On modifying normal numbers

Verónica Becher

Universidad de Buenos Aires

Uniform Distribution Theory - February 22, 2021

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.

Equivalently, a real number x is normal to base b if the fractional parts of $x, bx, b^2x, \ldots$ are uniformly distributed in the unit interval, Wall 1949.

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.

Equivalently, a real number x is normal to base b if the fractional parts of $x, bx, b^2x, \ldots$ are uniformly distributed in the unit interval, Wall 1949.

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

Modifying normal numbers

- Selection of subsequences
  Wall 1949, Agafonov 1968, Kamae and Weiss 1975
- Sums with rational number
  Volkonoff 1979, Aistleitner 2017
- Transformations by finite state tranducers
  Carton and Orduna 2020
- Insertion in positions in a set of density zero
  Becher and Figueira 2002, Aistleitner 2017
- Removal of one digit yields normality to **shrinked** alphabet
  Vandehey 2017

The question

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

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

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

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 with 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 with 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	
	...	
$\vdots$		

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)

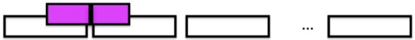
Let r coprime with b. The concatenation of blocks corresponding to the arithmetic sequence $0, r, 2r, ..., (b^n - 1)r$ yields an (n, n)-perfect necklace.

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

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* $\ell = 0, 1, \ldots n-1$ *for every block* x *of length* ℓ *and for every block* y *of length* $n - \ell$*, there is a unique block* w *of length* n *such that* $w(n-\ell \ldots n-1) = x$ *and* $(\sigma(w))(0 \ldots n-\ell-1) = y$*.*



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

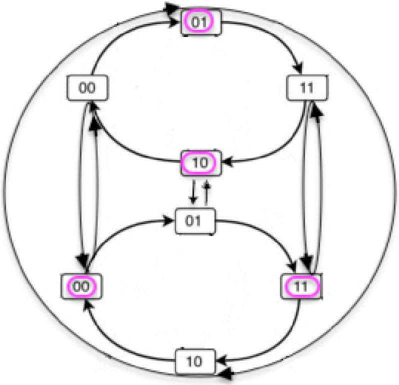
Astute graphs

Fix a b-alphabet. The astute graph $G(b, n, k)$ is directed, with kb^n vertices.

Vertices: pairs (w, m), w is length n block, m is integer between 0 and $k - 1$.

Edges: $(w, m) \to (w', m')$ if $w(2..n) = w'(1..n-1)$ and $(m+1) \bmod k = m'$.

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



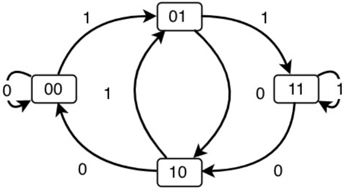
Astute graphs

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-alphabet.



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 be from various 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-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*,*
- $e(j) = (b!)^{j b^{n-1}} b^{-n}$ *is the number of Eulerian cycles in* $G(b, n-1, j)$
- φ *is Euler's totient function*

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.

The proof is a direct application of Piatetski-Shapiro (1951) theorem.

First result on insertion

Start with the concatenation of (n, k)-perfect necklaces over b-alphabet. Insert in each (n, k)-perfect necklace over b-alphabet to obtain (n, k)-perfect necklaces over $(b + 1)$-alphabet.

Insertion in perfect necklaces

Theorem (Becher and Cortés 2020)

For every (n,k)-perfect necklace v over a b-alphabet there is an (n,k)-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.

The small gap condition

Every consecutive $n + 2b - 1$ positions should have the new symbol.

For $b = 2$ and $n = 2$ every 5 positions should have the new symbol.

Insertion without the small gap condition

$v = \quad 00\ 01_{\blacktriangle} 10\ 11_{\blacktriangle}$

$w = \underbrace{00\ 01\ 0}_{5\ positions}2\ \ \underbrace{10\ 11\ 1}_{5\ positions}2\ 20\ 21\ 22$

The small gap condition

Every consecutive $n + 2b - 1$ positions should have the new symbol.

For $b = 2$ and $n = 2$ every 5 positions should have the new symbol.

Insertion without the small gap condition

$$v = \quad 00\ 01_{\blacktriangle}10\ 11_{\blacktriangle}$$

$$w = \underbrace{00\ 01\ 0}_{5\ positions}2\quad \underbrace{10\ 11\ 1}_{5\ positions}2\ 20\ 21\ 22$$

Insertion with the small gap condition

$$v = \quad 00_{\blacktriangle}01_{\blacktriangle}10\ 1_{\blacktriangle}1$$

$$w = \underbrace{00}_{4\ pos}\,2\,\underbrace{0\ 01}_{3\ pos}\ \ 22\ 2\underbrace{1\ 10\ 1}_{4\ pos}2\ \underset{\sim}{0}2\,\underbrace{11}_{4\ pos}$$

Proof of theorem on insertion in perfect necklaces

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

Proof of theorem on insertion in perfect necklaces

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

- $G(b,n,k)$ is a subgraph of $G(b+1,n,k)$
- The augmenting graph $G(b+1,n,k) \setminus G(b,n,k)$ is Eulerian
- Every Eulerian cycle is the union of disjoint cycles.

Proof of theorem on insertion in perfect necklaces

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

- $G(b,n,k)$ is a subgraph of $G(b+1,n,k)$
- The augmenting graph $G(b+1,n,k) \setminus G(b,n,k)$ is Eulerian
- Every Eulerian cycle is the union of disjoint cycles.
- **Without the small gap condition:**
  Use Euler-Hierholzer's algorithm for joining cycles

Proof of theorem on insertion in perfect necklaces

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

- $G(b,n,k)$ is a subgraph of $G(b+1,n,k)$
- The augmenting graph $G(b+1,n,k) \setminus G(b,n,k)$ is Eulerian
- Every Eulerian cycle is the union of disjoint cycles.
- **Without the small gap condition:**
  Use Euler-Hierholzer's algorithm for joining cycles

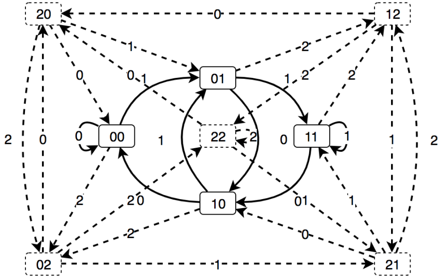
  With the small gap condition:
  Partition the augmenting graph in kb^n disjoint cycles, solve a matching problem and join cycles.

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

Augmenting graph

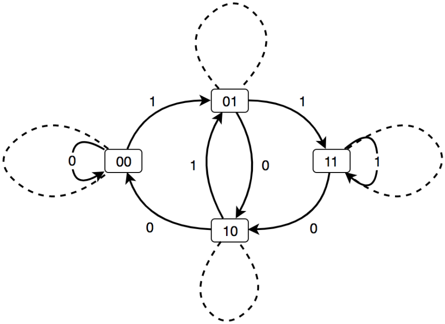
The augmenting graph $A(b+1, n, k)$ has exactly all the vertices of $G(b+1, n, k)$ and all the edges of $G(b+1, n, k) \setminus G(b, n, k)$.

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



Petals

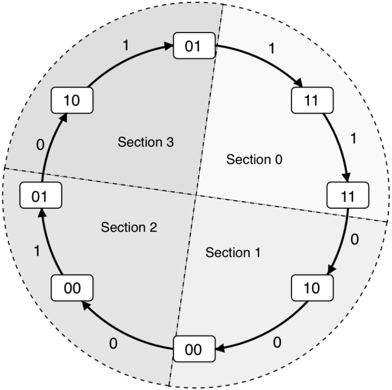
Partition the augmenting graph in kb^n disjoint cycles, called petals, such that in each of them there is one vertex in $G(b, n, k)$.



The small gap condition

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

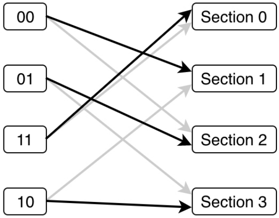
Pick corresponding Eulerian cycle in $G(b,n,k)$, with edges $e_1,\dots e_{kb^{n+1}}$.
Divide it in kb^n consecutive sections, each consisting of b edges.
Identify each section with the b target vertices in it.



A matching problem

The astute graph $G(b,n,k)$ has kb^n vertices.
An Eulerian cycle in $G(b,n,k)$ has kb^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 □

Second result on insertion

Theorem (Zylber 2017)

For every sequence x normal to a 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 $'b'$.

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, n non-decreasing

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, n non-decreasing

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_1a_2	a_3a_4	a_5a_6	a_7a_8	a_9a_{10}	$a_{11}a_{12}$			
insertion	a_1a_2	a_3a_4	$a_5 2$	a_6a_7	a_8a_9	$a_{10}\ 2$	$2a_{11}$	$2a_{12}$	22

We need to control the number of blocks before insertion and after insertion

Discrete discrepancy on aligned occurrences

We count aligned occurrences of blocks (occurrences at positions that are multiple of the block length, hence non-overlapping).

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

As proved by Pillai (1940), a sequence $a_1 a_2 \ldots$ is normal to b-alphabet if for every length ℓ

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

Insertion and discrete discrepancy

Lemma (crucial)

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

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

where

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

$e_n(u)$ *inserts b's in u according to ordered (n, n)-perfect necklace on $(b+1)$-alphabet*

Construction

It uses $(2^n, 2^n)$-perfect necklaces in the $(b+1)$-alphabet.
Given sequence x normal to b-alphabet,

- Determine $t_1, t_2, t_3 \ldots$ to partition x in

$$u_1, u_2, u_3, \ldots$$

so that for every n, $\Delta_{b,\ell_{2n}}(u_n)$ is small and $|u_n| = t_n \ell_{2^n}$, where ℓ_{2^n} is number of non-b's in $(2^n, 2^n)$-perfect necklace on $(b+1)$-alphabet

Construction

It uses $(2^n, 2^n)$-perfect necklaces in the $(b+1)$-alphabet.
Given sequence x normal to b-alphabet,

- Determine $t_1, t_2, t_3 \ldots$ to partition x in

$$u_1, u_2, u_3, \ldots$$

  so that for every n, $\Delta_{b,\ell_{2^n}}(u_n)$ is small and $|u_n| = t_n \ell_{2^n}$, where ℓ_{2^n} is number of non-b's in $(2^n, 2^n)$-perfect necklace on $(b+1)$-alphabet

- By the crucial Lemma

$$\Delta_{b+1,2^n}(e_{2^n}(u_n)) \text{ is small}$$

Construction

It uses $(2^n, 2^n)$-perfect necklaces in the $(b+1)$-alphabet.
Given sequence x normal to b-alphabet,

- Determine $t_1, t_2, t_3 \ldots$ to partition x in

$$u_1, u_2, u_3, \ldots$$

  so that for every n, $\Delta_{b,\ell_{2^n}}(u_n)$ is small and $|u_n| = t_n \ell_{2^n}$, where ℓ_{2^n} is number of non-b's in $(2^n, 2^n)$-perfect necklace on $(b+1)$-alphabet

- By the crucial Lemma

$$\Delta_{b+1,2^n}(e_{2^n}(u_n)) \text{ is small}$$

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

Construction

It uses $(2^n, 2^n)$-perfect necklaces in the $(b+1)$-alphabet.
Given sequence x normal to b-alphabet,

- Determine $t_1, t_2, t_3 \ldots$ to partition x in

$$u_1, u_2, u_3, \ldots$$

  so that for every n, $\Delta_{b,\ell_{2^n}}(u_n)$ is small and $|u_n| = t_n \ell_{2^n}$, where ℓ_{2^n} is number of non-b's in $(2^n, 2^n)$-perfect necklace on $(b+1)$-alphabet
- By the crucial Lemma

$$\Delta_{b+1,2^n}(e_{2^n}(u_n)) \text{ is small}$$

- Discrete discrepancy $\Delta_{b+1,P}$ also controls $\Delta_{b+1,p}$ when p divides P.
- The sequence $y = e_{2^1}(u_1)e_{2^2}(u_2)e_{2^3}(u_3)\ldots$ is normal to $(b+1)$-alphabet:
  In each $e(u)$ we controlled the number of aligned occurrences of blocks.
  In the concatenation we bound the number of non-aligned occurrences.
  We conclude applying Piatetski-Shapiro (1951) theorem. □

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 $(b^n x \mod 1)_{n\geq 0}$,

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.
- 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.
- 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 with two symbols to randomness with three symbols", *Tesis de Licenciatura en Ciencias de la Computación*, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires,November 2017.