Densities of rational languages

Dyadisc 8, Amiens

Valérie Berthé, Herman Goulet-Ouellet, Dominique Perrin

07.07.2025

FACULTY
OF INFORMATION
TECHNOLOGY
CTU IN PRAGUE

UNIVERSITÉ DE MONCTON
CAMPUS DE MONCTON

Faculté des sciences

Part 1

Introduction

Outline

1. Introduce the notion of **density**.
2. Introduce **Green's relations** and the structure theory of semigroups.
3. Define **minimal $\mathscr{J}$-classes** associated with ergodic measures.
4. Sketch the proof of **existence of densities** of all rational languages under all invariant measures.

Probability measures on shift spaces

Let μ be a (Borel) probability measure on $A^{\mathbb{Z}}$.

- We say that μ is **invariant** if $\mu(S^{-1}(B)) = \mu(B)$ for all Borel sets $B \subseteq A^{\mathbb{Z}}$.
- We say that μ is **ergodic** if it is invariant and $S^{-1}(B) = B \implies \mu(B) \in \{0, 1\}$ for all Borel sets B.

By Birkhoff's ergodic theorem, this is equivalent to convergence of **ergodic sums**:

$$\forall B, C, \quad \lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} \mu(B \cap S^{-n}C) = \mu(B)\mu(C).$$

The support of an invariant probability measure is a shift space. If the measure is ergodic then the shift space is irreducible.

Cylinders and probability measures

Let X be a shift space. We use the following notation for cylinders:

$$[u \cdot v]_X = \{x \in X \mid x_{[-|u|,|v|)} = uv\}.$$

Notation

Let μ be a probability measure with support X and let $w \in A^*$ and $L \subseteq A^*$. We write

$$\mu(w) = \mu([\varepsilon \cdot w]_X) \quad \text{and} \quad \mu(L) = \sum_{w \in L} \mu(w).$$

Probability measures have the properties that $\mu(\varepsilon) = 1$ and $\mu(w) = \sum_{a \in A} \mu(wa)$.

If μ is invariant then $\mu(w) = \mu([w \cdot \varepsilon]_X) = \sum_{a \in A} \mu(aw)$.

Michel's theorem

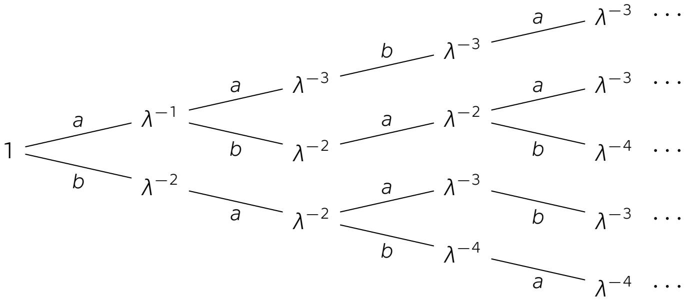
Theorem (Michel, 1974)

For every primitive substitution φ, there is a unique ergodic measure supported on the shift space generated by φ.

Here are some simple examples of where Michel's theorem can be applied.

1. $a \mapsto abc, b \mapsto abc, c \mapsto abc$ (three points example).
2. $a \mapsto ab, b \mapsto a$ (Fibonacci).
3. $a \mapsto ab, b \mapsto ba$ (Thue−Morse).
4. $a \mapsto aab, b \mapsto acb, c \mapsto ba$.

Fibonacci

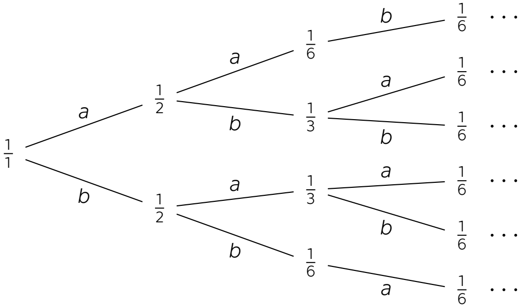


Fibonacci ergodic measure (λ = golden ratio)

The support of this measure is the **Fibonacci shift space**.

$$\sigma \colon a \mapsto ab, b \mapsto a, \qquad abaababaabaababaababa \cdots$$

Thue–Morse



Thue–Morse ergodic measure

The support of this measure is the **Thue–Morse shift space**.

$$\sigma\colon a \mapsto ab, b \mapsto ba, \qquad abbabaabbaababbabaab\cdots$$

Part 2

The notion of density

Density of a rational language

Definition (Berstel, 1972)

Let μ be a probability measure on $A^{\mathbb{Z}}$ and $L \subseteq A^*$. The **density of L with respect to μ** is the limit

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

We say that the density exists in the **strong sense** if $\lim_{n\to\infty} \mu(L \cap A^n)$ exists.

Our goal is to show that the density of a rational language exists for every invariant measure μ and to calculate it effectively under some conditions.

The existence result is known when μ is a **Bernoulli measure** (Berstel, 1972). We also considered recently the case of **group languages** (Berthé et al., 2024).

Basic properties

Let μ be a probability measure of support $X \subseteq A^{\mathbb{Z}}$. Recall the definition of density:

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

For every languages $L, K \subseteq A^*$, the following properties hold:

1. $0 \leq \delta_\mu(L) \leq 1$.
2. $\delta_\mu(L) = \delta_\mu(L \cap \mathcal{L}(X))$.
3. $\delta_\mu(L \cup K) = \delta_\mu(L) + \delta_\mu(K)$ if $L \cap K = \varnothing$.
4. $\delta_\mu(A^* \setminus L) = 1 - \delta_\mu(L)$.

Three points example

$$x = (abc)^\infty, \quad \mu(x) = \mu(Sx) = \mu(S^2x) = \tfrac{1}{3}, \quad X = \{x, Sx, S^2x\},$$
$$L = \{w \in \{a, b, c\}^* \mid |w|_a + |w|_b \equiv 0 \bmod 2\}.$$

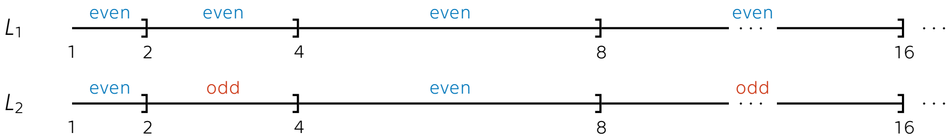
$$\begin{array}{llll}
|abc|_a + |abc|_b \equiv 0 & |bca|_a + |bca|_b \equiv 0 & |cab|_a + |cab|_b \equiv 0 & \mu(L \cap A^3) = 1 \\
|abca|_a + |abca|_b \equiv 1 & |bcab|_a + |bcab|_b \equiv 1 & |cabc|_a + |cabc|_b \equiv 0 & \mu(L \cap A^4) = \tfrac{1}{3} \\
|abcab|_a + |abcab|_b \equiv 0 & |bcabc|_a + |bcabc|_b \equiv 1 & |cabca|_a + |cabca|_b \equiv 1 & \mu(L \cap A^5) = \tfrac{1}{3}
\end{array}$$

$$\delta_\mu(L) = \lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} \mu(L \cap A^i) = \tfrac{1}{3}(1 + \tfrac{1}{3} + \tfrac{1}{3}) = \tfrac{5}{9}$$

The problem of existence

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

$$L_1 = \{w \in A^* \mid |w| \equiv 0 \mod 2\},$$
$$L_2 = \{w \in A^* \mid |w| \equiv \lfloor \log_2(|w|) \rfloor \mod 2\}.$$



In this example $\delta_\mu(L_1) = \delta_\mu(L_2) = 1/2$ but $\delta_\mu(L_1 \cap L_2)$ does not exist (no matter μ).

Part 3

Density of ideals

Density of right ideals

Definition

A language L is a **right ideal** if $LA^* = L$.

Proposition

Let μ be a probability measure on $A^{\mathbb{Z}}$ and let $L \subseteq A^*$ be a right ideal. Then $\delta_\mu(wA^*)$ exists in the strong sense and

$$\delta_\mu(L) = \mu(L \setminus LA^+).$$

$L \setminus LA^+$ is the unique prefix code D such that $L = DA^*$.

In particular for every $w \in A^*$ the set wA^* is a right ideal and $\delta_\mu(wA^*) = \mu(w)$.

Density of left ideals

Definition

A language L is a **left ideal** if $A^*L = L$.

Proposition

Let μ be a probability measure on $A^{\mathbb{Z}}$ and let $L \subseteq A^*$ be a left ideal. **If μ is invariant** then $\delta_\mu(L)$ exists in the strong sense and

$$\delta_\mu(L) = \mu(L \setminus A^+L).$$

$L \setminus A^+L$ is the unique suffix code G such that $L = GA^*$.

In particular for every $w \in A^*$ the set A^*w is a left ideal and $\delta_\mu(wA^*) = \mu(w)$, provided μ is invariant.

Quasi-ideals

Definition

A **quasi-ideal** is the intersection of a left and a right ideal.

Proposition

Let μ be a probability measure on $A^{\mathbb{Z}}$, let $L \subseteq A^*$ be a left ideal and let $K \subseteq A^*$ be a right ideal. **If μ is ergodic** then $\delta_\mu(L \cap K)$ exists and

$$\delta_\mu(L \cap K) = \mu(u)\mu(v).$$

The proof uses the **convergence of ergodic sums** in a key way.

In particular if $u, v \in A^*$ then $A^*u \cap vA^*$ is a quasi-ideal and $\delta_\mu(A^*u \cap vA^*) = \mu(u)\mu(v)$, provided μ is ergodic.

Zero-one law for two-sided ideals

Definition

A language L is a **two-sided ideal** if $A^*LA^* = L$.

Theorem

Let μ be a probability measure on A^* and let $w \in A^*$. **If μ is ergodic** then the density of every two-sided ideal exists in the strong sense and is equal to 0 or 1.

This follows from the formula for quasi-ideals. Set $D = L \setminus LA^+$ and $G = L \setminus A^+L$. Then $L = DA^* = A^*G = DA^* \cap A^*G$ and so

$$\delta_\mu(L) = \mu(G)\mu(D) = \delta_\mu(A^*G)\delta_\mu(DA^*) = \delta_\mu(L)^2.$$

Since $\delta_\mu(L) \in [0, 1]$ the result follows.

In particular, $\delta_\mu(L) = 1$ if $\mu(w) > 0$ for some $w \in L$ and 0 otherwise.

Recap

1. For right ideals, $\delta_\mu(L) = \mu(L \setminus LA^+)$ in the strong sense.
2. For left ideals, $\delta_\mu(L) = \mu(L \setminus A^+L)$ in the strong sense, provided μ is invariant.
3. For quasi-ideals, $\delta_\mu(L) = \mu(G)\mu(D)$, provided μ is ergodic.
4. For two-sided ideals, $\delta_\mu(L) \in \{0, 1\}$ in the strong sense if μ is ergodic.

Part 4

Finite monoids and Green's relations

The algebraic definition of rational languages

The following is equivalent to the classical automatic definition.

Definition

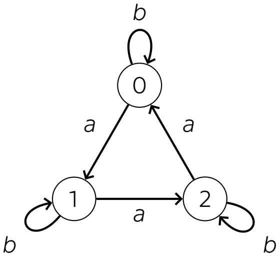
A language $L \subseteq A^*$ is **rational** if there exists a finite monoid M, a morphism $\varphi\colon A^* \to M$ and a subset $K \subseteq M$ such that $L = \varphi^{-1}(K)$.

We say that the monoid M recognizes the language L.

- The transition function of an automaton $\mathcal{A}$ with state set Q defines a morphism from A^* to a transformation monoid on Q recognizing the same languages as $\mathcal{A}$. We call this monoid the **transition monoid** of $\mathcal{A}$.
- Conversely, a morphism $\varphi\colon A^* \to M$ determines an automaton $\mathcal{A}$ with state set M recognizing the same languages as M.

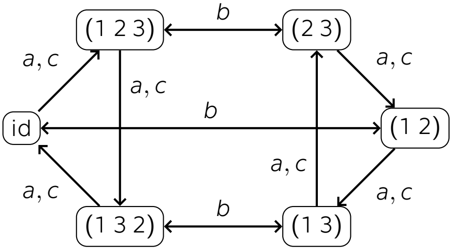
Modular letter counting

$$\varphi\colon \{a,b\}^* \to \mathbb{Z}/m\mathbb{Z}, \qquad \varphi(a) = 1, \varphi(b) = 0,$$
$$\varphi^{-1}(0) = \{w \mid |w|_a \equiv 0 \bmod m\}.$$



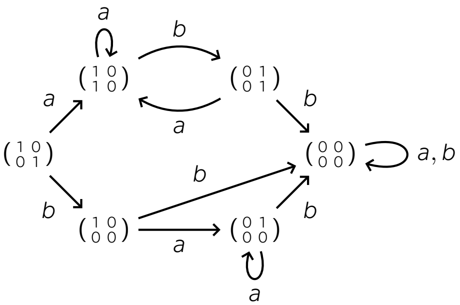
Automaton on the symmetric group

$$\varphi\colon \{a,b,c\}^* \to S_3, \quad a \mapsto (1\,2\,3), b \mapsto (1\,2), c \mapsto (1\,2\,3)$$



Automaton on 0-1 matrices

$$\varphi\colon \{a,b\}^* \to M \leq \{0,1\}^{2\times 2}, \quad a \mapsto \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix}, b \mapsto \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$$



(This is the transition monoid of the Rauzy graph of order 1 in the Fibonacci shift.)

Definition (Green, 1951)

The four Green's relations on a monoid M are defined by

$$s \mathscr{R} t \iff sM = tM$$
$$s \mathscr{L} t \iff Ms = Mt$$
$$s \mathscr{H} t \iff sM = tM, Ms = Mt$$
$$s \mathscr{J} t \iff MsM = MtM$$

So $\mathscr{R}$ is equality of principal right ideals, $\mathscr{L}$ of left ideals and $\mathscr{J}$ of two-sided ideals. The relation $\mathscr{H}$ is simply the intersection $\mathscr{L} \cap \mathscr{R}$.

We have the inclusions $\mathscr{H} \subseteq \mathscr{L}, \mathscr{R} \subseteq \mathscr{J}$.

Locating subgroups

Terminology

Let M be a monoid.

1. A **subgroup** of M is a subsemigroup which is in fact a group. It does not necessarily share the same identity element.
2. An $\mathscr{H}$-class of M is called **regular** if it contains an idempotent ($s^2 = s$).

Proposition (Green, 1951)

Let M be a monoid.

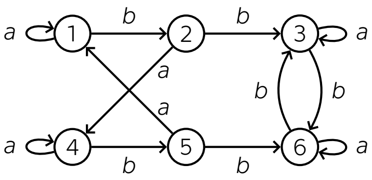
1. Every subgroup of M is contained in a regular $\mathscr{H}$-class.
2. Every regular $\mathscr{H}$-class is a subgroup of M.
3. If M is finite, all regular $\mathscr{H}$-classes in the same $\mathscr{J}$-class are isomorphic as groups.

Eggbox diagrams

In finite monoids, an $\mathscr{R}$-class R and an $\mathscr{L}$-class L which are in the same $\mathscr{J}$-class have non-empty intersection, and this intersection is an $\mathscr{H}$-class.

We represent finite monoid using an **eggbox diagram**. The boxes are $\mathscr{J}$-classes, the rows are $\mathscr{R}$-classes, the columns are $\mathscr{L}$-classes, and the cells are $\mathscr{H}$-classes.

Asterisks in the upper right corner indicate the regular $\mathscr{H}$-classes.



1^*

$ab, (ab)^2$ *	a, aba *
b, bab	$ba, (ba)^2$ *

b^2, b^3 *

Part 5

J-classes associated with shift spaces

The minimal $\mathcal{J}$-class

Let $\varphi\colon A^* \to M$ be a morphism onto a finite monoid M. Let μ be an ergodic measure and let X be its support.

Definition

We denote by $K_X(M)$ the intersection of all ideals of M which meet $\varphi(\mathcal{L}(X))$.

The minimal $\mathcal{J}$-class of M is the set

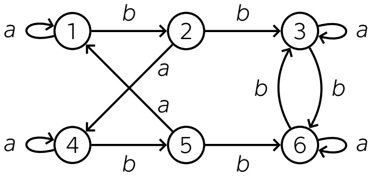
$$J_X(M) = \{m \in K_X(M) \mid MmM \cap \varphi(\mathcal{L}(X)) \neq \emptyset\}.$$

1. $J_X(M)$ is a regular $\mathcal{J}$-class.
2. $J_X(M)$ contains all $\leq_{\mathcal{J}}$-minimal elements of $\varphi(\mathcal{L}(X))$.

If X is substitutive or sofic, then $J_X(M)$ is computable for every finite monoid M.

Example of minimal J-class

Let M be the transition monoid of the automaton below on the left. The monoid has 11 elements, shown in the eggbox diagram on the right.



1^*

$ab, (ab)^2$ *	a, aba *
b, bab	$ba, (ba)^2$ *

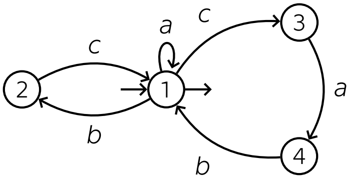
b^2, b^3 *

The minimal $\mathcal{J}$-class $J_X(M)$, where X is the Fibonacci shift space, is shown in yellow.

For the Thue–Morse shift space, the minimal $\mathcal{J}$-class is the set $\{b^2, b^3\}$. It is in fact the minimal ideal of M.

Three points example

Let L be the language recognized by the automaton below on the left. Let M be the transition monoid of this automaton.



abc^*	a	ab
bc	bca^*	b
c	ca	cab^*

Let X be the shift space $\{(abc)^\infty, (bca)^\infty, (cab)^\infty\}$ with the uniform distribution μ. The minimal $\mathscr{J}$-class $J_X(M)$ is shown on the right. The areas in yellow indicate the elements of $J_X(M)$ which are in the image of L.

Concentration of density

Theorem

Let μ be a probability measure on $A^{\mathbb{Z}}$ and let $\varphi\colon A^* \to M$ be a morphism onto a finite monoid M. If μ is ergodic then $\delta_\mu(\varphi^{-1}(J_X(M))) = 1$ in the strong sense.

In particular, $\delta_\mu(\varphi^{-1}(m)) = 0$ in the strong sense for all $m \notin J_X(M)$.

Indeed, $J_X(M) \cap \varphi(\mathcal{L}(X)) = K_X(M) \cap \varphi(\mathcal{L}(X))$, and by definition $K_X(M)$ is a two-sided ideal which meets the image of $\mathcal{L}(X)$. Then, we can apply the zero-one law for two-sided ideals to conclude.

It follows from the theorem that $\delta_\mu(L) = \delta_\mu(L')$ where $L' = \{w \in L \mid \varphi(w) \in J_X(X)\}$.

Density of aperiodic languages

A language L is **aperiodic** if it recognized by a monoid with only trivial subgroups. By a famous theorem of Schützenberger, those are exactly the star-free languages.

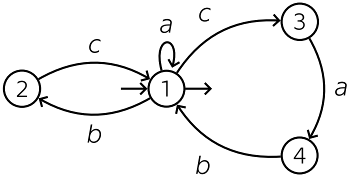
Theorem

Let μ be an ergodic measure on $A^{\mathbb{Z}}$. For every aperiodic language $L \subseteq A^*$, the density $\delta_\mu(L)$ exists.

Let $\varphi \colon A^* \to M$ be a morphism onto a finite aperiodic monoid. We may assume that $L = \varphi^{-1}(m)$ for some $m \in M$.

- If $m \notin J_X(M)$, then $\delta_\mu(L) = 0$ by concentration of the density on $J_X(M)$.
- If $m \in J_X(M)$, then we have $L \cap \mathcal{L}(X) = LA^* \cap A^*L \cap \mathcal{L}(X)$ and therefore $\delta_\mu(L) = \delta_\mu(LA^*)\delta_\mu(A^*L)$ by the formula for quasi-ideals.

Three points example



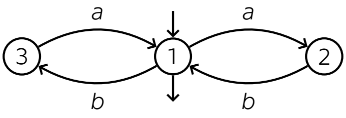
abc^*	a	ab
bc	bca^*	b
c	ca	cab^*

In this example, $\delta_\mu(\varphi^{-1}(m)) = \frac{1}{9}$ for each element $m \in J_X(M)$, therefore $\delta_\mu(L) = \frac{5}{9}$.

Note that L has the same intersection with $\mathcal{L}(X)$ as the non-aperiodic language

$$\{w \in A^* \mid |w|_a + |w|_b \equiv 0 \mod 2\}.$$

Consider the transition monoid M of the automaton below on the left, and let X be the Thue–Morse shift. The minimal $\mathcal{J}$-class $J_X(M)$ is represented on the right.



a^2	a^2b	a^2b^2 *
ba^2	ba^2b *	ab^2
b^2a^2 *	b^2a	b^2

Using the previous theorem, we can calculate that the density of the language $L = \{ab, ba\}^*$ under the Thue–Morse ergodic measure is $\delta_\mu(L) = \frac{1}{4}$.

Indeed, one has $LA^* = (ab)^* \cup (ba)^* \cup ((ab)^+b \cup (ba)^+a)A^*$, from which we can deduce that $\delta_\mu(LA^*) = \mu(abb, ababb, baa, babaa) = \frac{1}{2}$. Then $\delta_\mu(A^*L) = \frac{1}{2}$ for similar reasons, and $\delta_\mu(L) = \delta_\mu(A^*L)\delta_\mu(LA^*) = \frac{1}{4}$.

Part 6

Existence theorem

Main result

Existence Theorem

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

We fix a morphism $\varphi\colon A^* \to M$ onto a finite monoid M. The proof has three steps.

STEP 1 Use the **ergodic decomposition theorem** to reduce to the ergodic case.

STEP 2 Define a **skew product** $(R \cup \{0\}) \times X$, where R is an $\mathscr{R}$-class of $J_X(M)$.

STEP 3 Find an **ergodic lift** of μ on $(R \cup \{0\}) \times X$.

STEP 4 Express the density as **ergodic sums** in the skew product.

We will present details for step 2 and 4.

Step 2: The skew product

Let $\varphi\colon A^* \to M$ be a morphism onto a finite monoid M. Let μ be an ergodic measure with support X. Let R be an $\mathcal{R}$-class of $J_X(M)$ such that $R \cap \varphi(\mathcal{L}(X)) \neq \varnothing$.

Definition

In our setting, the **skew product** with the $\mathcal{R}$-class R is the dynamical system formed by $(R \cup \{0\}) \times X$ with the continuous transformation T defined by

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

where $r \cdot m = rm$ if $rm \in R$ and 0 otherwise.

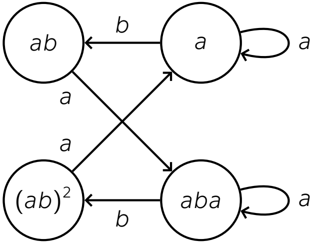
If $M = G$ is a group, then $R = G$ and we can get rid of 0 in the skew product.

Example

The skew product is a model for walks in the natural automaton over the chosen $\mathscr{R}$-class. The orbit under T of (r, x) correspond to an infinite walk starting at r.

$ab, (ab)^2$ *	a, aba *
b, bab	$ba, (ba)^2$ *

$\mathscr{R}$-class of $J_X(M)$ (in yellow).



Automaton on the $\mathscr{R}$-class.

ab	aba	$(ab)^2$	a	a	ab	aba	$(ab)^2$	a	a	ab	
a	b	a	a	b	a	b	a	a	b	a	$\ldots$

Step 4: Density as ergodic sums

- By step 1, we may assume that μ is ergodic.
- By step 3, we may consider an ergodic lift $\bar{\mu}$ of μ.

The Existence Theorem follows from the proposition below.

Proposition

For every $m \in M$, we have $\delta_\mu(\varphi^{-1}(m)) = 0$ if $m \notin J_X(M)$, and otherwise:

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

The key to this proposition is the fact that for $m \in J_X(M)$, the density of $\varphi^{-1}(m)$ can be rewritten as an **ergodic sum in the skew product** with respect to the lifted measure $\bar{\mu}$.

The weighted counting measure

A candidate for an ergodic lift of μ to the skew product is the following measure.

Definition

The **weighted counting measure** on $(R \cup \{0\}) \times X$ is the probability measure ν defined by defined by $\nu(\{0\} \times X) = 0$ and for $r \in R$ and $u, v \in \mathcal{L}(X)$:

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

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

The weighted counting measure is an invariant lift of μ. When M is a group G, the weighted counting measure is the product of the counting measure on G with μ.

Equidistributed densities

Definition

Let $\varphi \colon A^* \to M$ be a morphism onto a finite monoid M and let μ be an ergodic measure with support X. We say that φ is **equidistributed** if for all $m \in J_X(M)$,

$$\delta_\mu(\varphi^{-1}(m)) = \frac{1}{d}\delta_\mu(A^*\varphi^{-1}(m))\delta_\mu(\varphi^{-1}(m)A^*)$$

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

This means that the densities are uniform within the $\mathscr{H}$-classes of $J_X(M)$.

Theorem

If the weighted counting measure is ergodic, then φ has equidistibuted densities.

This generalizes a result for the group case from Berthé et al., 2024.

The dendric case

Theorem

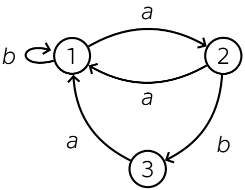
Let X be a substitutive dendric shift. Let μ be the unique ergodic measure on X. For every morphism $\varphi\colon A^* \to G$ onto a finite group G, the weighted counting measure on $G \times X$ is ergodic.

Therefore, if X is a substitutive dendric shift and $\varphi\colon A^* \to M$ is a group morphism, then φ has equidistributed densities.

Question: can we extend this beyond the substitutive case, or beyond the group case?

One last example

Let M be the transition monoid of the automaton below on the left. On the right we find the minimal $\mathscr{J}$-class $J_X(M)$, where X is the Fibonacci shift.



λ^{-2} * a, a^2	λ^{-3} * ab, a^2b
λ^{-3} * ba, ba^2	λ^{-4} b, bab

In this example, the weighted counting measure is ergodic.

The density of each $\mathscr{H}$-class is given in the upper left corner ($\lambda = \frac{1+\sqrt{5}}{2}$). These densities are then evenly distributed among the elements of the $\mathscr{H}$-class.

Some directions for future research:

- properties of weighted counting measures;
- equidistribution properties;
- cocycles and cobounding maps;
- relationship with return words;
- induced measures on profinite monoids.

A quick demo if I have time:

https://d0l.kam.fit.cvut.cz/