Normal numbers and perfect necklaces

Verónica Becher

Universidad de Buenos Aires & CONICET

70th birthday of Prof. Victor Selivanov
Fourth Workshop on Digitalization and Computable Models (WDCM-2022)
Sobolev Institute of Mathematics (Novosibirsk, Russia) and Kazan Federal University (Russia)

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.

Couterxamples:
$0.010010001000001\ldots$

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.

Couterxamples:
$0.010010001000001\ldots$
$0.0101010101010101\ldots$

In 1909 Borel gave this definition, proved that almost all real numbers are nomal to all integer bases, and he asked for an example.

Borel's question

All Martin-Löf random reals are normal to every base, in particular Ω-numbers.

Constructions

Lebesgue 1909, Sierpinski 1916, Champernowne 1933, Turing 1937, Copeland and Erdős 1946, Davenport and Erdős 1952, W. Schmidt 1960, M. B. Levin 1970, Stoneham 1973, M. B. Levin 1999, ...

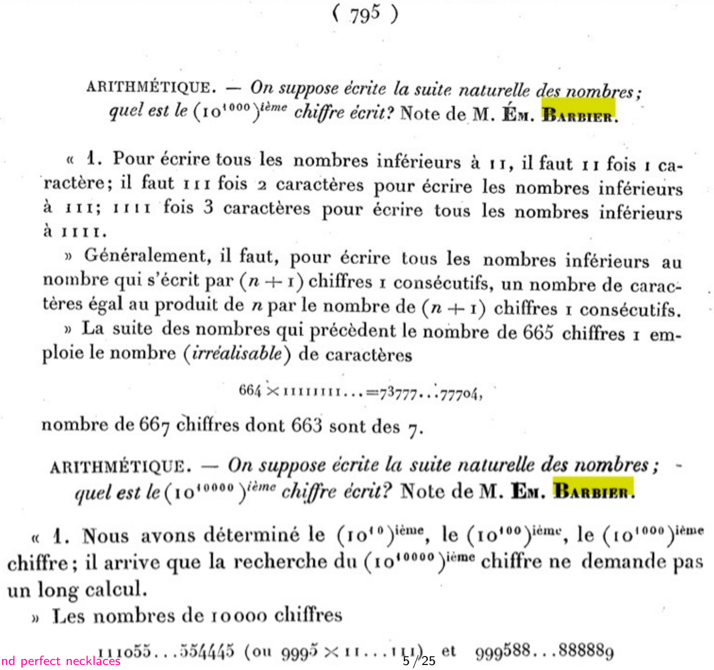
Borel's question is essentially open.

Champernowne's example

Theorem (Champernowne 1933)

$0.1234567891011121314151617181920212223\ldots$ *is normal to base* 10*.*

Comptes Rendus des Séances de l'Académie des Sciences Paris (105) pp 795--798 and 1238--1239, 1887.



(795)

ARITHMÉTIQUE. — *On suppose écrite la suite naturelle des nombres; quel est le* $(10^{1000})^{\text{ième}}$ *chiffre écrit?* Note de M. **Ém. Barbier**.

« 1. Pour écrire tous les nombres inférieurs à 11, il faut 11 fois 1 caractère; il faut 111 fois 2 caractères pour écrire les nombres inférieurs à 111; 1111 fois 3 caractères pour écrire tous les nombres inférieurs à 1111.

» Généralement, il faut, pour écrire tous les nombres inférieurs au nombre qui s'écrit par $(n+1)$ chiffres 1 consécutifs, un nombre de caractères égal au produit de n par le nombre de $(n+1)$ chiffres 1 consécutifs.

» La suite des nombres qui précèdent le nombre de 665 chiffres 1 emploie le nombre (*irréalisable*) de caractères

$$664 \times 1111111\ldots = 73777\ldots77704,$$

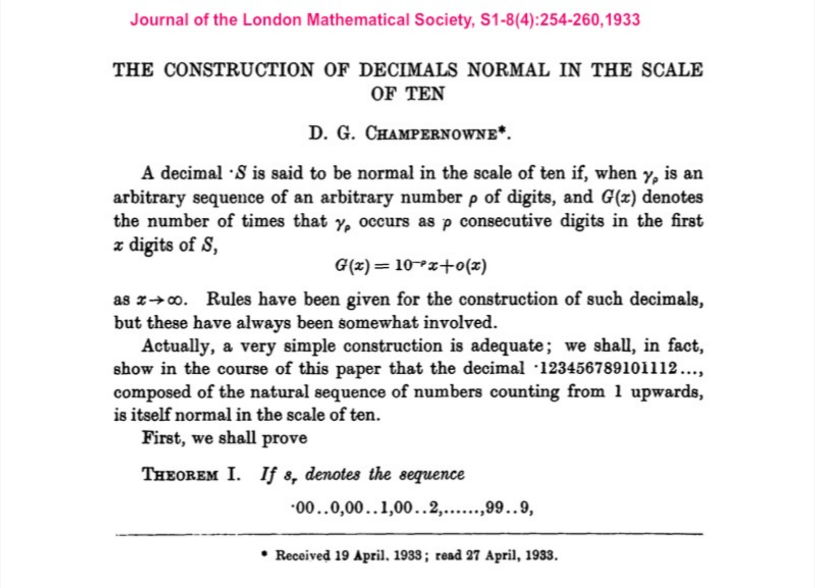
nombre de 667 chiffres dont 663 sont des 7.

ARITHMÉTIQUE. — *On suppose écrite la suite naturelle des nombres; quel est le* $(10^{10000})^{\text{ième}}$ *chiffre écrit?* Note de M. **Ém. Barbier**.

« 1. Nous avons déterminé le $(10^{10})^{\text{ième}}$, le $(10^{100})^{\text{ième}}$, le $(10^{1000})^{\text{ième}}$ chiffre; il arrive que la recherche du $(10^{10000})^{\text{ième}}$ chiffre ne demande pas un long calcul.

» Les nombres de 10000 chiffres

111055...554445 (ou 9995 × 11...111) et 999588...888889



Journal of the London Mathematical Society, S1-8(4):254-260,1933

THE CONSTRUCTION OF DECIMALS NORMAL IN THE SCALE OF TEN

D. G. CHAMPERNOWNE*.

A decimal $\cdot S$ is said to be normal in the scale of ten if, when γ_ρ is an arbitrary sequence of an arbitrary number ρ of digits, and $G(x)$ denotes the number of times that γ_ρ occurs as ρ consecutive digits in the first x digits of S,

$$G(x) = 10^{-\rho}x + o(x)$$

as $x \to \infty$. Rules have been given for the construction of such decimals, but these have always been somewhat involved.

Actually, a very simple construction is adequate; we shall, in fact, show in the course of this paper that the decimal $\cdot 123456789101112\ldots$, composed of the natural sequence of numbers counting from 1 upwards, is itself normal in the scale of ten.

First, we shall prove

THEOREM I. *If s_r denotes the sequence*

$$\cdot 00..0, 00..1, 00..2, ..., 99..9,$$

---

* Received 19 April, 1933; read 27 April, 1933.

The property of normality

A number $x = 0.a_1a_2, \ldots$ is normal to base b if

for every word w,

$$\lim_{n\to\infty} \frac{\text{the number occurrences of } w \text{ in } a_1, ..., a_n}{n} = b^{-|w|}.$$

Thus, we must count occurrences of word w in $a_1, ..a_n$.

Champernowne's proof

Instead, the concatenation of all blocks of n symbols in lexicographic order,

$$\underbrace{0123456789}_{megablock\ 1}\ \ \underbrace{00\ 01 \ldots 98\ 99}_{megablock\ 2}\ \ \underbrace{000\ 001\ \ \ldots\ \ 998\ 999}_{megablock\ 3} \ldots$$

Champernowne's proof counts:

- each digit
- each block of two digits

  . . .
- each block of n digits

The difficult part:

- to count occurrences in between blocks

  $000\ 001\ 002\ 003 \ldots 990\ 991\ 992\ 993\ 994\ 995\ 996\ 997\ 998\ 999$

  Inside a megablock for length n Champernowne just counts inside blocks and bounds the number of occurrences in between blocks.
- to count up to an arbitrary position within a megablock. $\square$

Our observation

For simplicity consider the alphabet $\{0, 1\}$.

In the megablock n viewed circularly, each block of length n occurs exactly n times at different positions modulo n.

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

Our observation

For simplicity consider the alphabet $\{0,1\}$.

In the megablock n viewed circularly, each block of length n occurs exactly n times at different positions modulo n.

position	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

Our observation

For simplicity consider the alphabet $\{0, 1\}$.

In the megablock n viewed circularly, each block of length n occurs exactly n times at different positions modulo n.

position $12\ \ 34\ \ 56\ \ 78$
$00\ 01\ 10\ 11$
$00\ 01\ 10\ 11$ 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

Our observation

For simplicity consider the alphabet $\{0, 1\}$.

In the megablock n viewed circularly, each block of length n occurs exactly n times at different positions modulo n.

position	$12\ 34\ 56\ 78$	
	$00\ 01\ 10\ 11$	
	$00\ 01\ 10\ 11$	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

Our observation

For simplicity consider the alphabet $\{0, 1\}$.

In the megablock n viewed circularly, each block of length n occurs exactly n times at different positions modulo n.

position	$12\ \ 34\ \ 56\ \ 78$	
	$00\ 01\ 10\ 11$	
	$00\ 01\ 10\ 11$	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

Neither Barbier nor Champernowne noticed this!

In the megablock n viewed circularly, each block of length n occurs exactly n times at different positions modulo n.

000 001 010 011 100 101 110 111 000 occurs three times,
000 001 010 011 100 101 110 111 at different positions 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 different postions modulo 3
000 001 010 011 100 101 110 111

. . .

However, 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-symbol alphabet is (n, k)-perfect if each block of length n occurs k times, at different position modulo k for any convention of the starting point.

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

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

Megablocks are 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, this is the magablock for length n.

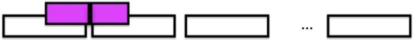
Megablocks are perfect necklaces

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, this is the magablock for length n.

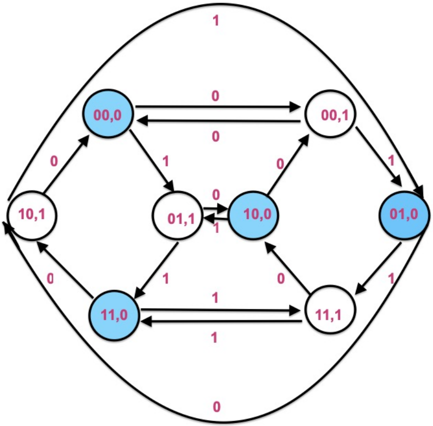


For every length-n block splitted in two parts, there is exactly one matching in the necklace (a tail of a block and the head of next block).

Astute graphs

Fix b-symbol alphabet. The astute graph $G_{b,n,k}$ is directed, with kb^n vertices.
The set if vertices is $\{0,..b-1\}^n \times \{0,..,.k-1\}$.
An edge $(w,m) \to (w',m')$ if $w(2,,n) = w'(1..n-1)$ and $(m+1) \mod k = m'$

This is $G_{2,2,2}$



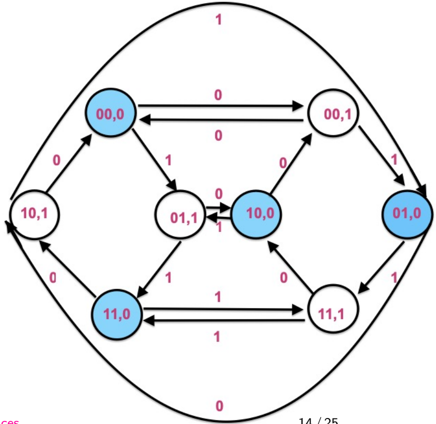
Astute graphs

Observation

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

Observation

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



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 (proved first by Ugalde 2000 for de Bruijn)

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

Proof of Theorem

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.

Proof of Theorem

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.

Instead, an equivalent formulation of normality but simpler to test:

Lemma (Piatetski-Shapiro 1951)

A sequence $a_1a_2a_3\ldots$ is normal to a b-symbol alphabet if and only if there is positive constant C such that for every word w,

$$\limsup_{n\to\infty} \frac{\textit{number of occurrences of } w \textit{ in } a_1\ldots a_n}{n} < C\, b^{-|w|}.$$

To prove that the sequence of megablocks is normal the count at an anbritrary position is bounded by considering the count at the end of the megablock.

□

The concatenation of (n, n)-perfect necklaces, n linearly increasing, is normal. Applying the same modification that Champernowne did we also obtain his result.

Corollary

Champernowne's sequence $0.12345678910112\ldots$ *is normal to base* 10.

Speed of convergence to normality

- A real x is normal to base b if the fractional parts of $x, bx, b^2x, \ldots$, that is $(b^n x \mod 1)_{n\geq 0}$, is uniformly distributed in the unit interval, Wall 1949.

Speed of convergence to normality

- A real x is normal to base b if the fractional parts of $x, bx, b^2x, \ldots$, that is $(b^n x \mod 1)_{n \geq 0}$, is uniformly distributed in the unit interval, Wall 1949.

- A sequence $(x_n)_{n \geq 1}$ is uniformly distributed in the unit interval if $D_N((x_n)_{n \geq 1}) = \sup_{[\alpha,\beta)} \left| \frac{\#\{n \leq N : x_n \in [\alpha,\beta)\}}{N} - (\beta - \alpha) \right|$ goes to 0 as N to ∞.

Speed of convergence to normality

- A real x is normal to base b if the fractional parts of $x, bx, b^2x, \ldots$, that is $(b^n x \mod 1)_{n\geq 0}$, is uniformly distributed in the unit interval, Wall 1949.

- A sequence $(x_n)_{n\geq 1}$ is uniformly distributed in the unit interval if $D_N((x_n)_{n\geq 1}) = \sup_{[\alpha,\beta)} \left| \frac{\#\{n \leq N : x_n \in [\alpha,\beta)\}}{N} - (\beta - \alpha) \right|$ goes to 0 as N to ∞.

- Schmidt 1972 proved that there is constant C such that for every $(x_n)_{n\geq 1}$ there are infinite Ns, $D_N((x_n)_{n\geq 1}) > C\frac{\log N}{N}$.
  This is optimal Van de Corput sequence has exactly this discrepancy.

Speed of convergence to normality

- A real x is normal to base b if the fractional parts of $x, bx, b^2x, \ldots$, that is $(b^n x \mod 1)_{n\geq 0}$, is uniformly distributed in the unit interval, Wall 1949.

- A sequence $(x_n)_{n\geq 1}$ is uniformly distributed in the unit interval if $D_N((x_n)_{n\geq 1}) = \sup_{[\alpha,\beta)} \left| \frac{\#\{n \leq N : x_n \in [\alpha,\beta)\}}{N} - (\beta-\alpha) \right|$ goes to 0 as N to ∞.

- Schmidt 1972 proved that there is constant C such that for every $(x_n)_{n\geq 1}$ there are infinite Ns, $D_N((x_n)_{n\geq 1}) > C\frac{\log N}{N}$.
  This is optimal Van de Corput sequence has exactly this discrepancy.

- It is still unknown whether the optimal order of discrepancy can be achieved by $(b^n x \mod 1)_{n\geq 0}$ for some real x, Korobov 1956.

Speed of convergence to normality

- A real x is normal to base b if the fractional parts of $x, bx, b^2x, \ldots$, that is $(b^n x \mod 1)_{n\geq 0}$, is uniformly distributed in the unit interval, Wall 1949.

- A sequence $(x_n)_{n\geq 1}$ is uniformly distributed in the unit interval if $D_N((x_n)_{n\geq 1}) = \sup_{[\alpha,\beta)} \left| \frac{\#\{n \leq N : x_n \in [\alpha,\beta)\}}{N} - (\beta - \alpha) \right|$ goes to 0 as N to ∞.

- Schmidt 1972 proved that there is constant C such that for every $(x_n)_{n\geq 1}$ there are infinite Ns, $D_N((x_n)_{n\geq 1}) > C\frac{\log N}{N}$.
  This is optimal Van de Corput sequence has exactly this discrepancy.

- It is still unknown whether the optimal order of discrepancy can be achieved by $(b^n x \mod 1)_{n\geq 0}$ for some real x, Korobov 1956.

- The lowest discrepancy known for $(b^n x \mod 1)_{n\geq 0}$ is exactly $C(\log N)^2/N$ for a real x constructed by M. Levin 1999 using the Pascal triangle matrix modulo 2. The lower bound by Hofer and Larcher 2022.

Nested perfect necklaces

Definition (Becher and Carton 2019)

A sequence over a b-symbol alphabet is a nested (n, k)-perfect necklace if it is (n, k)-perfect and, in case $n > 1$, it is the concatenation of b nested $(n-1, k)$-perfect necklaces.

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

$$0011\ 0110$$

Nested perfect necklaces

Definition (Becher and Carton 2019)

A sequence over a b-symbol alphabet is a nested (n,k)-perfect necklace if it is (n,k)-perfect and, in case $n>1$, it is the concatenation of b nested $(n-1,k)$-perfect necklaces.

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

$$0011\ 0110$$

The lexicographic order yields a perfect necklace but not nested,

$$\underbrace{0001}_{\text{not perfect}}\ \underbrace{1011}_{\text{not perfect}}$$

Nested perfect necklaces

Definition (Becher and Carton 2019)

A sequence over a b-symbol alphabet is a nested (n, k)-perfect necklace if it is (n, k)-perfect and, in case $n > 1$, it is the concatenation of b nested $(n-1, k)$-perfect necklaces.

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

$$0011\ 0110$$

The lexicographic order yields a perfect necklace but not nested,

$$\underbrace{0001}_{\text{not perfect}}\ \underbrace{1011}_{\text{not perfect}}$$

Nested (n, k)-perfect necklaces are pointed, which means an initial position

Nested perfect necklaces

These are nested $(2,4)$-perfect necklaces:

$$00001111\ 01011010$$

$$00111100\ 01101001$$

$$00011110\ 01001011$$

$$00101101\ 01111000$$

The concatenation of the first two is a nested $(3,4)$-perfect necklace.
The concatenation of the last two is a nested $(3,4)$-perfect necklace.
The concatenation of all of them is a nested $(4,4)$-perfect necklace.

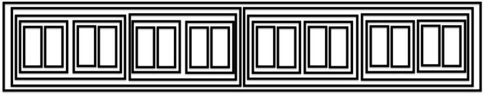
Nested perfect necklaces

Observation

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

• Since x is (n, n)-perfect, each block of length n occurs n times in x, at different positions modulo n.

• Since x is nested, for every $i = 1, \ldots n$, x is the concatenation of b^{n-i} nested (i, n)-perfect necklaces. So, in the prefix of x of length cnb^i each block of length i occurs $c \pm \epsilon$ times at positions with the same congruence modulo n, for ε equal to 0 or 1.



Levin's constant and nested perfect necklaces

Theorem (Becher and Carton 2019)

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

Theorem (Becher and Carton 2019)

For $n = 2^0, 2^1, 2^2, \ldots$. there are 2^{n-1} binary nested (n,n)-perfect necklaces obtained by column rotations of the Pascal triangle matrix modulo 2.

Nested perfect necklaces and low discrepancy

Theorem (Becher and Carton 2019)

Let b be a prime number. Every number x whose base-b expansion is the concatenation of nested (n, n)-perfect necklaces for $n = 2^0, 2^1, 2^2, \ldots.$ satifies $D_N((b^k x \mod 1)_{k\geq 0})$ is $O((\log N)^2/N)$.

Open problems

- Give a graph interpretation to nested perfect necklaces
- Study perfect necklaces in higher dimensions
- Is there a Martin-Löf random real x such that for every N, $D_N((2^n x \mod 1)_{n\geq 1})$ is $O((\log N)^2/N)$?

References

- 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 O. Carton. “Normal numbers and nested perfect necklaces”, Journal of Complexity 54:101403, 2019
- D. Champernowne. “The Construction of Decimals Normal in the Scale of Ten.” Journal London Mathematical Society, S1-8(4):254-260, 1933.
- M. B. Levin. On the discrepancy estimate of normal numbers. Acta Arithmetica 88(2), 99–111, 1999.
- R. Hofer and G.Larcher. The exact order of discrepancy for Levin’s normal number in base 2. arXiv:2205.01566, 2022.
- N. Korobov. “On completely uniform distributions and jointly normal numbers”. *Izv. AN SSSR, ser. matem.*, 20, 1956.
- M. B. Levin. “On the discrepancy estimate of normal numbers”. *Acta Arithmetica* 88(2):99–111, 1999.
- W. Schmidt. “ Irregularities of distribution VII”. *Acta Arithmetica*, 21:45–50, 1972.
- D.D.Wall. “Normal numbers”, Ph.D.Thesis University of California Berkeley, 1949