Group normality

A joint work in progress ...

Olivier Carton[1*] Vincent Delecroix[2]

[1]IRIF – Université de Paris & CNRS

[2]LaBRI – Université de Bordeaux & CNRS

[*]Supported by LIA SINFIN

2021 – STIC AmSud RAPA2

Outline

Deterministic automata

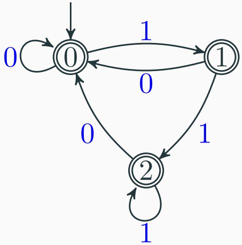
Group automata

Outline

Deterministic automata

Group automata

Deterministic automata

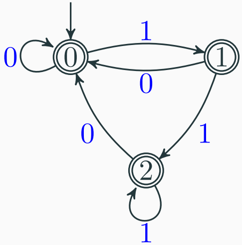


Deterministic automata are assumed to be complete. Each sequence $x \in A^{\mathbb{N}}$ is the label of exactly one run $\gamma(x)$ from the initial state.

Input: $x = 0100011 \cdots$

Run: $\gamma(x) = 0 \xrightarrow{0} 0 \xrightarrow{1} 1 \xrightarrow{0} 0 \xrightarrow{0} 0 \xrightarrow{0} 0 \xrightarrow{1} 1 \xrightarrow{1} 2 \cdots$

Deterministic automata



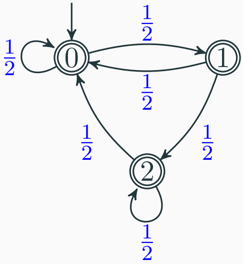
Deterministic automata are assumed to be complete. Each sequence $x \in A^{\mathbb{N}}$ is the label of exactly one run $\gamma(x)$ from the initial state.

$$\text{Input: } x = 0100011\cdots$$

$$\text{Run: } \gamma(x) = 0 \xrightarrow{0} 0 \xrightarrow{1} 1 \xrightarrow{0} 0 \xrightarrow{0} 0 \xrightarrow{0} 0 \xrightarrow{1} 1 \xrightarrow{1} 2 \cdots$$

Ergodic theorem for Markov chains

Uniform measure: $\mu(wA^{\mathbb{N}}) = \dfrac{1}{(\#A)^{|w|}}$ for each $w \in A^*$.



Stochastic matrix

$$P = \frac{1}{2}\begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \end{pmatrix}$$

Stationary distribution

$$\pi = \begin{pmatrix} \frac{1}{2} & \frac{1}{4} & \frac{1}{4} \end{pmatrix}$$

Stationary distribution: distribution π satisfying $\pi P = \pi$.

Theorem

For almost all (for the uniform measure) $x \in A^{\mathbb{N}}$ and for each state q,

$$\lim_{n\to\infty} \frac{|\gamma(x)[1:n]|_q}{n} = \pi_q$$

Normal sequences

A normal sequence is a sequence such that all finite words of the same length occur in it with the same limiting frequency.

If $x \in A^{\mathbb{N}}$ and $w \in A^*$, the frequency of w in x is defined by

$$\mathrm{freq}(x, w) = \lim_{n \to \infty} \frac{|x[1:n]|_w}{n}.$$

where $|z|_w$ denotes the number of occurrences of w in z.

A sequence $x \in A^{\mathbb{N}}$ is normal if for each finite word $w \in A^*$:

$$\mathrm{freq}(x, w) = \frac{1}{(\#A)^{|w|}}$$

where

- $\#A$ is the cardinality of the alphabet A
- $|w|$ is the length of the word w.

Normal sequences (continued)

Theorem (Borel, 1909)

The decimal expansion of almost every real number in $[0,1)$ *is a normal sequence over the alphabet* $\{0,1,\ldots,9\}$.

Nevertheless, not so many examples have been proved normal. Some of them are:

- Champernowne 1933 (natural numbers):

  $12345678910111213141516171819202122232425\cdots$

- Besicovitch 1935 (squares):

  $149162536496481100121144169196225256289324\cdots$

- Copeland and Erdős 1946 (primes):

  $23571113171923293137414347535961677173798 3\cdots$

Normality implies genericity

Theorem (Schnorr & Stimm)

Let $\gamma(x)$ be the run in a deterministic and strongly connected automaton. For each normal sequence $x \in A^{\mathbb{N}}$ and each state q,

$$\lim_{n\to\infty} \frac{|\gamma(x)[1:n]|_q}{n} = \pi_q.$$

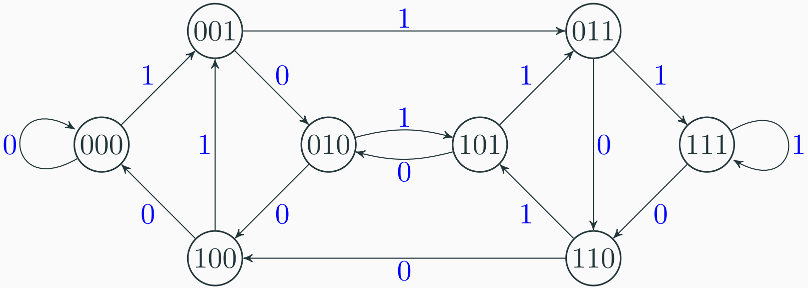
Corollary

Let $\gamma(x)$ be the run in a deterministic and strongly connected automaton. For each normal sequence $x \in A^{\mathbb{N}}$ and each finite run ξ starting from q,

$$\lim_{n\to\infty} \frac{|\gamma(x)[1:n]|_\xi}{n} = \frac{\pi_q}{(\#A)^{|\xi|}}.$$

Back to normality

Conversely, normality can be defined as genericity in the local automata *also known as* de Bruijn graphs.



Outline

Deterministic automata

Group automata

Genericity for a class of automata

For a class $\mathcal{C}$ of automata, a sequence $x \in A^{\mathbb{N}}$ is generic for $\mathcal{C}$ if for each state q of an automaton in $\mathcal{C}$

$$\lim_{n\to\infty} \frac{|\gamma(x)[1:n]|_q}{n} = \pi_q$$

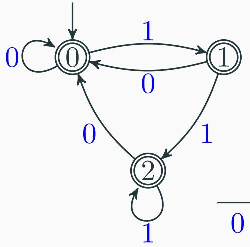
Rephrasing what is known:

- Normal sequences are generic for all automata.
- If the class $\mathcal{C}$ contains all local automata, then x is generic for $\mathcal{C}$ if and only if x is normal.

Otherwise, which sequences are generic for $\mathcal{C}$?

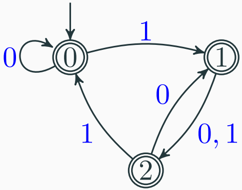
Group automata

An automaton is a group automaton if each symbol induces a permutation of the states. The stationnary distribution of a group automaton is the uniform distribution.



	0	1	2
0	0	0	0
1	1	2	2

Not a group automaton



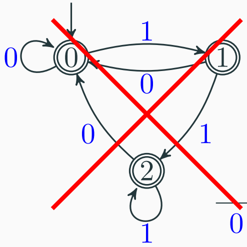
	0	1	2
0	0	2	1
1	1	2	0

A group automaton

Is there a sequence which is not normal but which is generic for the class of group automata ?

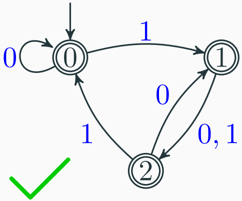
Group automata

An automaton is a group automaton if each symbol induces a permutation of the states. The stationnary distribution of a group automaton is the uniform distribution.



	0	1	2
0	0	0	0
1	1	2	2

Not a group automaton



	0	1	2
0	0	2	1
1	1	2	0

A group automaton

Is there a sequence which is not normal but which is generic for the class of group automata ?

Prouhet-Thue-Morse sequence

Consider the fixed point $x = \tau^\omega(0)$ where the substitution τ is

$$\tau \begin{cases} 0 \mapsto 01 \\ 1 \mapsto 10 \end{cases}$$

$$x = 0110100110010110100101 10\ldots$$

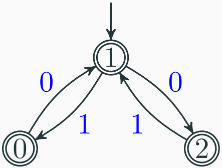
$$x_n = \begin{cases} 0 & \text{if the number of 1s in the binary expansion of } n \text{ is even} \\ 1 & \text{otherwise} \end{cases}$$

Prouhet-Thue-Morse sequence

Consider the fixed point $x = \tau^\omega(0)$ where the substitution τ is

$$\tau \begin{cases} 0 \mapsto 01 \\ 1 \mapsto 10 \end{cases}$$

$$x = 01\,10\,10\,01\,10\,01\,01\,10\,10\,01\,01\,10\ldots$$



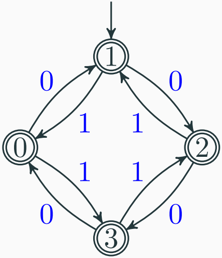
	0	1	2
0	1	2	
1		0	1

Prouhet-Thue-Morse sequence

Consider the fixed point $x = \tau^\omega(0)$ where the substitution τ is

$$\tau \begin{cases} 0 \mapsto 01 \\ 1 \mapsto 10 \end{cases}$$

$$x = 01\,10\,10\,01\,10\,01\,01\,10\,10\,01\,01\,10\ldots$$



	0	1	2	3
0	1	2	3	0
1	3	0	1	2

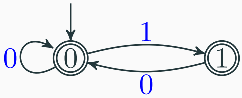
The Prouhet-Thue-Morse sequence is not generic for group automata.

Fibonacci sequence

Consider the fixed point $x = \tau^{\omega}(0)$ where the substitution τ is

$$\tau \begin{cases} 0 \mapsto 01 \\ 1 \mapsto 0 \end{cases}$$

$$x = 01\,0\,01\,01\,0\,01\,0\,01\,01\,0\,01\,01\,0\ldots$$

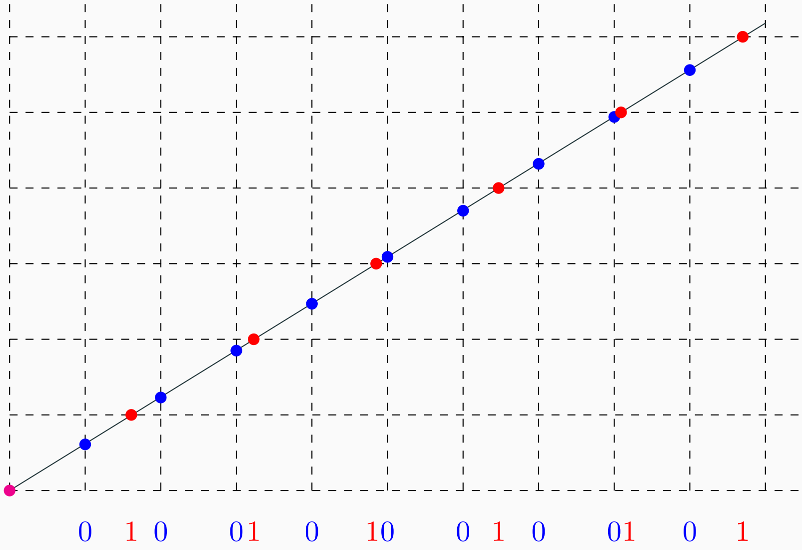


	0	1
0	0	0
1		1

The Fibonacci sequence might be generic for group automata. It is not possible to use the same trick as for the Prouhet-Thue-Morse sequence because $\{0, 01\}$ is a base of the free group F_2.

Sturmian sequences

Intersections of $y = \alpha x + \rho$ with the grid



Complexity function

The complexity function p_x of a sequence x is defined by

$$\begin{aligned} p_x(n) &= \text{number of distinct factors of length } n \text{ in } x \\ &= \#(F(x) \cap A^n). \end{aligned}$$

Theorem (Hedlung & Morse)
A sequence x is Sturmian iff $p_x(n) = n + 1$ for each $n \geqslant 1$.

Conjecture

Conjecture (C. & Delecroix)
Sturmian sequences are generic for group automata.

I have no idea why it really works but here are some hints:

- Sturmian sequences are a coding of rotations in $\mathbb{T}^1$ (a compact group).
- Sturmian are generared as S-adic sequences by the morphisms which generates the automorphisms of the free group F_2.
- ...

Conjecture

Conjecture (C. & Delecroix)
Sturmian sequences are generic for group automata.

I have no idea why it really works but here are some hints:

- Sturmian sequences are a coding of rotations in $\mathbb{T}^1$ (a compact group).
- Sturmian are generared as S-adic sequences by the morphisms which generates the automorphisms of the free group F_2.
- …

If you have better ideas, they are welcome.

Open questions

- What about larger alphabets (of size greater than 2) ?
- Sturmian sequences are not the only ones: factors of the form $0^{k!}$ can be inserted at sporadic positions. What else ?
- $\ldots$

Thank you