Departamento de Computación, Facultad de Ciencias Exactas y Naturales, UBA

Azar, Algoritmos y Autómatas

Clase 2:
Lindas secuencias normales

Notation.

An alphabet is a finite set of symbols. We call it A.
A^{ω} is the set of all infinite words.
A^{*} is the set of all finite words the set of all finite words.
$A^{\leq k}$ is the set of all words of length up to k.
A^{k} for the set of words of length exactly k.

If w is a word, $|w|$ is its length.

The positions of finite and infinite words are numbered starting at 1.

To denote the symbol at position i of a word w we write $w[i]$ and to denote the substring of w from position i to j we write $w[i \dots j]$.

The empty word is denoted by λ.

Notation

Number of occurrences, not-aligned and aligned,

$$
\begin{aligned}
|w|_u &= |\{i : w[i \ldots i + |u| - 1] = u\}|, \\
\|w\|_u &= |\{i : w[i \ldots i + |u| - 1] = u \text{ and } i \equiv 1 \mod |u|\}|.
\end{aligned}
$$

For example, $|aaaaa|_{aa} = 4$ and $\|aaaaa\|_{aa} = 2$.

Notice that the definition of aligned occurrences has the condition $i \equiv 1 \mod |u|$ instead of $i \equiv 0 \mod |u|$, because the positions are numbered starting at 1.

When a word u is just a symbol, $|w|_u$ and $\|w\|_u$ coincide.

Counting aligned occurrences

Aligned occurrences of a word of length r over alphabet A coincide with occurrences of the corresponding symbol over alphabet A^r.

Consider alphabet A, a length r and alphabet B with $|A|^r$ symbols.
The set of words of length r over alphabet A and the set B are isomorphic:

$$\pi : A^r \to B$$

induced by the lexicographic order in the respective sets.

Thus, for any $w \in A^*$ such that $|w|$ is a multiple of r,

$$|\pi(w)| = |w|/r.$$

Then,

$$\forall\ u \in A^r\ (\|w\|_u = |\pi(w)|_{\pi(u)}).$$

Example. Suppose $A = \{0, 1\}$, $r = 3$.
Consider B such that $|A|^r = |B|$, $B = \{0, 1, 2, 3, 4, 5, 6, 7\}$

100 100 111 000

4470

Representation of real numbers

A base is an integer greater than or equal to 2. For a positive real number x, the expansion of x in base b is a sequence $a_1 a_2 a_3 \dots$ of integers from $\{0, 1, \dots, b-1\}$ such that

$$x = \lfloor x \rfloor + \sum_{k \geq 1} a_k b^{-k} = \lfloor x \rfloor + 0.a_1 a_2 a_3 \dots$$

To have a unique representation of all rational numbers we require that expansions do not end with a tail of $b-1$.

We will abuse notation and whenever the base b is understood we will denote the first n digits in the expansion of x with $x[1 \dots n]$.

Normal numbers

Definition (Non-aligned normality, Borel first: Niven and Zuckerman 1951)

A real number x is normal to base b if for every block u,

$$\lim_{n\to\infty} \frac{|x[1\ldots n]|_u}{n} = \frac{1}{b^{|u|}}.$$

Normality as a seemingly weaker condition

Theorem (Piatetski-Shapiro 1957)

Let x be a real and let b be an integer greater than or equal to 2.
Let $A = \{0, \ldots, b-1\}$. The following conditions are equivalent,

1. *The real x is normal to base b.*
2. *There is a constant C such that for infinitely many lengths ℓ and for every w in A^ℓ*

$$\limsup_{n\to\infty} \frac{|x[1\ldots n]|_w}{n} < C \cdot b^{-\ell}.$$

3. *There is a constant C such that for infinitely many lengths ℓ and for every w in A^ℓ*

$$\limsup_{n\to\infty} \frac{\|x[1\ldots n\ell]\|_w}{n} < C \cdot b^{-\ell}.$$

Dos Secuencias normales

- Secuencias de Bruijn infinitas
- À la Champernowne

01 00 01 10 11 000 001 010 011 100 101 110 111 0000 . . .

Secuencias de Bruijn

Definition (de Bruijn 1946; Sainte-Marie 1894)

A de Bruijn necklace of order n over alphabet A is a cyclic sequence of length $|A|^n$ such that every word of length n occurs in it exactly once.

A (non cyclic) de Bruijn word of order n over alphabet A is a word of length $|A|^n + n - 1$ such that every word of length n occurs in it exactly once.

Examples of de Bruijn necklaces : 01; Examples of de Bruijn words :

01;

Secuencias de Bruijn infinitas

Definition

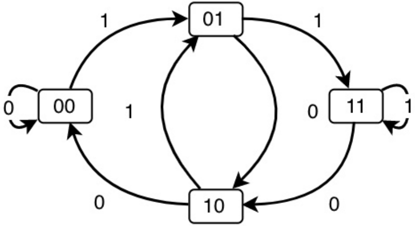
An infinite de Bruijn word $w = a_1 a_2 \ldots$ in an alphabet of at least three symbols is an infinite word such that, for every n, $a_1 \ldots a_{|A|^n+n-1}$ is a de Bruijn word of order n.

Example: 012;   can be extended to

In case the alphabet has two symbols, an infinite de Bruijn word $w = a_1 a_2 \ldots$ is such that, for every odd n, $a_1 \ldots a_{|A|^n+n-1}$ is a de Bruijn word of order n.

de Bruijn graph

A de Bruijn graph $G_A(n)$ is a directed graph whose vertices are the words of length n over alphabet A and whose edges are the pairs (v, w) where $v = au$ and $w = ub$, for some word u of length $n - 1$ and possibly two different symbols a, b.



The graph $G_A(n)$ has $|A|^n$ vertices and $|A|^{n+1}$ edges, it is strongly connected and every vertex has the same in-degree and out-degree.

Each de Bruijn sequence of order $n + 1$ over an alphabet of $|A|$ symbols can be constructed as an Eulerian cycle in $G_A(n)$.

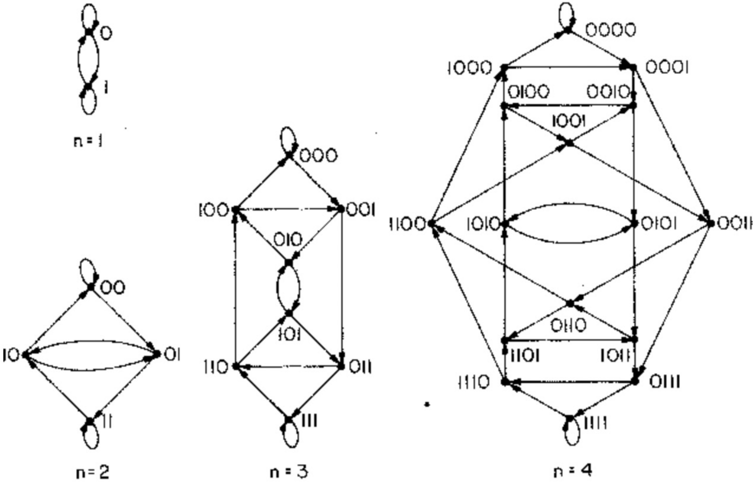


Fig. 3. The de Bruijn graphs of order $n = 1, 2, 3$, and 4.

Proposition (Becher and Heiber 2011)

Suppose alphabet A has at least three symbols. Every de Bruijn sequence of order n can be extended to a de Bruijn sequence of order $n+1$.

Demostracion: secuencias de Bruijn extendidas

Fix alphabet A. Suppose E is Eulerian in $G_A(n)$. Since $G_A(n+1)$ is the line graph of $G_A(n)$, E is Hamiltonian cycle in $G_A(n+1)$

We just need to argue that $G_A(n+1) \setminus G_A(n)$ is Eulerian, which means that it is strongly connected and regular (every vertex has the same in and out degree).
Clearly, every vertex in $G_A(n+1) \setminus G_A(n)$ has the same in and out degree.

A directed graph G in which each vertex has its in-degree equal to its out-degree is strongly connected if and only if its underlying undirected graph is connected.

It suffices to see that $G_A(n+1) \setminus G_A(n)$ is connected. $\square$

Computing extended de Bruijn words

There is an obvious algorithm to compute an infinite de Bruijn word which, for each $n \geq 1$, extends a Hamiltonian cycle in a de Bruijn graph of order n to an Eulerian cycle in the same graph.

This is done in time exponential in n. No efficient algorithm is known to compute the N-th symbol of an infinite de Bruijn word without computing the first N symbols.

Lemma

Fix $u \in A^\ell$. Then u occurs in a de Bruijn word of order n between $|A|^{n-\ell}$ and $|A|^{n-\ell} + n - \ell$ times.

Proof.

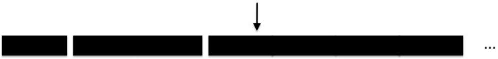
At each position of a de Bruijn word of length n starts a new word of length n. There are exactly $|A|^{n-\ell}$ words of length n whose first ℓ symbols are u.
Since the de Bruijn word of order n has length $|A|^n + n - 1$, There are exactly $n - \ell$ other positions in a de Bruijn word of order n where a word of length ℓ may start. □

Theorem (Ugalde 2000)

Infinite de Bruijn words are normal.

In case the alphabet A has two symbols, consider the instead the words in the alphabet A' of four symbols obtained by the morphism mapping blocks two symbols in A to one symbol in A' and prove normality for alphabet A'.

A sequence is normal if every block of digits occurs with the same limiting frequency as every other block of the same length.



To prove it, the count at an anbritrary position is bounded by considering the count at the end of the megablock. □.

Demostración: normalidad de las Bruijn infinitas

Let $x = a_1 a_2 \ldots$ be an infinite de Bruijn word over A.
Then, for each n, $a_1 \ldots a_{|A|^n+n-1}$ is a de Bruijn word or order n.
Fix a position N and let n be such that

$$|A|^n + n - 1 \leq N < |A|^{n+1} + n.$$

Then,

$$\frac{|a_1 \ldots a_N|_u}{N} \leq \frac{|a_1 \ldots a_{|A|^{n+1}+n}|_u}{|A|^n + n - 1} \leq \frac{|A|^{n+1-\ell} + n - \ell}{|A|^n + n - 1} < 2\ |A|^{-\ell+1}.$$

Thus,

$$\limsup_{N\to\infty} \frac{|a_1 \ldots a_N|_u}{N} < 2|A|^{-\ell+1}.$$

By Theorem 1, using $C = 2\ |A|$, x is normal. □

Our observation

Consider all blocks of length n, concatenated in lexicographical order, view it circularly. Each block of length n occurs exactly n times at positions with different modulo n.

For example, for alphabet $\{0, 1\}$

$n = 2$ position

12 34 56 78

00 01 10 11

0 0 0 1 10 11 00 occurs twice, at positions different modulo 2

00 01 10 11

00 01 1 0 1 1 01 occurs twice, at positions different modulo 2

00 01 10 11

0 0 01 10 1 1 10 occurs twice, at positions different modulo 2

00 0 1 1 0 11

00 01 10 11 11 occurs twice, at positions different modulo 2

Our observation

$n = 3$

000 001 010 011 100 101 110 111   000 occurs three times,
0 00 0 01 010 011 100 101 110 111   at positions different modulo 3
00 0 00 1 010 011 100 101 110 111

000 001 010 011 100 101 110 111   001 occurs three times
000 001 010 011 1 00 1 01 110 111   at postions different modulo 3
000 001 01 0 01 1 100 101 110 111

⋮

However ...

Not every permutation of the blocks of length n has the property:

00 10 11 01

000 101 001 010 011 100 110 111

Perfect necklaces

Definition (Alvarez, Becher, Ferrari and Yuhjtman 2016)

A necklace over a b-symbol alphabet is (n,k)-perfect if each block of length n occurs k times, at positions different modulo k, for any convention of the starting point.

De Bruijn necklaces are exactly the $(n,1)$-perfect necklaces.

The (n,k)-perfect necklaces have length kb^n.

Arithmetic progressions yield perfect necklaces

Identify the blocks of length n over a b-symbol alphabet with the set of non-negative integers modulo b^n according to representation in base b.

Theorem (Alvarez, Becher, Ferrari and Yuhjtman 2016)

Let r coprime with b. The concatenation of blocks corresponding to the arithmetic sequence $0, r, 2r, ..., (b^n - 1)r$ yields an (n, n)-perfect necklace.

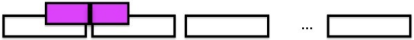
With $r = 1$ we obtain the lexicographically ordered sequence.

Arithmetic progressions yield perfect necklaces

Lemma

Let $\sigma : \{0, .., b-1\}^n \to \{0, .., b-1\}^n$ be such that for any block v of length n $\{\sigma^j(v) : j = 0, ..., b^n - 1\}$ is the set of all blocks of length n.

The necklace $[\sigma^0(v)\sigma^1(v) \ldots \sigma^{b^n-1}(v)]$ is (n, n)-perfect if and only if for every block u of length n, for every $\ell = 0, \ldots, n-1$ there is a unique block v of length n such that $v(n-\ell-1 \ldots n) = u(1 \ldots \ell)$ and $(\sigma(v))(1 \ldots n-\ell) = u(\ell+1 \ldots n)$.



For every length-n block splitted in two parts, there is exactly one matching (a tail of a block and the head of next block).

Demostración del Lema

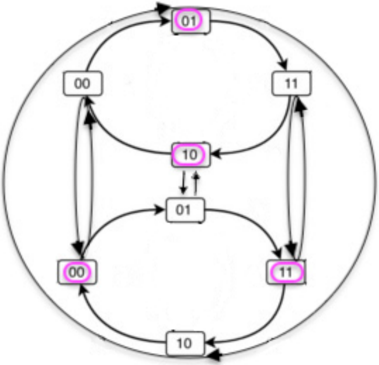
Assume s is (n,n)-perfect. Take ℓ such that $0 \leq \ell < n$, $x \in A^{\ell}$ and $y \in A^{n-\ell}$. Consider $\theta^{-\ell}s$, the $-\ell^{th}$ shift of s. Since s is ((n,n)-perfect, xy occurs exactly once in the decomposition of $\theta^{-\ell}s$ in consecutive words of length n. Thus, there is a unique word w in the decomposition of s in consecutive words of length n whose last ℓ symbols are equal to x and whose first $n-\ell$ symbols are equal to y.

Conversely, suppose s is not (n,n)-perfect. Then, there is some ℓ, $0 \leq \ell < n$, such that the decomposition of $\theta^{-\ell}(s)$ contains two equal words of length n. This contradicts that for every $x \in A^{\ell}$ and every $y \in A^{n-\ell}$, there is a unique $w \in A^{k}$ such that $w(n-\ell \ldots n-1) = x$ and $(\sigma(w))(0 \ldots n-\ell-1) = y$. $\square$

Astute graphs

Fix A an alphabet with b-symbols.
The astute graph $G_A(n,k)$ is directed, with kb^n vertices.
The set of vertices is $\{0,..b-1\}^n \times \{0,..,.k-1\}$.
There is an edge $(w,m) \rightarrow (w',m')$ if $w(2..n) = w'(1..n-1)$ and $(m+1) \mod k = m'$

$G_A(2,2)$



Astute graphs

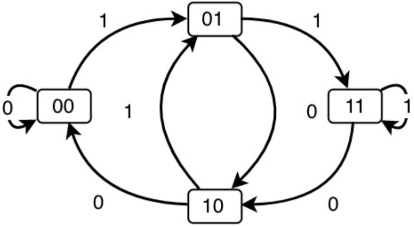
Observation

$G_{b,n,k}$ is Eulerian because it is strongly regular and strongly connected.

Observation

$G_{b,n,1}$ is the de Bruijn graph of blocks of length n over b-symbols.

$G_{2,2,1}$



Eulerian cycles in astute graphs

Each Eulerian cycle in $G_{b,n-1,k}$ gives one (n,k)-perfect necklace.

Each (n,k)-perfect necklace can come from many Eulerian cycles in $G_{b,n-1,k}$

Theorem (Alvarez, Becher, Ferrari and Yuhjtman 2016)

The number of (n,k)-perfect necklaces over a b-symbol alphabet is

$$\frac{1}{k}\sum_{d_{b,k}|j|k} e(j)\varphi(k/j)$$

where

- $d_{b,k} = \prod p_i^{\alpha_i}$, *such that $\{p_i\}$ is the set of primes that divide both b and k, and α_i is the exponent of p_i in the factorization of k,*
- $e(j) = (b!)^{jb^{n-1}} b^{-n}$ *is the number of Eulerian cycles in $G_{b,n-1,j}$*
- φ *is Euler's totient function*

Normal sequences as sequences of Eulerian cycles

Theorem

The concatenation of (n, k)-perfect necklaces over a b-symbol alphabet, for arithmetically increasing (n, k) is normal to the b-symbol alphabet.

Proof.

A number is normal to base b if in its base-b expansion every block of digits occurs with the same limiting frequency as every other block of the same length.

To prove that the sequence of megablocks is normal the count at an abritrary position is bounded by considering the count at the end of the megablock. The proof is a direct application of Piatetski-Shapiro theorem. □

La secuencia de Champernowne es normal

Corollary

The concatenation of lexicografically ordered (n, n)-perfect necklaces for $n = 1, 2, \ldots$ is normal; Champernowne's sequence is normal.

Nested perfect necklaces

Definition

An (n,k)-perfect necklace over a b-symbol alphabet is nested if $n=1$ or it is the concatenation of b nested $(n-1,k)$- perfect necklaces.

For example, for alphabet $\{0,1\}$, a nested $(2,2)$-perfect necklace

$$\underbrace{0011}_{\text{(1,2)-perfect}}\ \underbrace{0110}_{\text{(1,2)-perfect}}$$

The lexicographic order yields a perfect necklace but not nested,

$$\underbrace{00\ 01\ 02}_{\text{not (1,2)-perfect}}\ \underbrace{10\ 11\ 12}_{\text{not (1,2)-perfect}}\ \underbrace{20\ 21\ 22}_{\text{not (1,2)-perfect}}$$

Nested perfect necklaces

These following 8 blocks are $(1,4)$-perfect necklaces:

00001111	01011010
00111100	01101001
00011110	01001011
00101101	01111000

The concatenation in each row is a $(2,4)$-perfect necklace.
The concatenation of the first two rows is a nested $(3,4)$-perfect necklace.
The concatenation of the last two rows is a nested $(3,4)$-perfect necklace.
The concatenation of all rows is a nested $(4,4)$-perfect necklace.

Nested perfect necklaces

Supongamos x es nested $(2^m, 2^m)$ perfect necklace.
2 nested $(2^m - 1, 2^m)$ perfect necklaces.
4 nested $(2^m - 2, 2^m)$ perfect necklaces.
...
2^i nested $(2^m - i, 2^m)$ perfect necklaces.
2^{m-1} nested $(2^1, 2^m)$ perfect necklaces.
2^m nested $(1, 2^m)$ perfect necklaces.

Nested perfect necklaces

Lemma

Assume a b-symbol alphabet. For a nested (n, n)-perfect necklace x,

- *each block of length n occurs n times in x, at positions with different congruence modulo n.*
- *for every $i = 1, \dots n$, x is the concatenation of b^{n-i} nested (i, n)-perfect necklaces. So, in every segment of length nb^i starting at a position multiple of nb^i, each block of length i occurs $1 \pm 2\varepsilon$ times, for $\varepsilon \in \{0, 1\}$ at positions in each of the n congruence classes.*

Definition

$$\Delta_{\ell,N}(x) = \max_{|v|=\ell} \left| \frac{|x[1,N]|_v}{N} - \frac{1}{|A|^\ell} \right| .$$

Fix alphabet A.

Lemma

If x is a nested $(2^m, 2^m)$-perfect necklace then for every ℓ such that $1 \leq \ell \leq m$, and for every N such that $1 \leq N \leq 2^m 2^{2^m}$,

$$\Delta_{\ell,N}(x) = O((2^m)^2/N)$$

Fix alphabet A.

Lemma

Let x be concatenation of nested $(2^d, 2^d)$-perfect necklaces for $d = 0, 1, 2, \ldots$. Then, for every ℓ, there is N_ℓ such that for every $N \geq N_\ell$,

$$\Delta_{\ell,N}(x) = O\big((\log N)^2/N\big).$$

Nested perfect necklaces

Theorem (Levin 1999; Becher and Carton 2019)

The number whose expansion is the concatenation of nested $(2^d, 2^d)$-perfect necklaces for $d = 0, 1, 2, \ldots$ is normal.

Nested perfect necklaces

Theorem (Becher and Carton 2019)

The base b-expansion of the number defined by M. Levin 1999 for base b using the Pascal triangle matrix modulo 2 is the concatenation of nested $(2^d, 2^d)$-perfect necklaces for $d = 0, 1, 2, \ldots$.

Levin's number

For $d = 0, 1, 2, \ldots$ Levin defines the matrix M_d in $\mathbb{F}_2^{2^d \times 2^d}$ (siguiente diapositiva).
Fix base b. Consider the elements of $\mathbb{F}_b^{2^d}$ in increasing order

$$w_0, w_1, \ldots, w_{b^{2^d}-1}$$

Identify vectors of $\mathbb{F}_b$ with blocks of symbols in $\{0, .., b-1\}$. Thus, each $(M_d w_i)$ is identified with a block of length 2^d.
Levin's number the number in the unit interval whose base b-expansion is

$$\lambda = 0.\lambda_0 \lambda_1 \lambda_2 \ldots$$

where $d = 0, 1, 2, \ldots$ define λ_d as

$$\lambda_d = (M_d w_0) \ldots (M_d w_{b^{2^d}-1})$$

Pascal triangle matrices modulo 2

Define a family of matrices using Pascal triangle modulo 2,

$$\begin{array}{cccccc} \dots & 1 & 1 & 1 & 1 & 1 \\ \dots & 5 & 4 & 3 & 2 & 1 \\ \dots & 15 & 10 & 6 & 3 & 1 \\ \dots & 35 & 20 & 10 & 4 & 1 \\ \dots & 70 & 35 & 15 & 5 & 1 \\ & \vdots & \vdots & \vdots & \vdots & \vdots \end{array} \longrightarrow \begin{array}{cccccc} \dots & 1 & 1 & 1 & 1 & 1 \\ \dots & 1 & 0 & 1 & 0 & 1 \\ \dots & 1 & 0 & 0 & 1 & 1 \\ \dots & 1 & 0 & 0 & 0 & 1 \\ \dots & 0 & 1 & 1 & 1 & 1 \\ & \vdots & \vdots & \vdots & \vdots & \vdots \end{array}$$

Pascal triangle matrices modulo 2

Define a family of matrices using Pascal's triangle modulo 2,

$$\begin{array}{cccccc} \dots & 1 & 1 & 1 & 1 & 1 \\ \dots & 5 & 4 & 3 & 2 & 1 \\ \dots & 15 & 10 & 6 & 3 & 1 \\ \dots & 35 & 20 & 10 & 4 & 1 \\ \dots & 70 & 35 & 15 & 5 & 1 \\ & \vdots & \vdots & \vdots & \vdots & \vdots \end{array} \longrightarrow \begin{array}{cccccc} \dots & 1 & 1 & 1 & 1 & 1 \\ \dots & 1 & 0 & 1 & 0 & 1 \\ \dots & 1 & 0 & 0 & 1 & 1 \\ \dots & 1 & 0 & 0 & 0 & 1 \\ \dots & 0 & 1 & 1 & 1 & 1 \\ & \vdots & \vdots & \vdots & \vdots & \vdots \end{array}$$

For $d = 0$, M_d has dimension $2^0 \times 2^0$

$$M_0 = \begin{pmatrix} 1 \end{pmatrix}$$

Matrices de Pascal Módulo 2

Define a family of matrices using Pascal's triangle modulo 2,

$$\begin{array}{cccccc} \dots & 1 & 1 & 1 & 1 & 1 \\ \dots & 5 & 4 & 3 & 2 & 1 \\ \dots & 15 & 10 & 6 & 3 & 1 \\ \dots & 35 & 20 & 10 & 4 & 1 \\ \dots & 70 & 35 & 15 & 5 & 1 \\ & \vdots & \vdots & \vdots & \vdots & \vdots \end{array} \longrightarrow \begin{array}{cccccc} \dots & 1 & 1 & 1 & 1 & 1 \\ \dots & 1 & 0 & 1 & 0 & 1 \\ \dots & 1 & 0 & 0 & 1 & 1 \\ \dots & 1 & 0 & 0 & 0 & 1 \\ \dots & 0 & 1 & 1 & 1 & 1 \\ & \vdots & \vdots & \vdots & \vdots & \vdots \end{array}$$

For $d = 1$, M_d has dimension $2^1 \times 2^1$

$$M_1 = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$$

Matrices de Pascal Módulo 2

Define a family of matrices using Pascal's triangle modulo 2,

$$\begin{array}{cccccc}
\dots & 1 & 1 & 1 & 1 & 1 \\
\dots & 5 & 4 & 3 & 2 & 1 \\
\dots & 15 & 10 & 6 & 3 & 1 \\
\dots & 35 & 20 & 10 & 4 & 1 \\
\dots & 70 & 35 & 15 & 5 & 1 \\
 & \vdots & \vdots & \vdots & \vdots & \vdots
\end{array}
\longrightarrow
\begin{array}{cccccc}
\dots & 1 & 1 & 1 & 1 & 1 \\
\dots & 1 & 0 & 1 & 0 & 1 \\
\dots & 1 & 0 & 0 & 1 & 1 \\
\dots & 1 & 0 & 0 & 0 & 1 \\
\dots & 0 & 1 & 1 & 1 & 1 \\
 & \vdots & \vdots & \vdots & \vdots & \vdots
\end{array}$$

For $d = 1$, M_d has dimension $2^1 \times 2^1$

$$M_1 = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$$

Matrices de Pascal Módulo 2

Define a family of matrices using Pascal's triangle modulo 2,

$$\begin{array}{cccccc} \dots & 1 & 1 & 1 & 1 & 1 \\ \dots & 5 & 4 & 3 & 2 & 1 \\ \dots & 15 & 10 & 6 & 3 & 1 \\ \dots & 35 & 20 & 10 & 4 & 1 \\ \dots & 70 & 35 & 15 & 5 & 1 \\ & \vdots & \vdots & \vdots & \vdots & \vdots \end{array} \longrightarrow \begin{array}{cccccc} \dots & 1 & 1 & 1 & 1 & 1 \\ \dots & 1 & 0 & 1 & 0 & 1 \\ \dots & 1 & 0 & 0 & 1 & 1 \\ \dots & 1 & 0 & 0 & 0 & 1 \\ \dots & 0 & 1 & 1 & 1 & 1 \\ & \vdots & \vdots & \vdots & \vdots & \vdots \end{array}$$

For $d = 2$, M_d has dimension $2^2 \times 2^2$

$$M_2 = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

Alternative formulation Pascal triangle matrices modulo 2

$$M_0 = (1), \qquad M_{d+1} = \begin{pmatrix} M_d & M_d \\ 0 & M_d \end{pmatrix}$$

- M_d in $\mathbb{F}_2^{2^d \times 2^d}$.
- M_d is invertible.
- The first row of M_d is the vector of 1s
- The last column of M_d is the vector of 1s

Invertible submatrices

$$M_d = \left(\begin{array}{cc} & \boxed{} \\ & k \end{array}\right) \qquad M_d = \left(\begin{array}{cc} \boxed{} & \\ k & \\ & \end{array}\right)$$

Lemma (Levin 1999 from Bicknell and Hoggart 1978)

For $d \geq 0$, the following submatrices of M_d are invertible

- *k rows and the last k columns*
- *the first k rows and k columns*

Levin's number

$$
\begin{aligned}
\lambda \quad = \quad 0. \quad & \underbrace{0\,1}_{\lambda_0} \\
& \underbrace{00\;11\;10\;01}_{\lambda_1} \\
& \underbrace{0000\;1111\;1010\;0101\;1100\;0011\;0110\;1001\;1000\;0111\;0010\;1101\;0100\;1011\;1110\;0001}_{\lambda_2} \\
& \dots
\end{aligned}
$$

Observation

Assume $b = 2$. For every $d \geq 0$, λ_d is the concatenation of all blocks of length 2^d in some order.

Levin's number

Observation

Assume $b = 2$. For every d and for every even n, $M_d w_n$ and $M_d w_{n+1}$ are complementary blocks.

$$\begin{array}{lcll}
\lambda & = & 0. & 0\,1 \\
 & & & 00\,11\;\;10\,01 \\
 & & & 0000\,1111\;\;1010\,0101\;\;1100\,0011\;\;0110\,1001\;\;1000\,0111\;\;0010\,1101\;\;0100\,1011\;\;1110\,0001 \\
 & & & \ldots
\end{array}$$

Theorem

Assume $b = 2$. For each $d = 0, 1, 2, \ldots$ there are $2^{2^{d+1}-1}$ binary nested $(2^d, 2^d)$-perfect necklaces.

Normal and self similar

For a given finite or infinite word $x = a_1 a_2 a_3 \dots$ where each a_i is a symbol in alphabet A, define

$$\text{even}(x) = a_2 a_4 a_6 \cdots$$

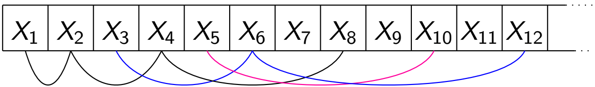
and

$$odd(x) = a_1 a_3 a_5 \cdots$$

Thus, $x = \text{even}(x)$ means that $a_n = a_{2n}$ for all n.

Theorem (Becher, Carton and Heiber 2017)

There is a normal word x such that $x = \text{even}(x)$.



Proof of Theorem

We consider an alphabet of 2 symbols. A finite word w is called ℓ-exact for an integer $\ell \geq 1$, if $|w|$ is a multiple of ℓ and all words of length ℓ have the same number of aligned occurrences in w.

Example: 000 001 010 011 100 101 110 111 is 3-exact

Lemma

Let w be an ℓ-exact word such that $|w|$ is a multiple of $\ell 2^{2\ell}$. Then, there exists a 2ℓ-exact word z of length $2|w|$ such that $\text{even}(z) = w$*.*

Proof.

Assume w is ℓ-exact and $|w|$ is a multiple of $\ell 2^{2\ell}$. Consider a factorization of $w = w_1 w_2 \cdots w_r$ such that for each i, $|w_i| = \ell$.
Thus, $r = |w|/\ell$.
Define z of length $2|w|$ as $z = z_1 z_2 \cdots z_r$ such that for each i, $|z_i| = 2\ell$, $even(z_i) = w_i$ and for all words u and u' of length ℓ, the set $\{i : z_i = u' \vee u\}$ has cardinality $r/2^{2\ell}$.
This latter condition is achievable because w is ℓ-exact, so for each word u of length ℓ, the set $\{i : \text{even}(z_i) = u\}$ has cardinality $r/2^{\ell}$ which is a multiple of 2^{ℓ}, the number of possible words u'. □

Corollary

Let w be an ℓ-exact word for some even integer ℓ. Then there exists an ℓ-exact word z of length $2|w|$ such that $\text{even}(z) = w$.

Proof.

Since w is ℓ-exact, it is also $\ell/2$-exact. Furthermore, if u and v are words of length $\ell/2$ and ℓ respectively then $\|w\|_u = 2^{\ell/2+1}\|w\|_v$. Thus, the hypothesis of Lemma 2 is fulfilled with $\ell/2$. □

Corollary

There exist a sequence $(w_n)_{n\geq 1}$ of words and a sequence of positive integers $(\ell_n)_{n\geq 1}$ such that $|w_n| = 2^n$, $\mathrm{even}(w_{n+1}) = w_n$, w_n is ℓ_n-exact and $(\ell_n)_{n\geq 1}$ is non-decreasing and unbounded. Furthermore, it can be assumed that $w_1 = 01$.

Proof.

We start with $w_1 = 01$, $\ell_1 = 1$, $w_2 = 1001$ and $\ell_2 = 1$. For each $n \geq 2$, if $\ell_n 2^{2\ell_n}$ divides $|w_n|$, then $\ell_{n+1} = 2\ell_n$ and w_{n+1} is obtained by Lemma 2. Otherwise, $\ell_{n+1} = \ell_n$ and w_{n+1} is obtained by Corollary 19. Note that the former case happens infinitely often, so $(\ell_n)_{n\geq 1}$ is unbounded. Also note that each ℓ_n is a power of 2. □

Proof of Theorem

Let $(w_n)_{n\geq 1}$ be a sequence given by Corollary 20. Let $x = 11w_1w_2w_3\cdots$
Note that for each $k \geq 1$,

$$x[2^k+1..2^{k+1}] = w_k \text{ and } x[1..2^{k+1}] = 11w_1\cdots w_k.$$

The fact that $w_n = \mathrm{even}(w_{n+1})$ implies $x[2n] = x[n]$, for every $n \geq 3$.
The cases for $n = 1$ and $n = 2$ hold because $x[1..4] = 1101$.
We prove that x is normal. Consider an arbitrary index n_0.
By construction, w_{n_0} is ℓ_{n_0}-exact and for each $n \geq n_0$, w_n is also ℓ_{n_0}-exact. For every word u of length ℓ_{n_0} and for every $n \geq n_0$,

$$\|x[1..2^{n+1}]\|_u \leq \|x[1..2^{n_0}]\|_u + \|w_{n_0}\ldots w_n\|_u.$$

Then, for every N such that $2^n \leq N < 2^{n+1}$ and $n \geq n_0$,

$$\begin{aligned}
\frac{\|x[1..N]\|_u}{N/\ell_{n_0}} &\leq \frac{\|x[1..2^{n+1}]\|_u}{N/\ell_{n_0}} \\
&\leq \frac{\|x[1..2^{n_0}]\|_u + \|w_{n_0} \ldots w_n\|_u}{N/\ell_{n_0}} \\
&\leq \frac{\|x[1..2^{n_0}]\|_u}{2^n/\ell_{n_0}} + \frac{\|w_{n_0} \ldots w_n\|_u}{2^n/\ell_{n_0}} \\
&= \frac{\|x[1..2^{n_0}]\|_u}{2^n/\ell_{n_0}} + \frac{(2^{n_0} + \ldots + 2^n)/(\ell_{n_0} 2^{\ell_{n_0}})}{2^n/\ell_{n_0}} \\
&< \frac{\|x[1..2^{n_0}]\|_u}{2^n/\ell_{n_0}} + \frac{2}{2^{\ell_{n_0}}}.
\end{aligned}$$

For large values of N and n such that $2^n \leq N < 2^{n+1}$, the expression $\|x[1..2^{n_0}]\|_u/(2^n/\ell_{n_0})$ becomes arbitrarily small. We obtain for every word u of length ℓ_{n_0},

$$\limsup_{N\to\infty} \frac{\|x[1..N]\|_u}{N/\ell_{n_0}} \leq 3\ 2^{-\ell_{n_0}}.$$

The choice of ℓ_{n_0} was arbitrary, so the above inequality holds for each ℓ_n. Since $(\ell_n)_{n\geq 1}$ is unbounded, the hypothesis of Theorem 1 is fulfilled, with $C = 3$, so we conclude that x is normal.

For a positive integer r and a set $P = \{p_1, \ldots, p_r\}$ of r prime numbers, let T_P be the set of all Toeplitz sequences, that is, the set of all sequences $t_1 t_2 t_3 \cdots$ in A^ω such that for every $n \geq 1$ and for every $i = 1, \ldots, r$,

$$t_n = t_{np_i}.$$

It is possible to compute a normal word x such that $x = \mathit{even}(x)$ in linear time.

Problem

Construct a normal infinite word $a_1 a_2 \ldots$ such that for every n, $a_n = a_{2n} = a_{3n}$.