangleleft$$

3.2 On aperiodic languages

The next theorem concerns the density of aperiodic languages. Recall that a language is *aperiodic* if it can be recognized by an aperiodic monoid. By Schützenberger's theorem [38], aperiodic languages are precisely the *star-free* languages, as well as the languages defined by first-order formulas (see [30, Theorem 4.1]). The density of aperiodic languages was also studied, from a logic perspective, by Lynch [26]. Lynch's definition of density involves sequences of Bernoulli measures, thus it is both more general than our definition (since we use only one measure) and much more particular, since we use more general measures than Bernoulli measures.

▶ **Theorem 3.6.** *An aperiodic language L has a density with respect to an ergodic measure μ. More precisely, there exist a finite number of pairs $(D_i, G_i)_{1\leq i\leq k}$ of a prefix code and a suffix code such that $L \cap \mathcal{L}(X) = \bigcup_{i=1}^k (D_i A^* \cap A^* G_i) \cap \mathcal{L}(X)$ and*

$$\delta_\mu(L) = \sum_{i=1}^k \mu(D_i)\mu(G_i). \tag{6}$$

Moreover, if the measure is mixing, then the density exists in the strong sense.

Proof. Let μ be an ergodic measure and $\varphi\colon A^* \to M$ be a morphism onto an aperiodic monoid M. Let $J = J_X(M)$, let $m \in M$ and let $L = \varphi^{-1}(m)$. It is enough to prove that Equation (6) holds for such L. If $m \notin J$, then $\delta_\mu(L) = 0$ in the strong sense by Corollary 3.5, so we may assume that $m \in J$. We claim that

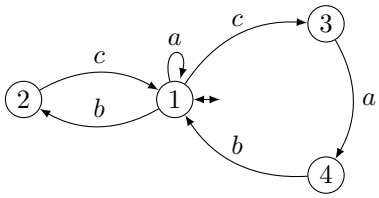
$$L \cap \mathcal{L}(X) = LA^* \cap A^*L \cap \mathcal{L}(X).$$

Indeed, assume that $u \in (LA^* \cap A^*L) \cap \mathcal{L}(X)$. Then $\varphi(u) \in mM \subseteq K_X(M)M = K_X(M)$, and thus $\varphi(u) \in J$ by the second part of Proposition 3.2. But then $\varphi(u) \in mM \cap Mm \cap J$ and since M is aperiodic, we have $mM \cap Mm \cap J = \{m\}$. It follows that $\varphi(u) = m$ and $u \in L$. This concludes the proof of the claim, as the other inclusion is trivial. Finally, we can use Proposition 2.4 to deduce that $\delta_\mu(L) = \mu(D)\mu(G)$, where $D = L \setminus LA^+$ and $G = L \setminus A^+L$, and that the density exists in the strong sense if μ is mixing. ◀

Let us give two examples to illustrate Theorem 3.6, starting with an aperiodic language.

▶ **Example 3.7.** Let $L = \{ab, ba\}^*$, whose minimal automaton $\mathcal{A}$ may be found in [33, Chapter 4, Example 2.1]. Let M stand for its transition monoid. Note that L is aperiodic and thus star-free, which is not obvious. A star-free expression of L may be found in [33, Chapter 4, Example 2.1]. Let X be Thue–Morse shift and let μ be the unique invariant measure on X (Example 1.2). The $\mathcal{J}$-class $J_X(M)$ is the $\mathcal{J}$-class of α^2 in [33, Chapter 4, Example 2.1]. Combining Equation (6) with the expression of L in [33, Chapter 4, Example 2.1], we get

$$\delta_\mu(L) = \mu((ab)^+b \cup (ba)^+a)\mu(a(ab)^+ \cup b(ba)^+).$$



$\alpha\beta\gamma^*$	α	$\alpha\beta$
$\beta\gamma$	$\beta\gamma\alpha^*$	β
γ	$\gamma\alpha$	$\gamma\alpha\beta^*$

Figure 2 The minimal automaton of B^* and the $\mathcal{J}$-class $J_X(M)$ from Example 3.8, where α, β and γ are the transformations induced respectively by a, b and c. The $\mathcal{H}$-classes in yellow represent elements in the image of B^*.

With the values of μ from [14, Example 3.8.20], we find

$$\mu((ab)^+b \cup (ba)^+a) = \mu(\{abb, ababb, baa, babaa\}) = 1/2$$

and similarly $\mu(a(ab)^+ \cup b(ab)^+) = 1/2$. Thus $\delta_\mu(L) = 1/4$.

The next example is a group language whose intersection with the shift space behaves like an aperiodic language.

▶ **Example 3.8.** Let X be the orbit of the periodic sequence $x = (abc)^\infty$. Thus X consists of three elements and has for unique ergodic measure the uniform probability measure μ. Let $\varphi\colon A^* \to \mathbb{Z}/2\mathbb{Z}$ be defined by $\varphi(a) = 0$ and $\varphi(b) = \varphi(c) = 1$. Consider the rational language $L = \varphi^{-1}(0)$. Then one sees

$$L \cap \mathcal{L}(X) = (abc)^*\{\varepsilon, a\} \cup (bca)^*\{\varepsilon, bc\} \cup (cab)^*\{\varepsilon\},$$

and thus $\delta_\mu(L) = \frac{1}{3}\left(1 + \frac{1}{3} + \frac{1}{3}\right) = \frac{5}{9}$.

The same result can be obtained using Theorem 3.6. First, observe that L has the same intersection with $\mathcal{L}(X)$ as the language B^* where $B = \{a, bc, cab\}$. The language B^* is recognized by the automaton $\mathcal{A}$ depicted in Figure 2. The transition monoid of $\mathcal{A}$ is an aperiodic monoid M with 27 elements, hence B^* turns out to be star-free. The $\mathcal{J}$-class $J_X(M)$, depicted in Figure 2, consists of 9 elements, 5 of which belong to the image of B^* (indicated in yellow). Taking for instance $m = \beta\gamma = \varphi(bc)$, we find that $\varphi^{-1}(m)A^* = bcA^*$ and $A^*\varphi^{-1}(m) = A^*bc$, and thus $\delta_\mu(\varphi^{-1}(m)) = \mu(bc)^2 = 1/9$ by Equation (6); likewise $\delta_\mu(\varphi^{-1}(m)) = 1/9$ for all other elements $m \in J_X(M)$ using Equation (6). Thus we recover the fact that $\delta_\mu(B^*) = \delta_\mu(B^* \cap \mathcal{L}(X)) = \delta_\mu(L \cap \mathcal{L}(X)) = 5/9$.

3.3 Decidability and degree

Next we consider the question of whether membership in $J_X(M)$ is decidable. We say that a shift space X is *rationally decidable* if the emptiness of $\mathcal{L}(X) \cap L$ is decidable for every rational language L. Sofic shifts are obviously rationally decidable. The class of rationally decidable shift spaces also includes all substitution shifts by a result of [35, Lemma 3] (see also [4, Lemma 3.15] and [12]).

▶ **Proposition 3.9.** *Let X be a rationally decidable shift space. Then for every morphism $\varphi\colon A^* \to M$, there is an algorithm which decides which elements $m \in M$ are in $J_X(M)$.*

Proof. We have $m \in K_X(M)$ if and only if $m \in MnM$ for every $n \in M$ such that $\varphi^{-1}(n) \cap \mathcal{L}(X) \neq \varnothing$. Since X is rationally decidable, this is decidable for every $n \in M$. Therefore, we can decide whether $m \in K_X(M)$, and then whether $m \in J_X(M)$. ◀

The $\mathcal{J}$-class is also linked with the following notion of degree. Let $\mathcal{A}$ be an automaton on A and $\varphi\colon A^* \to M$ be its transition morphism. The minimal rank of the elements of $\varphi(\mathcal{L}(X))$ viewed as partial mappings is called the X-*degree* of the automaton, denoted $d_X(\mathcal{A})$. Note that [28, Proposition 3.2] states that $J_X(M)$ contains all elements of rank $d_X(\mathcal{A})$. For instance the automaton $\mathcal{A}$ from Example 3.8 has X-degree 2, while the one in Example 3.3 has X-degree 1. This result also implies that $d_X(\mathcal{A})$ is computable if and only if membership in $J_X(M)$ is decidable (cf. Proposition 3.9 for cases where the latter is decidable).

The notion of X-degree has also been studied in relation with codes [6, 3]. Let X be an irreducible shift, let $C \subseteq \mathcal{L}(X)$ be a rational prefix code and let $\mathcal{A}$ be the minimal automaton of C^*. If C is X-maximal, then its X-degree (that is, $d_X(\mathcal{A})$) is larger than or equal to 1. When C is a finite X-maximal bifix code (i.e. both prefix and suffix), then the X-degree of C can also be defined in terms of the number of *parses* of elements of $\mathcal{L}(X)$ [6].

4 Existence of the density

The goal of this section is to establish the existence of density of regular languages under all invariant measures. More precisely, we will prove the following, which is our main theorem.

▶ **Theorem 4.1.** *Let μ be an invariant measure on $A^{\mathbb{Z}}$. Then every rational language L has a density with respect to μ.*

The main ingredient for the proof is the dynamical system defined as follows, called a skew product. Let X be an irreducible shift space and $\varphi\colon A^* \to M$ a morphism onto a finite monoid. Fix R an $\mathcal{R}$-class of $J_X(M)$ such that $R \cap \varphi(\mathcal{L}(X)) \neq \varnothing$. Let M act on the right of $R \cup \{0\}$ by $r \cdot m = rm$ if $rm \in R$ and $r \cdot m = 0$ otherwise. The specific choice of an $\mathcal{R}$-class class R plays no role here, as long as R satisfies $R \cap \varphi(\mathcal{L}(X)) \neq \varnothing$.

▶ **Definition 4.2.** *The skew product $R \cup \{0\} \rtimes X$ is the system $((R \cup \{0\}) \times X, T)$ where*

$$T(r, x) = (r \cdot \varphi(x_0), Sx).$$

Note that the projection $\pi_X\colon (R \cup \{0\}) \times X \to X$ satisfies $\pi_X \circ T = S \circ \pi_X$. It follows that if ν is an invariant (respectively ergodic) measure on $(R \cup \{0\}) \rtimes X$, then $\nu \circ \pi_X^{-1}$ is an invariant (respectively ergodic) measure on X. When $M = G$ is a group, then $R = G$ and $G \rtimes X$ forms a subsystem of $(G \cup \{0\}) \rtimes X$, which in effect means we can get rid of 0. In particular we recover the type of skew products studied in the preprint [9]. Next, we introduce a natural probability measure on $(R \cup \{0\}) \rtimes X$ induced by a measure μ on X. In the group case, we recover the product of μ with the normalized counting measure on the group considered in [9].

▶ **Definition 4.3.** *Fix an invariant probability measure μ on X. The* weighted counting measure *on the skew product $(R \cup \{0\}) \rtimes X$ is the Borel measure ν defined by $\nu(\{0\} \times X) = 0$, and for every $u, v \in \mathcal{L}(X)$ and $r \in R$,*

$$\nu(\{r\} \times [u \cdot v]_X) = \frac{1}{d} \sum_{s, s\varphi(u) = r} \mu(G_s uv),$$

where d is the cardinality of the $\mathcal{H}$-classes of $J_X(M)$ and G_s, for $s \in R$, is the suffix code such that $\varphi^{-1}(Ms) = A^* G_s$.

That ν is a well-defined Borel measure on $(R \cup \{0\}) \rtimes X$ follows from Carathéodory's extension theorem, since the above formula defines a pre-measure on the Boolean algebra of clopen sets of $(R \cup \{0\}) \times X$. Note that if $r \mathcal{H} s$, then $G_r = G_s$. For an $\mathcal{H}$-class $H \subseteq R$, let

G_H be the common value of G_r for $r \in H$. The following technical lemma will be useful. In its proof, we use the following notion: we call a monoid M *stable* if for every $s, t \in M$, the implications $s \mathcal{J} st \implies s \mathcal{R} st$ and $s \mathcal{J} ts \implies s \mathcal{L} ts$ hold. Every finite monoid is stable [18, Lemma 1.1, Chapter V].

▶ **Lemma 4.4.** *The union $G = \bigcup_{H \subseteq R} G_H$ over all $\mathcal{H}$-classes of R is a suffix code such that $\mu(G) = 1$.*

Proof. The fact that G is a suffix code is a consequence of the fact that the monoid M is stable. Moreover, G is X-maximal since, for all $w \in \mathcal{L}(X)$, there exists u such that $\varphi(uw) \in J_X(M)$, by irreducibility and the fact that $R \cap \varphi(\mathcal{L}(X)) \neq \varepsilon$. Note that a word $w \in \mathcal{L}(X)$ such that $\varphi(w) \in J_X(M)$ cannot be a proper factor of any element of G, since G is a suffix code. Thus there must exist $w \in \mathcal{L}(X)$ such that $A^*wA^* \cap G = \varnothing$ (simply fix $u \in G$ and then take $w \in \mathcal{L}(X) \cap A^+u$). It follows that $\mu(G) = 1$ by Proposition 2.7. ◀

Next we establish some properties of the weighted counting measure.

▶ **Proposition 4.5.** *The weighted counting measure is an invariant probability measure on the skew product $(R \cup \{0\}) \rtimes X$ which satisfies $\nu \circ \pi_X^{-1} = \mu$.*

Proof. First we show that ν is invariant under the map T. Note that it suffices to check invariance for sets of the form $B = \{r\} \times [u \cdot v]_X$. First, assume that $u \neq \varepsilon$. If we let $u = u'a$, $a \in A$, then $T^{-1}(\{r\} \times [u \cdot v]_X)$ can be expressed as the disjoint union

$$T^{-1}(\{r\} \times [u \cdot v]_X) = \bigcup_{s, s\varphi(a)=r} \{s\} \times [u' \cdot av]_X,$$

and from the definition of ν we find

$$\begin{aligned}\nu(T^{-1}(\{r\} \times [u \cdot v]_X)) &= \sum_{s, s\varphi(a)=r} \nu(\{s\} \times [u' \cdot av]_X) = \sum_{s, s\varphi(a)=r} \frac{1}{d} \sum_{s', s'\varphi(u')=s} \mu(G_{s'}u'av) \\ &= \frac{1}{d} \sum_{s, s\varphi(u)=r} \mu(G_s uv) = \nu(\{r\} \times [u \cdot v]_X).\end{aligned}$$

Next let us treat the case where $u = \varepsilon$, or in other words where $B = \{r\} \times [v]_X$. Observe that

$$T^{-1}(\{r\} \times [v]_X) = \bigcup_{s\varphi(a)=r} \{s\} \times [av]_X, \qquad [G_r v \cdot \varepsilon]_X = \bigcup_{s\varphi(a)=r} [G_s av \cdot \varepsilon]_X,$$

where the unions are taken over pairs $(s, a) \in R \times A$ such that $s\varphi(a) = r$. Since $G = \bigcup_{s \in R} G_s$ is a suffix code by Lemma 4.4, the second union in the above equation is disjoint (as is the first, for obvious reasons). Thus we have

$$\begin{aligned}\nu(T^{-1}(\{r\} \times [v]_X)) &= \sum_{s\varphi(a)=r} \nu(\{s\} \times [av]_X) = \sum_{s\varphi(a)=r} \frac{1}{d}\mu(G_s av) \\ &= \frac{1}{d}\mu(G_r v) = \nu(\{r\} \times [v]_X).\end{aligned}$$

Finally, let us show that $\nu(R \times U) = \mu(U)$ for every Borel set U. Using Lemma 4.4 together with the invariance of μ, we have

$$\nu(R \times U) = \sum_{r \in R} \mu([G_r \cdot \varepsilon]_X \cap U)/d = \sum_{H \subseteq R} \mu([G_H \cdot \varepsilon]_X \cap U) = \mu([G \cdot \varepsilon]_X \cap U) = \mu(U),$$

where the last equality follows since $\mu([G \cdot \varepsilon]_X) = \mu(G) = 1$. Taking $U = X$, we find that $\nu(R \times X) = \mu(X) = 1$, which shows that ν is indeed a probability measure. ◀

The following lemma will also be needed for the proof of Theorem 4.1. Its proof uses standard arguments from ergodic theory. We include it for the sake of completeness.

▶ **Lemma 4.6.** *For every ergodic measure μ on X, there exists an ergodic measure $\bar{\mu}$ on $(R \cup \{0\}) \rtimes X$ such that $\bar{\mu}(\{0\} \times X) = 0$ and $\bar{\mu} \circ \pi_X^{-1} = \mu$.*

Proof. Let $\mathcal{M}$ be the set of all invariant probability measures on $(R \cup \{0\}) \rtimes X$ and $\mathcal{M}_\mu$ the subset of those measures ζ such that $\zeta(\{0\} \times X) = 0$ and $\zeta \circ \pi_X^{-1} = \mu$. Note that $\mathcal{M}_\mu$ contains the weighted counting measure by Proposition 4.5 and thus $\mathcal{M}_\mu \neq \varnothing$. Moreover, $\mathcal{M}_\mu$ is a closed convex subspace of $\mathcal{M}$, so it must contain an extreme point $\bar{\mu}$ by the Krein–Milman theorem. Let us show that $\bar{\mu}$ is also an extreme point of $\mathcal{M}$. Suppose that $\bar{\mu} = s\zeta' + (1-s)\zeta''$ for some $\zeta', \zeta'' \in \mathcal{M}$ and $0 < s < 1$. It is clear that both ζ' and ζ'' must give to $\{0\} \times X$ zero measure, and the fact that μ is an extreme point in the convex set of invariant measures on X implies that $\zeta' \circ \pi_X^{-1} = \zeta'' \circ \pi_X^{-1} = \mu$. Thus $\zeta', \zeta'' \in \mathcal{M}_\mu$, contradicting the fact that $\bar{\mu}$ is an extreme point of $\mathcal{M}_\mu$. Since the extreme points of $\mathcal{M}$ are precisely the ergodic measure on $(R \cup \{0\}) \rtimes X$, we are done. ◀

The next proposition is a more precise form of Theorem 4.1 for the case where μ is ergodic. Note that the statement uses Lemma 4.6 implicitly for the existence of $\bar{\mu}$.

▶ **Proposition 4.7.** *Let μ be an ergodic measure on $A^{\mathbb{Z}}$ with support a shift space X. Let $\varphi\colon A^* \to M$ be a morphism onto a finite monoid and let $L = \varphi^{-1}(m)$ for some $m \in M$. Fix an $\mathcal{R}$-class R of the $\mathcal{J}$-class $J_X(M)$. We set the notation $U_{r,V} = \{r\} \times V$. Then $\delta_\mu(L) = 0$ if $m \notin J_X(M)$ and otherwise*

$$\delta_\mu(L) = \sum_{r, rm \in R} \bar{\mu}(U_{r,[L]_X})\bar{\mu}(U_{rm,X}) \tag{7}$$

where $\bar{\mu}$ is any ergodic measure on $(R \cup \{0\}) \rtimes X$ such that $\bar{\mu}(\{0\} \times X) = 0$ and $\bar{\mu} \circ \pi_X^{-1} = \mu$.

Proof. If $m \notin J_X(M)$, then $\delta_\mu(L) = 0$ by Corollary 3.5. We may now assume that $m \in J_X(M)$. Let C be the prefix code such that $LA^* = CA^*$. For $i \geq 0$, let

$$C_{\leq i} = \{u \in C \mid |u| \leq i\}, \quad C_{>i} = \{u \in C \mid |u| > i\}.$$

We claim that for every $r \in R$ such that $rm \in R$, one has

$$U_{r,[L \cap A^i]_X} = U_{r,[C_{\leq i}]_X} \cap T^{-i}(U_{rm,X}).$$

The left-to-right inclusion is obvious. For the converse, we take x such that $(r, x) \in U_{r,[C_{\leq i}]_X} \cap T^{-i}(U_{rm,X})$. This means that there exists some j, with $0 \leq j \leq i$, such that $\varphi(x_0 \cdots x_{j-1}) = m$ and furthermore $rm' = rm$ where $m' = \varphi(x_0 \cdots x_{i-1})$. Clearly $m \geq_{\mathcal{R}} m'$, hence $m \mathrel{\mathcal{R}} m'$ by Proposition 3.2 and stability of M. Moreover our choice of r guarantees that $rm \mathrel{\mathcal{J}} m$, thus $rm \mathrel{\mathcal{L}} m$ by stability. By Green's lemma, $x \mapsto rx$ is a bijection between the $\mathcal{R}$-classes of m and rm. Since $rm = rm'$ it follows that $m = m'$, i.e. $\varphi(x_0 \cdots x_{i-1}) = m$. Thus $x \in [L \cap A^i]_X$ which proves the claim.

As a result, we have

$$\mu(L \cap A^i) = \bar{\mu}(R \times [L \cap A^i]_X) = \sum_{r, rm \in R} \bar{\mu}\Big(U_{r,[C_{\leq i}]_X} \cap T^{-i}(U_{rm,X})\Big).$$

Next we claim that for $\epsilon > 0$, there is $i_0 \geq 0$ such that $\bar{\mu}(U_{r,[C_{>i_0}]_X}) < \epsilon$ for every $r \in R$. First observe that

$$\sum_{r \in R} \bar{\mu}(U_{r,[C_{>i_0}]_X}) = \bar{\mu}(R \times [C_{>i_0}]_X) = \mu([C_{>i_0}]_X),$$

since $\bar{\mu}$ projects to μ. Moreover since C is a prefix code, the cylinders $[c]_X$ for $c \in C$ are all disjoint, and thus we may write

$$\mu([C_{>i_0}]_X) = \mu(C_{>i_0}) = \sum_{i>i_0} \sum_{c \in C \cap A^i} \mu(c).$$

But note that this is a tail of the series

$$\sum_{i \geq 0} \sum_{c \in C \cap A^i} \mu(c) = \mu(C) \leq 1,$$

and thus the claim immediately follows.

Using the above claim together with the ergodicity of $\bar{\mu}$, this gives

$$\begin{aligned}
\delta_\mu(L) &= \lim_{n\to\infty} \frac{1}{n} \sum_{i=i_0}^{n-1} \mu(L \cap A^i) = \sum_{r,rm \in R} \lim_{n\to\infty} \frac{1}{n} \sum_{i=i_0}^{n-1} \bar{\mu}\big(U_{r,[C_{\leq i}]_X} \cap T^{-i}(U_{rm,X})\big) \\
&\geq \sum_{r,rm \in R} \lim_{n\to\infty} \frac{1}{n} \sum_{i=i_0}^{n-1} \bar{\mu}\big(U_{r,[C]_X} \cap T^{-i}(U_{rm,X})\big) - \epsilon \\
&\geq \sum_{r,rm \in R} \bar{\mu}(U_{r,[L]_X})\bar{\mu}(U_{rm,X}) - \epsilon.
\end{aligned}$$

On the other hand, we have

$$\begin{aligned}
\delta_\mu(L) &= \sum_{r,rm \in R} \lim_{n\to\infty} \frac{1}{n} \sum_{i=i_0}^{n-1} \bar{\mu}\big(U_{r,[C_{\leq i}]_X} \cap T^{-i}(U_{rm,X})\big) \\
&\leq \sum_{r,rm \in R} \lim_{n\to\infty} \frac{1}{n} \sum_{i=i_0}^{n-1} \bar{\mu}\big(U_{r,[C]_X} \cap T^{-i}(U_{rm,X})\big) \leq \sum_{r,rm \in R} \bar{\mu}(U_{r,[L]_X})\bar{\mu}(U_{rm,X}),
\end{aligned}$$

concluding the proof. ◀

In order to conclude the proof of Theorem 4.1, we need to deduce the existence of densities for a general invariant measure. This is done using the ergodic decomposition discussed at the end of Section 1.1.

Proof of Theorem 4.1. Let L be a rational language and μ be an invariant measure. Let $\mathcal{E}$ be the set of ergodic measure on the support of μ and consider the ergodic decomposition of μ as in Equation (4). By Proposition 4.7, $\delta_\nu(L)$ exists for every $\nu \in \mathcal{E}$. Finally by the dominated convergence theorem

$$\begin{aligned}
\delta_\mu(L) &= \lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} \mu(L \cap A^i) = \lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} \int_{\mathcal{E}} \nu(L \cap A^i)\, d\tau(\nu) \\
&= \int_{\mathcal{E}} \lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} \nu(L \cap A^i)\, d\tau(\nu) = \int_{\mathcal{E}} \delta_\nu(L)\, d\tau(\nu).
\end{aligned}$$

◀

When the weighted counting measure is ergodic, we can apply Proposition 4.7 to obtain formulas based on the density of ideals, which are easily computable (Section 2). This generalizes a result of [7] where the same formula is proved for a Bernoulli measure.

▶ **Theorem 4.8.** *Let μ be an ergodic measure on $A^{\mathbb{Z}}$ with support a shift space X. Let $\varphi\colon A^* \to M$ be a morphism onto a finite monoid. Fix R an $\mathcal{R}$-class of $J_X(M)$. If the weighted counting measure is an ergodic measure on the skew product $(R \cup \{0\}) \rtimes X$, then, for every $m \in J$, the density of $L = \varphi^{-1}(m)$ is*

$$\delta_\mu(L) = \delta_\mu(A^*L)\delta_\mu(LA^*)/d, \tag{8}$$

where d is the cardinality of the $\mathcal{H}$-classes of $J_X(M)$.

Proof. Let ν be the weighted counting measure on $(R \cup \{0\}) \rtimes X$. Let C be the prefix code such that $LA^* = CA^*$. Then Equation (7) reduces to

$$\begin{aligned}\delta_\mu(L) &= \sum_{r,rm\in R} \nu(U_{r,[L]_X})\nu(U_{rm,X}) = \frac{1}{d^2}\sum_{r,rm\in R} \mu([G_r \cdot C]_X)\mu(G_{rm}) \\ &= \frac{1}{d^2}\mu(G_m)\sum_{r,rm\in R} \mu([G_r \cdot C]_X) = \frac{1}{d}\mu(G_m)\sum_{H\subseteq R} \mu([G_H \cdot C]_X)\end{aligned}$$

where H runs over the $\mathcal{H}$-classes of R and G_H is the common value of G_r for the d elements $r \in H$. By Lemma 4.4 and since C is a prefix code,

$$\sum_{H\subseteq R} \mu([G_H \cdot C]_X) = \mu([C]_X) = \mu(C)$$

and therefore by Propositions 2.1 and 2.2

$$\delta_\mu(L) = \frac{1}{d}\mu(G_m)\mu(C) = \frac{1}{d}\delta_\mu(A^*L)\delta_\mu(LA^*).$$ ◀

There are known examples where the skew product has more than one ergodic measure. In Example 3.8, the skew product of X with $\mathbb{Z}/2\mathbb{Z}$ has two orbits, each one being the support of an ergodic measure, and the weighted counting measure is non-ergodic.

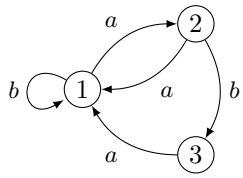
5 On algebraic properties of the density

Let us highlight a corollary which generalizes the fact that, when μ is a Bernoulli measure with rational values, the density of a rational language is rational [5].

▶ **Corollary 5.1.** *Let $\mathbb{K}$ be an extension of $\mathbb{Q}$ such that $\mu(L) \in \mathbb{K} \cup \infty$ for every rational language L. Then, for every rational language L on the alphabet A which satisfies the hypotheses of Theorem 4.8, the density of L belongs to $\mathbb{K}$.*

The hypothesis that $\mu(L) \in \mathbb{K} \cup \infty$ is satisfied when μ is a Markov measure (or, more generally, a sofic measure) defined by a transition matrix and an initial vector with coefficents in $\mathbb{K}$. More details can be found in Section 6. It is also satisfied when all the values of μ on A^* are in $\mathbb{K}$ and the support X of μ is minimal. Indeed, in this case, consider a rational language L. If every word of $\mathcal{L}(X)$ is a factor of some word of L, then $\delta_\mu(L) > 0$ by Equation (8), which implies $\mu(L) = \infty$. Otherwise, the intersection $L \cap \mathcal{L}(X)$ is finite and thus the conclusion follows. Let us finish with an example.

▶ **Example 5.2.** Let X be the Fibonacci shift (Example 1.1) and $\mathcal{A}$ be the automaton depicted in Figure 3. Let μ be the unique ergodic measure on X. Let $\varphi\colon A^* \to M$ be the transition morphism of $\mathcal{A}$ and $\alpha = \varphi(a)$, $\beta = \varphi(b)$. Let R be the $\mathcal{R}$-class of α. Note that α^2 is an idempotent which belongs to the $\mathcal{J}$-class $J_X(M)$.



α, α^2 *	$\alpha\beta, \alpha^2\beta$ *
$\beta\alpha, \beta\alpha^2$ *	$\beta, \beta\alpha\beta$

Figure 3 The automaton $\mathcal{A}$ in Example 5.2 next to the $\mathcal{J}$-class of $J_X(M)$, where M is the transition monoid of $\mathcal{A}$. The image of a in M is denoted by α while the image of b is denoted by β.

Let ν be the weighted counting measure on $(R \cup \{0\}) \rtimes X$. We claim that ν is an ergodic measure. First observe that $(R \cup \{0\}) \rtimes X$ has two closed invariant subsets

$$Y_0 = (\{\alpha\beta, \alpha^2\beta\} \times [b]_X) \cup (\{0\} \times X), \qquad Y_1 = (\{\alpha\beta, \alpha^2\beta\} \times [a]_X) \cup (\{\alpha, \alpha^2\} \times X).$$

Note that $\nu(Y_0) = 0$ and $\nu(Y_1) = 1$ (in fact Y_1 is the support of ν). Consider the morphism $\psi\colon A^* \to \mathbb{Z}/2\mathbb{Z}$ defined by $\psi(a) = 1$ and $\psi(b) = 0$ and the corresponding skew product $(\mathbb{Z}/2\mathbb{Z} \times X, \tilde{T})$. The weighted counting measure $\tilde{\nu}$ on this skew product is ergodic by [9, Corollary 8.12]. Furthermore, one can show that

$$\pi(g, x) = \begin{cases} (\alpha^2, x) & \text{if } g = 0 \text{ and } x_{-1} = a \\ (\alpha, x) & \text{if } g = 1 \text{ and } x_{-1} = a \\ ((\alpha\beta)^2, x) & \text{if } g = 0 \text{ and } x_{-1} = b \\ (\alpha\beta, x) & \text{if } g = 1 \text{ and } x_{-1} = b \end{cases}$$

is a continuous map $\mathbb{Z}/2\mathbb{Z} \times X \to Y_1$ such that $\pi \circ \tilde{T} = T \circ \pi$ and $\tilde{\nu} \circ \pi^{-1} = \nu$. Therefore the weighted counting measure on $(R \cup \{0\}) \rtimes X$ is ergodic, and the hypotheses of Theorem 4.8 are satisfied.

The values $\mu(w)$ for $w \in A^*$ are all in the quadratic extension $\mathbb{K} = \mathbb{Q}(\lambda)$ where λ is the golden ratio [8, Theorem 2]. By Corollary 5.1 it follows that $\delta_\mu(\varphi^{-1}(m)) \in \mathbb{K}$ for every $m \in J_X(M)$. For instance the language $L = \varphi^{-1}(\alpha)$ satisfies $\mu(LA^*) = \mu(A^*L) = \mu(a)$. Since $\mu(a) = \lambda$ (cf. [14, Example 3.8.19]), we get that $\delta_\mu(L) = \frac{1}{2\lambda^2}$ from Theorem 4.8 (noting that the $\mathcal{H}$-classes of $J_X(M)$ have size 2). Notice also that the language $\psi^{-1}(0)$ has the same intersection with $\mathcal{L}(X)$ as the submonoid C^* generated by $C = \{aa, aba, b\}$. The language C^* is recognized by $\mathcal{A}$ with 1 as initial and terminal state. Since $\mathbb{Z}/2\mathbb{Z} \rtimes X$ is uniquely ergodic, we have $\delta_\mu(\psi^{-1}(0)) = 1/2$.

6 Markov and sofic measures

This section discusses how our results may be applied to the case of Markov, and more generally sofic measures. The main result, Proposition 6.4, shows that sofic measures satisfy the condition on field extensions from Corollary 5.1.

First, recall that a *Markov measure* μ (also called a *Markov chain*) is given by an $A \times A$-stochastic matrix M (its transition matrix) and a stochastic A-vector v (its initial vector) which is a left eigenvector of M for the eigenvalue 1. Then, for $w = a_0a_1 \cdots a_{n-1}$, with $a_i \in A$, we define

$$\mu(w) = v_{a_0} M_{a_0a_1} \cdots M_{a_{n-2},a_{n-1}}$$

The measure is invariant because $vM = v$. It is ergodic if the matrix M is irreducible and it is mixing if M is primitive (see [31]). A shift space X has *finite type* if $\mathcal{L}(X) = A^* \setminus A^*FA^*$ for some finite set $F \subseteq A^*$, called the set of forbidden blocks. The support of a Markov

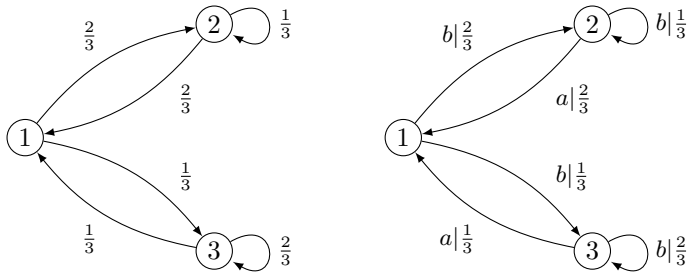


Figure 4 A sofic measure.

measure μ is the shift of finite type defined by the set of forbidden blocks ab where $a, b \in A$ are such that $M_{a,b} = 0$. In terms of probability theory, a Markov measure is defined by a sequence of random variables ζ_n, as for a Bernoulli measure, but this time ζ_n depends on ζ_{n-1}. When ζ_n depends on $\zeta_{n-k}, \ldots, \zeta_{n-1}$ we say that μ is a Markov measure of *order* k.

A *sofic measure* ν (also called a *hidden Markov chain*) is given by a shift of finite type X on the alphabet B, a Markov measure μ on X, and a map $\phi\colon B \to A$ from B onto an alphabet A. We extend ϕ to a morphism from B^* to A^*. Its extension to a map from $A^{\mathbb{Z}}$ to $B^{\mathbb{Z}}$ is called a 1-block map. Then, for $w \in A^*$, we define

$$\nu(w) = \mu(\phi^{-1}(w)).$$

The support of ν is contained in the sofic shift $Y = \phi(X)$ (a shift X is called *sofic* if $\mathcal{L}(X)$ is rational). A sofic measure is invariant. It is ergodic if μ is ergodic (see [11]).

▶ **Example 6.1.** Consider the Markov measure on $B = \{1, 2, 3\}$ defined by the pair

$$v = \begin{bmatrix} 1/3 & 1/3 & 1/3 \end{bmatrix}, \quad M = \begin{bmatrix} 0 & 2/3 & 1/3 \\ 2/3 & 1/3 & 0 \\ 1/3 & 0 & 2/3 \end{bmatrix}$$

The support of μ is the shift of finite type X represented in Figure 4 on the left. Using the 1-block map $\phi(1) = a$ and $\phi(2) = \phi(3) = b$, we obtain an invariant sofic measure defined by the diagram of Figure 4 on the right. It can be shown that this sofic measure is not a Markov measure of order k for any k (see [11]).

A map $\mu\colon A^* \to \mathbb{R}_+$ is $\mathbb{R}_+$*-rational* if there is a morphism $\varphi\colon A^* \to M_n(\mathbb{R}_+)$ from A^* into the monoid of $n \times n$-matrices with coefficients in $\mathbb{R}_+$, a row vector λ and a column vector γ such that

$$\mu(w) = \lambda\mu(w)\gamma,$$

for every $w \in A^*$. The triple $(\lambda, \varphi, \gamma)$ is called a *linear representation* of μ. The following is from [20] (see also the survey [11]).

▶ **Proposition 6.2.** *A probability measure on $A^{\mathbb{Z}}$ is sofic if and only if the associated probability distribution is $\mathbb{R}_+$-rational.*

The probability measure will be invariant if and only if the linear representation $(\lambda, \varphi, \gamma)$ can be chosen such that the matrix $P = \sum_{a \in A} \varphi(a)$ is stochastic, with $\lambda P = \lambda$ and $\gamma = \begin{bmatrix} 1 & 1 & \ldots & 1 \end{bmatrix}^t$.

▶ **Example 6.3.** A linear representation for the sofic measure of Example 6.1 is given by

$$\varphi(a) = \begin{bmatrix} 0 & 0 & 0 \\ 2/3 & 0 & 0 \\ 1/3 & 0 & 0 \end{bmatrix}, \quad \varphi(b) = \begin{bmatrix} 0 & 2/3 & 1/3 \\ 0 & 1/3 & 0 \\ 0 & 0 & 2/3 \end{bmatrix},$$

with

$$\lambda = \begin{bmatrix} 1/3 & 1/3 & 1/3 \end{bmatrix}, \quad \gamma = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}.$$

Let $L \subseteq A^*$ be a language and μ be a probability measure on $A^{\mathbb{Z}}$. The *generating series* of L is the formal series

$$f_L(z) = \sum_{n \geq 0} \mu(L \cap A^n) z^n.$$

Therefore, the density of L (if it exists) is the limit in average of the coefficients of $f_L(z)$.

A formal series $f(z) = \sum_{n \geq 0} f_n z^n$, with $f_n \geq 0$, is $\mathbb{R}_+$-rational if the map $a^n \mapsto f_n$ is $\mathbb{R}_+$-rational. In this case, we have $f(z) = p(z)/q(z)$ for two polynomials p, q with real coefficients. The following statement, originally due to Berstel, is well known (see [15] for example).

▶ **Proposition 6.4.** *Let L be a rational language. If μ is a sofic measure, then $f_L(z)$ is $\mathbb{R}_+$-rational. Let $\mathbb{K}$ be a subfield of $\mathbb{R}$ containing the values of μ on A^*. Then $\mu(L) \in \mathbb{K} \cup \{\infty\}$ and $\delta_\mu(L) \in \mathbb{K}$.*

Proof. Let $g \colon A^* \to \mathbb{R}_+$ be defined by

$$g(w) = \begin{cases} \mu(w) & \text{if } w \in L, \\ 0 & \text{otherwise.} \end{cases}$$

Thus, $g(w)$ is the product of the values of μ and of the characteristic function χ_L of L. The map $w \mapsto \mu(w)$ is $\mathbb{R}_+$-rational by Proposition 6.2 and χ_L is $\mathbb{R}_+$-rational since L is rational. Therefore, by [15, Theorem 5.2], the map g is $\mathbb{R}_+$-rational. Set $h_n = \sum_{w \in A^n} g(w)$. Then $h = \sum_{n \geq 0} h_n z^n$ is $\mathbb{R}_+$-rational. Since $h_n = \mu(L \cap A^n)$, this proves that $f_L(z)$ is $\mathbb{R}_+$-rational.

By [15, Theorem 7.2], there is an integer $p \geq 0$ such that $r_i = \lim_{n \to \infty} h_{np+i}$ exists for $0 \leq i < p$. Moreover $r_i \in \mathbb{K}$ whenever $h_n \in \mathbb{K}$. This implies that $\delta_\mu(L) = \frac{1}{n} \sum_{0 \leq i < p} r_i$ exists and is in $\mathbb{K}$. Finally, we have $f_L(z) = p(z)/q(z)$ for two polynomials $p, q \in \mathbb{R}[z]$. Since $h_n \in \mathbb{K}$, we may assume $p, q \in \mathbb{K}[z]$ by [15, Proposition 3.2]. Thus, if $\mu(L) = \sum_{n \geq 0} h_n$ is finite, the radius of convergence of $f_L(z)$ is > 1 and the sum is equal to $p(1)/q(1)$, which is in $\mathbb{K}$. ◀

▶ **Example 6.5.** Let μ be the Bernoulli measure on $\{a, b\}^{\mathbb{Z}}$ defined by $\mu(a) = p$, $\mu(b) = q$. Let $L = \{a, b\}^* ab$. Then

$$f_L(z) = \sum_{n \geq 0} pq z^{n+2} = \frac{pqz^2}{1 - z}$$

and consequently $\delta_\mu(L) = pq$.

References

1 Jean-Paul Allouche and Jeffrey Shallit. *Automatic Sequences*. Cambridge University Press, 2003. doi:10.1017/cbo9780511546563.

2 Jorge Almeida. Profinite semigroups and applications. In V. Kudryavtsev and I. G. Rosenberg, editors, *Structural Theory of Automata, Semigroups, and Universal Algebra*, volume 207 of *NATO Science Series II: Mathematics, Physics and Chemistry*, pages 1–45. Springer, 2005. doi:10.1007/1-4020-3817-8_1.

3 Jorge Almeida, Alfredo Costa, Revekka Kyriakoglou, and Dominique Perrin. On the group of a rational maximal bifix code. *Forum Math.*, 32(3):553–576, 2020. doi:10.1515/forum-2018-0270.

4 Marie-Pierre Béal, Dominique Perrin, and Antonio Restivo. Decidable problems in substitution shifts. *J. Comput. System Sci.*, 143, 2024. doi:10.1016/j.jcss.2024.103529.

5 Jean Berstel. Sur la densité asymptotique de langages formels. In M. Nivat, editor, *Automata Languages and Programming*, pages 345–358. North-Holland, 1972.

6 Jean Berstel, Clelia De Felice, Dominique Perrin, Christophe Reutenauer, and Giuseppina Rindone. Bifix codes and Sturmian words. *J. Algebra*, 369:146–202, 2012.

7 Jean Berstel, Dominique Perrin, and Christophe Reutenauer. *Codes and Automata*. Cambridge University Press, 2009.

8 Valérie Berthé. Fréquences des facteurs des suites sturmiennes. *Theor. Comput. Sci.*, 165(2):295–309, 1996. doi:10.1016/0304-3975(95)00224-3.

9 Valérie Berthé, Herman Goulet-Ouellet, Carl-Fredrik Nyberg-Brodda, Dominique Perrin, and Karl Petersen. Density of group languages in shift spaces, 2024. doi:10.48550/arXiv.2403.17892.

10 Manuel Bodirsky, Tobias Gärtner, Timo von Oertzen, and Jan Schwinghammer. Efficiently computing the density of regular languages. In *LATIN 2004: Theoretical Informatics*, pages 262–270. Springer Berlin Heidelberg, 2004. doi:10.1007/978-3-540-24698-5_30.

11 Mike Boyle and Karl Petersen. *Hidden Markov processes in the context of symbolic dynamics*, pages 5–71. Cambridge University Press, 2011. doi:10.1017/cbo9780511819407.002.

12 Olivier Carton and Wolfgang Thomas. The monadic theory of morphic infinite words and generalizations. *Inf. Comput.*, 176(1):51–65, 2002. doi:10.1006/inco.2001.3139.

13 Thomas Colcombet. Green's relations and their use in automata theory. In Adrian-Horia Dediu, Shunsuke Inenaga, and Carlos Martín-Vide, editors, *Language and Automata Theory and Applications*, volume 6638 of *Lecture Notes in Computer Science*, pages 1–21. Springer, 2011. doi:10.1007/978-3-642-21254-3_1.

14 Fabien Durand and Dominique Perrin. *Dimension Groups and Dynamical Systems*. Cambridge University Press, 2022. doi:10.1017/9781108976039.

15 Samuel Eilenberg. *Automata, Languages and Machines, vol. A*. Pure and applied mathematics. Academic Press, 1974.

16 Samuel Eilenberg. *Automata, Languages and Machines, vol. B*. Pure and applied mathematics. Academic Press, 1976.

17 N. Pytheas Fogg. *Substitutions in dynamics, arithmetics and combinatorics*, volume 1794 of *Lecture Notes in Mathematics*. Springer-Verlag, 2002.

18 Pierre A. Grillet. *Semigroups: An Introduction to the Structure Theory*. Marcel Dekker, 1995. doi:10.4324/9780203739938.

19 G. Hansel and D. Perrin. Codes and Bernoulli partitions. *Math. Syst. Theory*, 16(1):133–157, 1983. doi:10.1007/BF01744574.

20 G. Hansel and D. Perrin. Rational probability measures. *Theor. Comput. Sci.*, 65(2):171–188, 1989. doi:10.1016/0304-3975(89)90042-x.

21 John M. Howie. *Fundamentals of semigroup theory*. Oxford science publications. Oxford University Press, 1995. doi:10.1093/oso/9780198511946.001.0001.

22 Toshihiro Koga. On the density of regular languages. *Fundam. Inform.*, 168(1):45–49, 2019. doi:10.3233/fi-2019-1823.

23 Jakub Kozik. Conditional densities of regular languages. *Electron. Notes Theor. Comput. Sci.*, 140:67–79, 2005. doi:10.1016/j.entcs.2005.06.023.

24 G. Lallement. *Semigroups and Combinatorial Applications.* Wiley, 1979.

25 M. Lothaire. *Combinatorics on Words.* Cambridge University Press, 1997. doi:10.1017/cbo9780511566097.

26 James F. Lynch. Convergence laws for random words. *Australas. J. Combin.*, 7:145–156, 1993. URL: http://ajc.maths.uq.edu.au/pdf/7/ocr-ajc-v7-p145.pdf.

27 Pierre Michel. Stricte ergodicité d'ensembles minimaux de substitution. *C. R. Acad. Sci.*, 278:811–813, 1974.

28 Dominique Perrin. Codes and automata in minimal sets. In Florin Manea and Dirk Nowotka, editors, *Combinatorics on Words*, volume 9304 of *Lecture Notes in Computer Science*, pages 35–46. Springer, 2015. doi:10.1007/978-3-319-23660-5_4.

29 Dominique Perrin and Paul E. Schupp. Automata on the integers, recurrence distinguishability, and the equivalence and decidability of monadic theories. In *Proceedings of the First Annual IEEE Symposium on Logic in Computer Science (LICS 1986)*, pages 301–304. IEEE Computer Society Press, 1986.

30 Dominique Perrin and Jean Éric Pin. *Infinite Words: Automata, Semigroups, Logic and Games.* Elsevier, 2004.

31 Karl E. Petersen. *Ergodic Theory.* Cambridge University Press, 1983. doi:10.1017/cbo9780511608728.

32 Robert R. Phelps. *Lectures on Choquet's theorem.* Lecture Notes in Mathematics. Springer Berlin Heidelberg, 2001. doi:10.1007/b76887.

33 Jean-Éric Pin. *Varieties of formal languages.* Springer New York, 1986.

34 M. Queffélec. *Substitution Dynamical Systems: Spectral Analysis*, volume 1294 of *Lecture Notes in Mathematics.* Springer-Verlag, 2010. doi:10.1007/978-3-642-11212-6.

35 Ville Salo. Decidability and universality of quasiminimal subshifts. *J. Comput. Syst. Sci.*, 89:288–314, 2017. doi:10.1016/j.jcss.2017.05.017.

36 Arto Salomaa and Matti Soittola. *Automata-Theoretic Aspects of Formal Power Series.* Springer New York, 1978. doi:10.1007/978-1-4612-6264-0.

37 M. P. Schützenberger. Sur certains sous-monoïdes libres. *Bull. Soc. Math. Fr.*, 93:209–223, 1965. doi:10.24033/bsmf.1623.

38 Marcel Paul Schützenberger. On finite monoids having only trivial subgroups. *Inform. Control*, 8(2):190–194, 1965. doi:10.1016/s0019-9958(65)90108-7.

39 Ryoma Sin'ya. An automata theoretic approach to the zero-one law for regular languages: Algorithmic and logical aspects. *Electron. Proc. Theor. Comput. Sci.*, 193:172–185, 2015. doi:10.4204/eptcs.193.13.

40 William A. Veech. Strict ergodicity in zero dimensional dynamical systems and the Kronecker–Weyl theorem mod 2. *Trans. Amer. Math. Soc.*, 140:1, 1969. doi:10.2307/1995120.

41 William A. Veech. Finite group extensions of irrational rotations. *Israel J. Math.*, 21(2):240–259, 1975. doi:10.1007/BF02760801.

42 Peter Walters. *An Introduction to Ergodic Theory*, volume 79 of *Graduate Texts in Mathematics.* Springer, 1982.