Insertion in normal numbers

Verónica Becher

Universidad de Buenos Aires

SASW09 Isaac Newton Institute for Mathematical Sciences
International Conference on Computability, Complexity and Randomness
June 7, 2022

Borel normal numbers

Let b be an integer greater than or equal to 2.

A real 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.

Borel normal numbers

Let b be an integer greater than or equal to 2.

A real 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.

Instead of expansions of numbers we talk about sequences of digits/symbols that are normal to a b-alphabet.

Modifying normal sequences

- Selection of subsequences
  Wall 1949, Agafonov 1968, Kamae and Weiss 1975
- Sums
  Rauzy 1976; Volkonoff 1979, Aistleitner 2017
- Transformations by finite state tranducers
  Carton and Orduna 2020
- Insertion in positions in a set of density zero
  Figueira 2002, Aistleitner 2017

Modifying effective Hausdorff dimension of sequences

The effective Hausdorff dimension of an infinite sequence x is a real number between 0 and 1 which measures the asymptotic information density of x (Lutz 2000, Mayordomo 2002).

Given a sequence x of effective Hausdorff dimension s, how much do we need to change x to obtain some y of dimension t?

For binary sequences,

Greenberg, J.S.Miller, Shen and Westrick, 2018

Le Goh, J. S. Miller, M. Soskova and Westrick, 2022

Modifying effective Hausdorff dimension of sequences

The effective Hausdorff dimension of an infinite sequence x is a real number between 0 and 1 which measures the asymptotic information density of x (Lutz 2000, Mayordomo 2002).

Given a sequence x of effective Hausdorff dimension s, how much do we need to change x to obtain some y of dimension t?

For binary sequences,

Greenberg, J.S.Miller, Shen and Westrick, 2018

Le Goh, J. S. Miller, M. Soskova and Westrick, 2022

How can we change dimension but from binary sequences to ternary ?

Today's question

Enlarge the alphabet

Given a normal sequence, how can we insert symbols so that the expanded sequence is normal to the enlarged alphabet?

Today's question

Enlarge the alphabet

Given a normal sequence, how can we insert symbols so that the expanded sequence is normal to the enlarged alphabet?

In other words, given a sequence of finite-state dimension 1, how can we insert symbols to obtain dimension 1 in the enlarged alphabet?

Two theorems

First theorem
Any-symbol insertion in some constructed normal sequences.

Second theorem (Zylber 2017,2021)
Just the new symbol insertion in arbitrary normal sequences.

Example of insertion

Theorem (Champernowne 1933)

The concatenation of all blocks in length-lexicographic order is normal

0 1 00 01 10 11 000 001 010 011 100 101 110 111 0000 0001 ...

Example of insertion

Theorem (Champernowne 1933)

The concatenation of all blocks in length-lexicographic order is normal

0 1 00 01 10 11 000 001 010 011 100 101 110 111 0000 0001 ...

If we enlarge the alphabet with one greater symbol,

0 1 00 01 10 11 000 001 010 011 100 101 110 111 0000 0001 ...

0 1 2 00 01 02 10 11 12 20 21 22 000 001 002 010 011 012 020 021 022 100 101 102 110 111 112 120 121 122 200 201 202 210 211 212 220 221 222 0000 0001...

Observation

Consider concatenation in lexicographic order of all blocks of length n. Viewed circularly, each block of length n occurs exactly n times at positions different modulo n.

positions

n=2 $\quad {\scriptstyle 12 \ \ 34 \ \ 56 \ \ 78}$

$00\ 01\ 10\ 11$

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

Observation

Consider concatenation in lexicographic order of all blocks of length n. Viewed circularly, each block of length n occurs exactly n times at positions different modulo n.

	positions	
n=2	$12\ 34\ 56\ 78$	
	$00\ 01\ 10\ 11$	
	$00\ 01\ 10\ 11$	00 occurs twice, at positions different modulo 2
	$00\ 01\ 10\ 11$	
	$00\ 01\ 10\ 11$	01 occurs twice, at positions different modulo 2

Observation

Consider concatenation in lexicographic order of all blocks of length n. Viewed circularly, each block of length n occurs exactly n times at positions different modulo n.

positions

n=2 $\quad$ 12 34 56 78

00 01 10 11

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

00 01 10 11

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

00 01 10 11

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

Observation

Consider concatenation in lexicographic order of all blocks of length n. Viewed circularly, each block of length n occurs exactly n times at positions different modulo n.

	positions	
n=2	$12\ \ 34\ \ 56\ \ 78$	
	$00\ 01\ 10\ 11$	
	$00\ 01\ 10\ 11$	00 occurs twice, at positions different modulo 2
	$00\ 01\ 10\ 11$	
	$00\ 01\ 10\ 11$	01 occurs twice, at positions different modulo 2
	$00\ 01\ 10\ 11$	
	$00\ 01\ 10\ 11$	10 occurs twice, at positions different modulo 2
	$00\ 01\ 10\ 11$	
	$00\ 01\ 10\ 11$	11 occurs twice, at positions different modulo 2

Observation

Consider the concatenation in lexicographic order of all blocks of length n. Viewed circularly, each block of length n occurs exactly n times at positions different modulo n.

$n = 3$	000 001 010 011 100 101 110 111	000 occurs three times,
	000 001 010 011 100 101 110 111	at positions different modulo 3
	000 001 010 011 100 101 110 111	
	000 001 010 011 100 101 110 111	001 occurs three times
	000 001 010 011 100 101 110 111	at positions different modulo 3
	000 001 010 011 100 101 110 111	
	...	
⋮		

Neither Barbier (1887) nor Champernowne (1933) noticed this.

Observation

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

$$00\ 10\ 11\ 01$$

Perfect necklaces

Definition (Alvarez, Becher, Ferrari and Yuhjtman 2016)

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

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

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

Arithmetic progressions yield perfect necklaces

Identify the blocks of length n over the b-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.

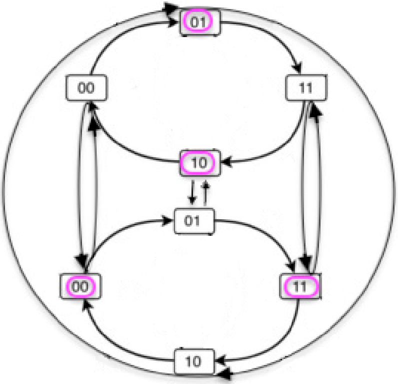
With $r = 1$ we have the lexicographically ordered (n, n)-perfect necklace.

Astute graphs

Fix b-alphabet A.
The astute graph $G(b, n, k)$ is directed, with kb^n vertices.
Vertices: $A^n \times \{0, \ldots, k-1\}$
Edges: $(w, m) \to (w', m')$ if

$$w(2..n) = w'(1..n-1) \text{ and } m' = (m+1) \mod k$$

This is $G(2, 2, 2)$



Astute graphs

Each Hamiltonian cycle in $G(b,n,k)$ gives one (n,k)-perfect necklace.

The line graph of $G(b,n-1,k)$ is $G(b,n,k)$.

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

Each (n,k)-perfect necklace possibly gives various Eulerian cycles in $G(b,n-1,k)$.

Normal sequences and perfect necklaces

Theorem (proved first by Ugalde 2000 for de Bruijn)

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

A seemingly easier criterion for normality

Theorem (Piatetski-Shapiro 1951)

Let x be an infinite sequence in a b-symbol alphabet.
If there is a constant C such that for all words w,

$$\limsup_{n\to\infty} \frac{\#\text{occurrences of } w \text{ in } x[1,n]}{n} < Cb^{-|w|}$$

then x is normal.

First result insertion

Theorem (Becher 2022)

Let x be a concatenation of $(n+1,n)$-perfect necklaces over b-alphabet, $n \geq 1$. There is computable sequence y normal to $(b+1)$-alphabet such that y is the concatenation of $(n+1,n)$-perfect necklaces over $(b+1)$-alphabet for $n \geq 1$, and x is a subsequence of y.

Moreover, for every integer N greater than b, in between the occurrences of the new symbol in y just before and just after position N there are at most $2b + \log_{b+1}(N)$ symbols.

Insertion in perfect necklaces

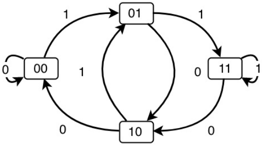
Theorem (follows from Becher and Cortés 2020)

For every $(n+1,n)$-perfect necklace v over a b-alphabet there is an $(n+1,n)$-perfect necklace w over $(b+1)$-alphabet such that v is a subsequence of w.

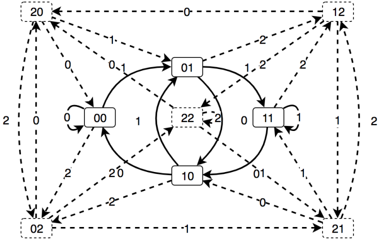
Moreover, for each such v there is w satisfying that for any $n+2b-1$ consecutive symbols there is at least one occurrence of the new symbol.

Proof of theorem on insertion in perfect necklaces

Given $(n+1,n)$-perfect necklace over b-alphabet,
pick an Eulerian cycle in $G(b,n,n)$ corresponding to it.

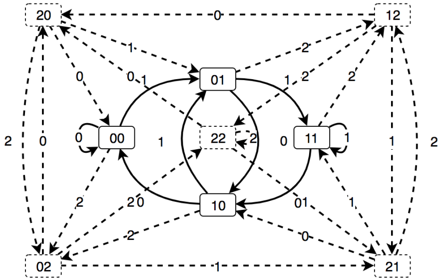


$G(b,n,n)$ is a subgraph of $G(b+1,n,n)$



Pictures here are $G(b,n,1)$ and $G(b+1,n,1)$ instead of $G(b,n,n)$ and $G(b+1,n,n)$

Consider $G(b+1,n,n) \setminus G(b,n,n)$ (the dotted lines)



Every Eulerian cycle is the union of disjoint cycles.

Without the small gap condition:
just use Euler-Hierholzer's algorithm for joining cycles

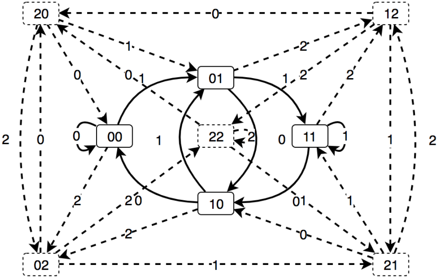
With the small gap condition is more delicate.

The resulting Eulerian cycle in $G(b+1,n,n)$ is an $(n+1,n)$-perfect necklace on $(b+1)$-alphabet with the wanted properties.

Augmenting graph

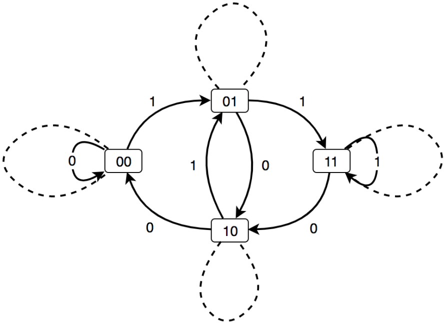
The augmenting graph $A(b+1,n,n)$ has exactly all the vertices of $G(b+1,n,n)$ and all the edges of $G(b+1,n,n) \setminus G(b,n,n)$.

Picture for $k=1$, de Bruijn case.



Petals

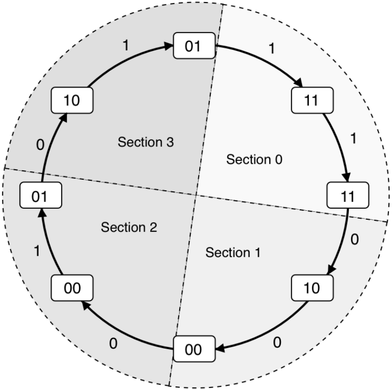
Partition the augmenting graph in nb^n disjoint cycles, called petals, such that in each of them there is one vertex in $G(b, n, n)$.



The small gap condition

Consider an $(n+1,n)$-perfect necklace and a starting position.

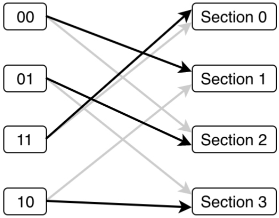
Pick corresponding Eulerian cycle in $G(b,n,n)$, with edges $e_1,\dots e_{nb^{n+1}}$.
Divide it in nb^n consecutive sections, each consisting of b edges.
Identify each section with the b target vertices in it.



A matching problem

The astute graph $G(b, n, n)$ has nb^n vertices.
An Eulerian cycle in $G(b, n, n)$ has nb^n sections with b vertices each.

We need to choose one vertex in each section, and they are all different.
We pose it as a matching problem.



By Hall's marriage theorem there is a perfect matching.
Find it computing the maximum flow in a network, Edmonds-Karp algorithm □

Any symbol insertion in constructed normal sequences

We have just showed:

Starting from some particular constructed normal sequences, we can insert symbos and obtain a normal sequence in the enlarged alphabet, with the small gap condition.

Second result on insertion

Theorem (Zylber 2017)

For every sequence x normal to b-alphabet there exists a sequence y normal to $(b+1)$-alphabet such that $retract(y) = x$, where $retract(\ .)$ removes the occurrences of the symbol $'b'$.

In general, the sequence y is not computable from x, but from an upper bound of the speeed of convergence to normality of x.

Proof idea

Given a sequence over b-alphabet, insert $'b'$ in the positions prescribed by the concatenation of lexicographically ordered (n,n)-perfect necklaces over $(b+1)$-alphabet

perfect	00	01	02	10	11	12	20	21	22
wildcards	$\star\star$	$\star\star$	$\star 2$	$\star\star$	$\star\star$	$\star 2$	$2\star$	$2\star$	22

Proof idea

Given a sequence over b-alphabet, insert $'b'$ in the positions prescribed by the concatenation of lexicographically ordered (n, n)-perfect necklaces over $(b + 1)$-alphabet

perfect	00	01	02	10	11	12	20	21	22
wildcards	$\star\star$	$\star\star$	$\star 2$	$\star\star$	$\star\star$	$\star 2$	$2\star$	$2\star$	22
arbitrary	$a_1 a_2$	$a_3 a_4$	$a_5 a_6$	$a_7 a_8$	$a_9 a_{10}$	$a_{11} a_{12}$			
insertion	$a_1 a_2$	$a_3 a_4$	$a_5 2$	$a_6 a_7$	$a_8 a_9$	$a_{10}\ 2$	$2 a_{11}$	$2 a_{12}$	22

Discrete discrepancy on aligned occurrences

$$\Delta_{b,\ell}(u) = \max_{\text{block } v \text{ of lenght } \ell} \left| \frac{\text{number of aligned occurrence of } v \text{ in } u}{\lfloor |u|/\ell \rfloor} - \frac{1}{b^\ell} \right|.$$

Fact

If p divides P and $\Delta_{b,P} < \epsilon$ then $\Delta_{b,p} < bound(b, \epsilon, p, P)$.

Proved by Pillai (1940):
A sequence $a_1 a_2 \ldots$ is normal to b-alphabet if for every length ℓ

$$\lim_{n \to \infty} \Delta_{b,\ell}(a_1 \ldots a_{\ell n}) = 0$$

In general, given $a_1 a_2 \ldots$ the function $f : \mathbb{Q} \to \mathbb{N}$ such that $f(\varepsilon) = n$ when n is the minimum such that $\Delta_{b,\ell}(a_1 ... a_n) < \varepsilon$ is not computable.

Insertion and discrete discrepancy

Let ℓ_n be the number of non-b symbols in (n, n)-perfect necklace on $(b+1)$-alphabet.

Let $e_n(u)$ be the sequence that results form inserting b's in u according to the ordered (n, n)-perfect necklace on $(b+1)$-alphabet.

Lemma (crucial)

For every n, there is c_n such that for every finite u over b-alphabet

if $\Delta_{b,\ell_n}(u) < \epsilon$

then $\Delta_{b+1,n}(e_n(u)) < c_n\epsilon$

Zylber's construction

Given sequence x normal to b-alphabet,

• Partition x in $u_1, u_2, u_3, \ldots$ where
$|u_n| = t_n \ell_{2^n}$ and t_n is large enough so that $\Delta_{b,\ell_{2^n}}(u_n)$ is small.

Recall ℓ_{2^n} is the number of non-b's in $(2^n, 2^n)$-perfect necklace on $(b+1)$-alphabet.

• By the crucial Lemma $\Delta_{b+1,2^n}(e_{2^n}(u_n))$ is small.

• Discrete discrepancy $\Delta_{b+1,P}$ also controls $\Delta_{b+1,p}$ when p divides P.

The wanted sequence is $y = e_{2^1}(u_1)e_{2^2}(u_2)e_{2^3}(u_3)\ldots$

By Piatetski-Shapiro theorem, y is normal to $(b+1)$-alphabet. □

Open problems

- What are other forms of insertion transfering normality from b-alphabet to $(b+1)$-alphabet?

- Discrepancy analysis for different forms of insertion

  Compare discrepancy $(b^n x \mod 1)_{n\geq 0}$ and $((b+1)^n y \mod 1)_{n\geq 0}$ where y results from insertion in x.

  Fukuyama and Hiroshima in 2012 gave metric discrepancy results for subsequences of $(b^n x \mod 1)_{n\geq 0}$,

- How can we change effective Hausdorff dimension while enlarging the alphabet?

References

- C. Aistleitner. "On modifying normal numbers". *Uniform Distribution Theory* 6 (2):2, 49–58, 2011.
- N. Alvarez, V. Becher, P. Ferrari and S. Yuhjtman. "Perfect necklaces", *Advances of Applied Mathematics* 80:48–61, 2016.
- V. Becher. "Insertion in constructed normal numbers", Uniform Distribution Theory 17(1): 55-76, 2022. .
- É. Barbier. "On suppose écrite la suite naturelle des nombres; quel est le $(10^{1000})^{\text{ième}}$ chiffre écrit?" *Comptes Rendus des Séances de l'Académie des Sciences Paris* 105:795–798, 1887.
- É. Barbier. "On suppose écrite la suite naturelle des nombres; quel est le $(10^{10000})^{\text{ième}}$ chiffre écrit?" *Comptes Rendus des Séances de l'Académie des Sciences Paris* 105:1238–1239, 1887.
- V. Becher and L. Cortés. "Extending de Bruijn sequences to larger alphabets", *Information Processing Letters*, Volume 168: 106085, 2021.
- D. Champernowne. "The Construction of Decimals Normal in the Scale of Ten." *Journal London Mathematical Society*, S1-8(4):254, 1933.
- K. Fukuyama and N. Hiroshima. "Metric discrepancy results for subsequences of $\{\theta^k x\}$", *Monatshefte für Mathematik*, 165: 199-215, 2012.
- N. Greenberg, J. S. Miller, A. Shen, and L. Brown Westrick. Dimension 1 sequences are close to randoms. *Theoretical Computer Science* 705:99–112, 2018.
- J. Vandehey. "Uncanny subsequence selections that generate normal numbers'. *Uniform Distribution Theory* 12(2), 65–75, 2017.
- D.D.Wall. "Normal numbers" *Ph.D.Thesis* University of California Berkeley, 1949
- A. Zylber. From Randomness in Two Symbols to Randomness in Three Symbols Uniform Distribution Theory 16:109-128, 2021.