Word combinatorics and symbolic substit'ns

1. Intro, Def.'s, Basics
2. Kolakoski sequences
3. Sturmian sequences & selfdual substitutions

1. Alphabet $\mathcal{A} = \{a, b\}$ [or $\{a,b,c\}$ or... here mostly 2 letters]

$\mathcal{A}^* := \{ w = w_1 w_2 \dots w_n \mid w_i \in \mathcal{A} \}$

„words"

$\mathcal{A}^{\mathbb{N}} := \{ w_1 w_2 \dots \mid w_i \in \mathcal{A} \}$ infinite words

$\mathcal{A}^{\mathbb{Z}} := \{ \dots w_{-1} w_0 w_1 \dots \mid w_i \in \mathcal{A} \}$ biinfinite words

Complexity function: $p_w(m)$: nb of different subwords of w of length m.

Ex. $w = \dots abbabbabbabb \dots$ (periodic word: $\forall i \in \mathbb{Z}: w_i = w_{i+3}$)

here: $p_w(1) = 2$; $p_w(2) = 3$; $p_w(3) = 3$
(a, b) (ab, bb, ba) (abb, bba, bab)

In fact $p_w(m) = 3$ for $m \geq 2$ [constant!]

Ex. random word: $p_w(m) = 2^m$ [exponential!]
(with probability 1)

Q Is there a „square-free" word in $\mathcal{A}^{\mathbb{Z}}$?
i.e. a word not containing uu for any $u \in \mathcal{A}^*$

Try: ~~aa~~ ~~abb~~ ~~abab~~ No.

②

Q: Is there a cube-free word in $\mathcal{A}^{\mathbb{Z}}$?
(i.e. no uuu) Yes!

Thue-Morse-sequence. [One way to define it:]

A (word-)substitution is a map $\sigma : \mathcal{A} \to \mathcal{A}^*$

Here: $\sigma_{TM} : a \mapsto ab ; \ b \mapsto ba$

$$a \to ab \to abba \to abbabaab \to \dots$$

or better: $a|a \to ab|ab \to abba|abba \to$

$abbabaab|abbabaab \to abbabaabbaababba|abbabaabb \dots$

$(\sigma_{TM})^2(a|a)$; $(\sigma_{TM})^4(a|a)$; $(\sigma_{TM})^6(a|a)$, ...

converges to some $w_{TM} \in \mathcal{A}^{\mathbb{Z}}$ (with $(\sigma_{TM})^2(w) = w$)

and w_{TM} is cube free [see B.G. p. 100]

In particular, w_{TM} is nonperiodic. (!)

Complexity? $p_w(m)$: $\underset{(m=1)}{2}, 4, 6, 10, \underset{(m=5)}{12}, 16, 20, \dots$

linear! $p_w(m) \in O(m)$

Frequency? [ratio #a's : #b's]

Can be obtained from the substitution matrix

$$M_\sigma = \begin{pmatrix} \#a\text{'s in } \sigma(a) & \#a\text{'s in } \sigma(b) \\ \#b\text{'s in } \sigma(a) & \#b\text{'s in } \sigma(b) \end{pmatrix}$$

Here: $M_{\sigma_{TM}} = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$

③ **Thm.** (Perron): A matrix $M \geq 0$ [i.e. all entries ≥ 0] which is primitive (i.e. $\exists k: M^k > 0$) has a real, positive, simple eigenvalue λ such that $|\lambda| > |\lambda'|$ for all eigenvalues $\lambda' \neq \lambda$ of M. Moreover, λ has a strictly positive eigenvector.

Thm If σ is a substitution, M_σ is primitive, then the Perron-eigenvalue λ is the inflation factor of σ; and the normalised eigenvector of λ contains the frequencies of the letters.

Ex.: $M_{\sigma_{TM}} = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$ eigenvalues $\underline{\underline{2}}, 0$

(normed) eigenvector of 2: $\begin{pmatrix} 1/2 \\ 1/2 \end{pmatrix}$

So $\text{freq}(a) = \frac{1}{2} = \text{freq}(b)$.

2. Kolakoski sequences [OEIS: A000002]

K: 2211212212211211221211... ($\mathcal{A} = \{1,2\}$)

2 2 1 1 2 1 2 2 1 2 2 1 1 ...

[uniquely def. up to leading 1]

Q
- $\text{freq}(1) = ?$
- If v subword in K, does v occur again? infinitely many often? with bounded gaps?

Open! Best bounds: $0.499999322\ldots < \text{freq}(1)$

$< 0.500000677\ldots$

[Johan Nilsson, Bielefeld]

④ [One may also consider similar sequences]

$K_{1,3}$: 3331113331313331113331333...

$K_{2,4}$: 4444222244442222442244224444....

freq (1) in $K_{1,3}$? [solve by substitution!]

Let $A = 33$; $B = 31$, $C = 11$

$\sigma: A \mapsto ABC, \ B \mapsto AB, \ C \mapsto B$

ABCABBABCABABABCABB...
3331113331313331111.....

$$M_\sigma = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 1 \\ 1 & 0 & 0 \end{pmatrix}$$ primitive, $\lambda = 2{,}205569...$
[eigenvector of λ yields]

freq(A) = 0,376... freq(B) = 0,453... freq(C) = 0,171...

$\Rightarrow$ freq(1) = 0,397... freq(3) = 0,603

Similar: $K_{2,4}$: $A = 44$, $B = 22$

$\sigma: A \mapsto AABB, \ B \mapsto AB$

$$M_\sigma = \begin{pmatrix} 2 & 1 \\ 2 & 1 \end{pmatrix}$$ $\lambda = 3$, freq(A) $= \frac{1}{2} =$ freq(B)

$\Rightarrow$ freq(2) $= \frac{1}{2} =$ freq(4).

General:

- $K_{n,m}$ for n and m even: freq(n) $= \frac{1}{2} =$ freq(m)
- $K_{n,m}$ for n and m odd: freq can be computed in each case.
- $K_{n,m}$ for n even, m odd: open.

[Bernd Sing]
[until 2007: Bielefeld
now Barbados]

Sturmian seq. and dual substitutions

If $w \in \mathcal{A}^{\mathbb{Z}}$ is periodic, then [for the complexity holds] $\exists C : \quad p_w(m) \leq C$ for all m

($p_w \in O(1)$).

If $w \in \mathcal{A}^{\mathbb{Z}}$ is non-periodic, then $p_w(m) \geq m+1$

($p_w \in O(m)$) [In particular, there is no w s.t. $p_w \in O(\log m)$ or $O(\sqrt{m})$!]

<u>Def</u> w is called <u>Sturmian</u> [seq. or word], if $p_w(m) = m+1$, and w is <u>repetitive</u>

Consider ....aaaaabbbbb..... has $p_w(m) = m+1$

but looks boring. Hence:

<u>repetitive</u>: each subword of w occurs infinitely many often in w with bounded gaps.

Mignosi–Séébold '93

<u>Thm</u> Let σ: be a primitive substitution (on $\mathcal{A} = \{a, b\}$) and $u \in \mathcal{A}^{\mathbb{Z}}$ such that $\sigma(u) = u$

σ invertible $\Leftrightarrow$ u Sturmian

<u>„invertible"</u>: $\sigma : F_2 \to F_2$ as a group endomorphism.

[in other words: $\sigma \in Aut(F_2)$, or: σ bijection]

Then we have $\det M_\sigma = \pm 1$

⑥ <u>Ex</u> Fibonacci (squared):

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

invertible? Try: • $b = \sigma^{-1}(ab) = \sigma^{-1}(a)\sigma^{-1}(b)$ ⊛

• $a = \sigma^{-1}(aba) = \sigma^{-1}(ab)\sigma^{-1}(a) = b\,\sigma^{-1}(a)$

$\Rightarrow b^{-1}a = \sigma^{-1}(a)$, and from

⊛ $\Rightarrow b = b^{-1}a\,\sigma^{-1}(b) \Rightarrow a^{-1}bb = \sigma^{-1}(b)$

Hence σ invertible, and the Fibonacci sequence

$$a \to aba \to abaababa \to abaababaabaababaabab \ldots$$

[limit of $\sigma^n(a|a)$] is a Sturmian seq.

[Here something more happens: Consider σ^{-1} as a substitution]

$$\sigma^{-1}: \{a^{-1}, b\} \to \{a^{-1}, b\}^*,$$

$$\sigma^{-1}: a^{-1} \mapsto a^{-1}b \ ; \ b \mapsto a^{-1}bb$$

Fibo: $\ldots \,|\, abaababaabaababa \ldots$

σ^{-1}: $\ldots\, \bar{a} \,|\, b\bar{a}bb\bar{a}b\bar{a}bb\bar{a}bb\bar{a}b\bar{a}b \ldots$

essentially the same sequence! (up to $\bar{a} \mapsto b ; b \to \bar{a}$)

<u>Def</u> $\sigma \in Aut(F_2)$ <u>selfdual</u>, if $\sigma \sim \tau\sigma\tau$

with $\tau: a \mapsto a^{-1} ; b \mapsto b$ or $a \mapsto b^{-1}, b \mapsto a$

<u>Thm</u> (Berthé-F-Sirvent) Let $\sigma \in Aut(F_2)$, $\det M_\sigma = 1$

$$\sigma \text{ selfdual} \Leftrightarrow M_\sigma = \begin{pmatrix} m & K \\ \frac{m^2-1}{K} & m \end{pmatrix} \text{ or } \begin{pmatrix} m & K \\ K & \frac{m^2+1}{K} \end{pmatrix}$$

$(K, m \geq 1 ; K = n \cdot (m^2 \pm 1))$