Density of rational languages and equidistribution

V. Berthé, H. Goulet-Ouellet, C.-F. Nyberg-Brodda, D. Perrin and K. Petersen

Funded by the European Union | European Research Council

Rencontres Mathématiques de Rouen

Density of a language

The density of a language $L \subseteq \mathcal{A}^*$ with respect to a probability measure μ on $\mathcal{A}^{\mathbb{Z}}$ is

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

whenever the limit exists

Our aim is

- to show that the density of a rational/regular language exists for every invariant measure μ
- to give a way to compute it (via ergodicity)

Symbolic dynamics

- Let $\mathcal{A}$ be a finite alphabet endowed with the discrete topology
- Let $\mathcal{A}^*$ be the set of all finite words on $\mathcal{A}$
- Let $\mathcal{A}^{\mathbb{Z}}$ be the space of bi-infinite words endowed with the product topology
- Let $S : \mathcal{A}^{\mathbb{Z}} \to \mathcal{A}^{\mathbb{Z}}$, $(Sx)_n = x_{n+1}$ be the shift map

Symbolic dynamics

- Let $\mathcal{A}$ be a finite alphabet endowed with the discrete topology
- Let $\mathcal{A}^*$ be the set of all finite words on $\mathcal{A}$
- Let $\mathcal{A}^{\mathbb{Z}}$ be the space of bi-infinite words endowed with the product topology
- Let $S : \mathcal{A}^{\mathbb{Z}} \to \mathcal{A}^{\mathbb{Z}}$, $(Sx)_n = x_{n+1}$ be the shift map

- We consider shifts (X, S) over $\mathcal{A}$, i.e., closed shift-invariant subsets of $\mathcal{A}^{\mathbb{Z}}$
- A shift is minimal if it contains no nontrivial subshift
- Minimality implies that all elements have the same language, i.e., the same set of subwords/factors, that we denote as $\mathcal{L}(X)$
- The topology is generated by cylinders

$$[w] = \{x \in X \mid x_0 = w_1, \ldots, x_{n-1} = w_n\} \text{ for } w = w_1 \cdots w_n \in \mathcal{A}^*$$

Density of regular patterns

Let (X, S) be a shift, $X \subset \mathcal{A}^{\mathbb{Z}}$

Can we define a notion of frequency/density for a regular pattern in (X, S)?

Example Words with an even number of a given letter

How often do they occur in a given shift?

Density

A shift $X \rightsquigarrow$ A measure μ

- Let μ be a (Borel) probability measure on $X \subseteq A^{\mathbb{Z}}$
- We say that μ is invariant if $\mu(S^{-1}(B)) = \mu(B)$ for all Borel sets $B \subseteq X$

$$[w] = \{x \in X \mid x_0 \cdots x_{|w|-1} = w\}$$

$$\sum_{a \in \mathcal{A}} \mu[aw] = \sum_{a \in \mathcal{A}} \mu[wa] = \mu[w] \text{ for all } w \in \mathcal{A}^*$$

- The support of an invariant probability measure is a shift space

Density

Let (X, S, μ) be a shift with a shift invariant probability measure μ

Let L be a rational language on $\mathcal{A}$ (i.e., a language recognised by a finite automaton)

The density of L under the measure μ is defined as the following limit whenever it exists

$$\begin{aligned}\delta_\mu(L) &= \lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} \mu(L \cap \mathcal{A}^i) \\ &= \lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} \mu\{x \in \mathcal{A}^{\mathbb{Z}} \mid x_0 \cdots x_{i-1} \in L\}\end{aligned}$$

- Since $\mu(w) = 0$ when $w \notin \mathcal{L}(X)$, $\delta_\mu(L) = \delta_\mu(L \cap \mathcal{L}(X))$

Some history

On the densities of rational languages under probability measures

1969 Work by Veech on a variation of the Kronecker–Weyl theorem

1972 Introduced by Berstel for Bernoulli measures

1989 Hansel and Perrin study densities for hidden Markov measures

1993 Studied by Lynch in connection with logic

2015 Work of Sin'ya on rational languages satisfying a zero-one law

The Fibonacci shift

Definition

A substitution σ is a morphism of the free monoid

$$\sigma(uv) = \sigma(u)\sigma(v)$$

The Fibonacci substitution

$$\sigma : a \mapsto ab \;\; b \mapsto a$$

$$\begin{array}{l} a \\ ab \\ aba \\ abaab \\ abaababa \end{array}$$

$$\sigma^{\infty}(a) = abaababaabaabab\cdots$$

The Fibonacci shift

Definition

A substitution σ is a morphism of the free monoid

$$\sigma(uv) = \sigma(u)\sigma(v)$$

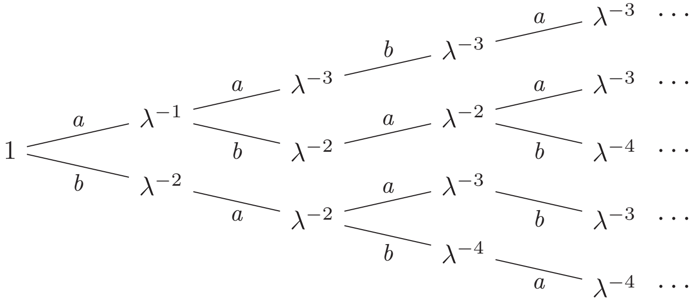
The Fibonacci substitution

$$\sigma : a \mapsto ab \;\; b \mapsto a$$

$$\sigma^\infty(a) = abaababaabaabab\cdots$$

- Let (X_σ, S) be the set of two-sided words having the same language as $\sigma^\infty(a)$
  The shift (X_σ, S, μ) is minimal and uniquely ergodic (unique invariant measure)
- The Fibonacci word $\sigma^\infty(a)$ is Sturmian and codes the rotation $R_\alpha : x \mapsto x + \alpha$ modulo 1, where $\alpha + 1$ is the golden ratio

Fibonacci measure



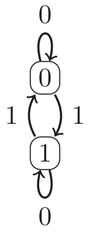
The support of this measure is the Fibonacci shift space

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

Counting modulo 2

Let (X_σ, S) be the Fibonacci shift

Question What is the density of finite words having an even number of a's in the shift (X_σ, S)? Does it exist?



Algebraically Let $\varphi\colon \{a, b\}^* \to \mathbb{Z}/2\mathbb{Z}$, $\varphi(a) = 1$, $\varphi(b) = 0$

$$L = \varphi^{-1}(0) = \{w \mid |w|_a \equiv 0 \mod 2\}$$

Group languages

We say that $L \subseteq A^*$ is a group language if

$$L = \varphi^{-1}(K)$$

where $\varphi\colon A^* \to G$ is a morphism onto a finite group and $K \subseteq G$

Group languages

We say that $L \subseteq A^*$ is a group language if

$$L = \varphi^{-1}(K)$$

where $\varphi\colon A^* \to G$ is a morphism onto a finite group and $K \subseteq G$

- Equivalently L is recognized by an automaton where letters act as permutations of the set of states
- In some sense they are opposite of star-free/aperiodic languages (Schützenberger theorem)

Skew product

- Let (X, S) be a minimal shift
- Let G be a finite group
- Let $\varphi\colon \mathcal{A}^* \to G$ a morphism which is onto G

Skew product

- Let (X, S) be a minimal shift
- Let G be a finite group
- Let $\varphi\colon \mathcal{A}^* \to G$ a morphism which is onto G

Skew product $G \times_\varphi X = (G \times X, T_\varphi)$

$$T_\varphi\colon G \times X \to G \times X, \quad T_\varphi(g, x) = (g\varphi(x_0), Sx)$$

Example Let X be the Fibonacci shift, $G = \mathbb{Z}/2\mathbb{Z}$ and

$$\varphi\colon \{a, b\}^* \to \mathbb{Z}/2\mathbb{Z}, a \mapsto 1, \; b \mapsto 0$$

$\rightsquigarrow$ counting modulo 2 in the Fibonacci shift

Skew products and cocycles

One has

$$T_\varphi \colon G \times X \to G \times X, \quad T_\varphi(g, x) = (g\varphi(x_0), Sx)$$

This gives for $n \geq 0$

$$T_\varphi^n(g, x) = (g\varphi(x_0 \cdots x_{n-1}), S^n x)$$

An ergodic viewpoint

We are given a minimal shift (X, S, μ), a group G and a surjective morphism $\varphi : X \to G$

$$T_\varphi \colon G \times X \to G \times X, \quad T_\varphi(g, x) = (g\varphi(x_0), Sx)$$

An example of an invariant measure

- Let μ be a shift- invariant probability measure on X
- Let ν be the uniform probability distribution on G
- The measure $\nu \times \mu$ is an invariant probability measure on $(G \times X, T_\varphi)$

An ergodic viewpoint

We are given a minimal shift (X, S, μ), a group G and a surjective morphism $\varphi : X \to G$

$$T_\varphi \colon G \times X \to G \times X, \quad T_\varphi(g, x) = (g\varphi(x_0), Sx)$$

- Let ν be an invariant measure on the skew product $(G \times X, T_\varphi)$ that projects to μ (the measure on the shift X)
- We fix $g \in G$

An ergodic viewpoint

We are given a minimal shift (X, S, μ), a group G and a surjective morphism $\varphi : X \to G$

$$T_\varphi \colon G \times X \to G \times X, \quad T_\varphi(g, x) = (g\varphi(x_0), Sx)$$

- Let ν be an invariant measure on the skew product $(G \times X, T_\varphi)$ that projects to μ (the measure on the shift X)
- We fix $g \in G$

The pointwise ergodic theorem applied to the skew product T_φ gives

$$\frac{1}{n} \sum_{i=0}^{n-1} 1_g(T_\varphi^i(h, x)) = \frac{1}{n} \sum_{i=0}^{n-1} 1_g(h\varphi(x_0 \cdots x_{i-1})) \to \text{ some } \bar{f}_g(h, x) \ \nu \text{ a.e.}$$

An ergodic viewpoint

We are given a minimal shift (X, S, μ), a group G and a surjective morphism $\varphi : X \to G$

$$T_\varphi \colon G \times X \to G \times X, \quad T_\varphi(g, x) = (g\varphi(x_0), Sx)$$

- Let ν be an invariant measure on the skew product $(G \times X, T_\varphi)$ that projects to μ (the measure on the shift X)
- We fix $g \in G$

The pointwise ergodic theorem applied to the skew product T_φ gives

$$\frac{1}{n}\sum_{i=0}^{n-1} 1_g(T_\varphi^i(h, x)) = \frac{1}{n}\sum_{i=0}^{n-1} 1_g(h\varphi(x_0 \cdots x_{i-1})) \to \text{ some } \bar{f}_g(h, x) \ \nu \text{ a.e.}$$

Take $h = 1$. One checks that

$$\frac{1}{n}\sum_{i=0}^{n-1} 1_g(\varphi(x_0 \cdots x_{i-1})) \to \text{ some } \bar{\psi}_g(x) \ \mu \text{ a.e.}$$

An ergodic viewpoint

$$\frac{1}{n}\sum_{i=0}^{n-1} 1_g(\varphi(x_0 \cdots x_{i-1})) \to \text{ some } \bar{\psi}_g(x)\ \mu \text{ a.e.}$$

$$\int_X \bar{\psi}_g(x)\, d\mu(x) = \int_X \lim \frac{1}{n}\sum_{i=0}^{n-1} 1_g(\varphi(x_0 \cdots x_{i-1}))\, d\mu(x)$$

An ergodic viewpoint

$$\frac{1}{n}\sum_{i=0}^{n-1} 1_g(\varphi(x_0 \cdots x_{i-1})) \to \text{ some } \bar{\psi}_g(x) \ \mu \text{ a.e.}$$

$$\begin{aligned}\int_X \bar{\psi}_g(x)\, d\mu(x) &= \int_X \lim \frac{1}{n}\sum_{i=0}^{n-1} 1_g(\varphi(x_0 \cdots x_{i-1}))\, d\mu(x) \\ &= \lim \frac{1}{n}\sum_{i=0}^{n-1} \int_X 1_g(\varphi(x_0 \cdots x_{i-1}))\, d\mu(x)\end{aligned}$$

An ergodic viewpoint

$$\frac{1}{n}\sum_{i=0}^{n-1} 1_g(\varphi(x_0 \cdots x_{i-1})) \to \text{ some } \bar{\psi}_g(x)\ \mu \text{ a.e.}$$

$$\begin{aligned}
\int_X \bar{\psi}_g(x)\, d\mu(x) &= \int_X \lim \frac{1}{n}\sum_{i=0}^{n-1} 1_g(\varphi(x_0 \cdots x_{i-1}))\, d\mu(x) \\
&= \lim \frac{1}{n}\sum_{i=0}^{n-1} \int_X 1_g(\varphi(x_0 \cdots x_{i-1}))\, d\mu(x) \\
&= \lim \frac{1}{n}\sum_{i=0}^{n-1} \sum_{w\in\mathcal{A}^i} \int_{[w]} 1_g(\varphi(x_0 \cdots x_{i-1}))\, d\mu(x)
\end{aligned}$$

An ergodic viewpoint

$$\frac{1}{n}\sum_{i=0}^{n-1} 1_g(\varphi(x_0\cdots x_{i-1})) \to \text{ some } \bar{\psi}_g(x)\ \mu \text{ a.e.}$$

Consider the rational language $L_g = \varphi^{-1}(g)$

$$\begin{aligned}
\int_X \bar{\psi}_g(x)\,d\mu(x) &= \int_X \lim \frac{1}{n}\sum_{i=0}^{n-1} 1_g(\varphi(x_0\cdots x_{i-1}))\,d\mu(x) \\
&= \lim \frac{1}{n}\sum_{i=0}^{n-1} \int_X 1_g(\varphi(x_0\cdots x_{i-1}))\,d\mu(x) \\
&= \lim \frac{1}{n}\sum_{i=0}^{n-1} \sum_{w\in\mathcal{A}^i} \int_{[w]} 1_g(\varphi(x_0\cdots x_{i-1}))\,d\mu(x) \\
&= \lim \frac{1}{n}\sum_{i=0}^{n-1} \sum_{w\in L_g\cap\mathcal{A}^i} \mu[w] = \lim \frac{1}{n}\sum_{i=0}^{n-1} \mu(L_g\cap\mathcal{A}^i)
\end{aligned}$$

Skewing the Fibonacci shift

Consider the Fibonacci shift (X_σ, S)

$$\sigma\colon a \mapsto ab,\ \ b \mapsto a$$

$$\sigma^\infty(a) = abaababaabaababaababaabaababaabaababaababaa\cdots$$

- The shift (X_σ, S, μ) is a minimal and uniquely ergodic shift
- Let $\varphi\colon \{a, b\}^* \to G$ be a morphism onto the finite group G

We consider the skew product

$$T_\varphi\colon G \times X \to G \times X, \quad T_\varphi(g, x) = (g\varphi(x_0), Sx)$$

Theorem The Fibonacci shift has minimal and uniquely ergodic skew products $G \times X_\sigma$ with all finite groups

Counting modulo 2 in the Fibonacci shift

Question What is the density of finite words having an even number of a's in the Fibonacci shift?

Counting modulo 2 in the Fibonacci shift

Question What is the density of finite words having an even number of a's in the Fibonacci shift?

We consider the language $L = \varphi^{-1}(0)$

$$\varphi\colon \{a, b\}^* \to \mathbb{Z}/2\mathbb{Z},\ \varphi(a) = 1,\ \varphi(b) = 0$$

$$L = \{w \in \{a, b\}^* \mid |w|_a \equiv 0 \mod 2\}$$

The sequence $(\mu(L \cap \mathcal{A}^n))_{n\in\mathbb{N}}$ does not have a limit, as

$$\lim_{n\to\infty} \mu(L \cap \mathcal{A}^{F_{4n}}) = 1, \quad \lim_{n\to\infty} \mu(L \cap \mathcal{A}^{F_{4n+2}}) = 0$$

but its Cèsaro mean does

$$\frac{1}{n}\sum_{i=0}^{n-1} \lim_{n\to\infty} \mu(L \cap \mathcal{A}^i) = 1/2$$

Counting modulo 2 in the Fibonacci shift

Question What is the density of finite words having an even number of a's in the Fibonacci shift?

We consider the language $L = \varphi^{-1}(0)$

$$\varphi\colon \{a,b\}^* \to \mathbb{Z}/2\mathbb{Z},\ \varphi(a) = 1,\ \varphi(b) = 0$$

$$L = \{w \in \{a,b\}^* \mid |w|_a \equiv 0 \mod 2\}$$

The sequence $(\mu(L \cap \mathcal{A}^n))_{n\in\mathbb{N}}$ does not have a limit, as

$$\lim_{n\to\infty} \mu(L \cap \mathcal{A}^{F_{4n}}) = 1, \quad \lim_{n\to\infty} \mu(L \cap \mathcal{A}^{F_{4n+2}}) = 0$$

but its Cèsaro mean does

$$\frac{1}{n}\sum_{i=0}^{n-1} \lim_{n\to\infty} \mu(L \cap \mathcal{A}^i) = 1/2$$

On the existence of the density

Theorem [Group language] [B., Goulet-Ouellet, Nyberg-Brodda, Petersen, Perrin, 2024] Let X be a shift space on a finite alphabet $\mathcal{A}$ with an ergodic measure μ and let $\varphi\colon \mathcal{A}^* \to G$ be a morphism onto a finite group G with uniform probability measure ν.
If the product measure $\nu \times \mu$ is ergodic on $G \times X$, then for every group language $L = \varphi^{-1}(K)$ with $K \subseteq G$

$$\delta_\mu(L) = |K|/|G|$$

Theorem [Rational language] [B.-Goulet-Ouellet-Perrin, 2025] Let μ be an invariant measure on $\mathcal{A}^{\mathbb{Z}}$. Then every rational language on the alphabet $\mathcal{A}$ has a density with respect to μ.

Toolbox

- Group language $L = \varphi^{-1}(g),\ g \in G$
  - Minimality can be characterized in terms of return words
  - Ergodicity via Anzai's criterium (coboundaries) and essential values
- Monoid case $L = \varphi^{-1}(m),\ m \in M$
  - Green's relations and J-class

Back to the Fibonacci shift

Consider the Fibonacci shift (X_σ, S)

$$\sigma\colon a \mapsto ab,\ \ b \mapsto a$$

$$\sigma^\infty(a) = abaababaabaababaababaabaababaabaababaababaa\cdots$$

- The shift (X_σ, S, μ) is a minimal and uniquely ergodic shift
- Let $\varphi\colon \mathcal{A}^* \to G$ be a morphism onto the finite group G

We consider the skew product

$$T_\varphi\colon G \times X \to G \times X, \quad T_\varphi(g, x) = (g\varphi(x_0), Sx)$$

Theorem The Fibonacci shift has minimal and uniquely ergodic skew products $G \times X_\sigma$ with all finite groups

Back to the Fibonacci shift

- Let (X, S) be the Fibonacci shift over $\{a, b\}$
- Let $G = \mathbb{Z}/m\mathbb{Z}$
- Let $\varphi : \{a, b\}^* \to \mathbb{Z}/m\mathbb{Z},\ a \mapsto 1,\ b \mapsto 0$

Theorem The skew product $\mathbb{Z}/m\mathbb{Z} \times X$ is uniquely ergodic

$\rightsquigarrow$ equidistribution results on the congruence of the number of visits of $R_\alpha : x \mapsto x + \alpha \mod 1$ to the interval $[0, \alpha)$, $\alpha = \frac{\sqrt{5}-1}{2}$

Corollary For every $x \in X$, $k \in \mathbb{Z}/m\mathbb{Z}$ and $a \in \{0, 1\}$, one has

$$\frac{1}{N} \mathrm{Card}\{0 \leq n \leq N-1 \mid |x_0 \cdots x_{n-1}|_a \equiv k \mod m\} \to \frac{1}{m}$$

$$\frac{1}{N} \mathrm{Card}\{0 \leq n \leq N-1 \mid \mathrm{Card}\{i \mid 0 \leq i < n, i\alpha \in [0, \alpha)\} \equiv k \mod m\} \to \frac{1}{m}$$

Continued fractions

- Let (X, S) be the Fibonacci shift X over the alphabet $\{1, 2\}$
- Let $G = \mathrm{GL}(2, \mathbb{Z}/2\mathbb{Z})$
- Let
$$\varphi\colon \{1,2\}^* \to \mathrm{GL}(2,\mathbb{Z}/2\mathbb{Z}), \quad k \mapsto \begin{pmatrix} 0 & 1 \\ 1 & \overline{k} \end{pmatrix},$$
where $\overline{k}$ stands for the congruence class of the integer k modulo 2
- The map $\varphi : \{1,2\}^* \to \mathrm{GL}(2,\mathbb{Z}/2\mathbb{Z})$ is onto

Continued fractions cocycles

For any $x \in X$, consider the real number in $[0,1]$ that admits $(x_n)_{n\geq 1}$ as its sequence of partial quotients

$$x = (x_n)_n \in X \rightsquigarrow \text{ sequence of partial quotients} \rightsquigarrow [0; x_1, x_2, \dots]$$

$$\varphi\colon \mathcal{A}^* \to \mathrm{GL}(2, \mathbb{Z}/2\mathbb{Z}), \quad k \mapsto \begin{pmatrix} 0 & 1 \\ 1 & \overline{k} \end{pmatrix}$$

Let $(p_n(x)/q_n(x))_n$ stand for the associated sequence of rational approximations in the corresponding continued fraction expansion. One has

$$q_{-1}(x) = 0, p_{-1}(x) = 1,\ q_0(x) = 1,\ p_0(x) = 0$$

$$q_{n+1}(x) = x_{n+1} q_n(x) + q_{n-1}(x),\ p_{n+1}(x) = x_{n+1} p_n(x) + p_{n-1}(x) \text{ for all } n$$

For $n \geq 0$, one has

$$\varphi^{(n)}(x) = \begin{pmatrix} p_{n-1}(x) & p_n(x) \\ q_{n-1}(x) & q_n(x) \end{pmatrix} \in \mathrm{GL}(2, \mathbb{Z}/2\mathbb{Z})$$

- Let (X, S) be the Fibonacci shift over the alphabet $\{1, 2\}$
- The skew product $\mathrm{GL}(2, \mathbb{Z}/2\mathbb{Z}) \times X$ is uniquely ergodic
- By ergodicity

$$\begin{pmatrix} p_{n-1}(x) & p_n(x) \\ q_{n-1}(x) & q_n(x) \end{pmatrix}$$

  equidistributes in the group $\mathrm{GL}(2, \mathbb{Z}/2\mathbb{Z})$.
- For every $x \in X$

$$\lim_{N\to\infty} \frac{1}{N} \mathrm{Card}\{1 \leq n \leq N \mid q_n(x) \equiv 0 \mod m\} = \frac{1}{3},$$
$$\lim_{N\to\infty} \frac{1}{N} \mathrm{Card}\{1 \leq n \leq N \mid q_n(x) \equiv 1 \mod m\} = \frac{2}{3}.$$

- One recovers the behaviour of a random irrational number [Jager-Liardet] Certain residue classes are attained more frequently than others.

- Let (X, S) be the Fibonacci shift over the alphabet $\{1, 2\}$
- Real numbers having as a sequence of partial quotients elements of the Fibonacci shift behave like a.e. real number
- One has for a.e. real number in $[0, 1]$

$$\lim_n \frac{\log q_n}{n} = \frac{\pi^2}{12 \log 2} \text{ a.e. (Lévy's theorem)}$$

- We take now the cocycle with values in $\mathrm{GL}(2, \mathbb{Z})$
- For every $x \in (X, S)$

$$\lim_n \frac{\log q_n}{n} \text{exists}$$

[Walters, Fan-Wu]