Subsequences of digitally defined functions

Lukas Spiegelhofer

March 25, 2022, TU Wien

Outline

Section 0

Digital expansions

In the simplest case, a *digital expansion* Φ assigns to each natural number a finite string of *digits*. This usually happens in a monotone way —

$$\text{if } n \le m, \text{ then } \Phi(n) \le_{\text{lex}} \Phi(m).$$

n	$\Phi(n)$	n	$\Phi(n)$
0	0	10	1010
1	1	11	1011
2	10	12	1100
3	11	13	1101
4	100	14	1110
5	101	15	1111
6	110	16	10000
7	111	17	10001
8	1000	18	10010
9	1001	19	10011

This is the *binary* expansion $[n]_2$ of a nonnegative integer n.

The sum-of-digits function

The *sum-of-digits* function s_q in base q simply sums all the digits in base q.

n	$[n]_2$	$\mathrm{s}_2(n)$	n	$[n]_2$	$\mathrm{s}_2(n)$
0	0	0	10	1010	2
1	1	1	11	1011	3
2	10	1	12	1100	2
3	11	2	13	1101	3
4	100	1	14	1110	3
5	101	2	15	1111	4
6	110	2	16	10000	1
7	111	3	17	10001	2
8	1000	1	18	10010	2
9	1001	2	19	10011	3

Legendre's formula

The base-p sum-of-digits function, p prime, appears in the prime factor decomposition of $n!$ by Legendre's formula:

$$(p-1)\nu_p(n!) = n - \mathsf{s}_p(n).$$

This links combinatorics to number theory. For me, this link is the strongest motivation for studying sum-of-digits functions.

Legendre's formula

The base-p sum-of-digits function, p prime, appears in the prime factor decomposition of $n!$ by Legendre's formula:

$$(p-1)\nu_p(n!) = n - \mathsf{s}_p(n).$$

This links combinatorics to number theory. For me, this link is the strongest motivation for studying sum-of-digits functions.

The Zeckendorf expansion

Every nonnegative integer n is the sum of different, non-consecutive Fibonacci numbers F_i, $i \geq 2$, and such a representation is unique $\rightsquigarrow$ Zeckendorf expansion.

0	0	8	10000	16	100100
1	1	9	10001	17	100101
2	10	10	10010	18	101000
3	100	11	10100	19	101001
4	101	12	10101	20	101010
5	1000	13	100000	21	1000000
6	1001	14	100001	22	1000001
7	1010	15	100010	23	1000010

- The number of 1s needed is the *Zeckendorf sum of digits* $\mathrm{z}(n)$ of n.

The Zeckendorf expansion

Every nonnegative integer n is the sum of different, non-consecutive Fibonacci numbers F_i, $i \geq 2$, and such a representation is unique $\rightsquigarrow$ Zeckendorf expansion.

0	0	0	8	10000	1	16	100100	2
1	1	1	9	10001	2	17	100101	3
2	10	1	10	10010	2	18	101000	2
3	100	1	11	10100	2	19	101001	3
4	101	2	12	10101	3	20	101010	3
5	1000	1	13	100000	1	21	1000000	1
6	1001	2	14	100001	2	22	1000001	2
7	1010	2	15	100010	2	23	1000010	2

- The number of 1s needed is the *Zeckendorf sum of digits* $\mathrm{z}(n)$ of n.

Section 1

Sparse arithmetic subsequences of sum-of-digits functions

The Thue–Morse sequence

The parity of the number of ones in the binary expansion yields the *Thue–Morse sequence*

$$\mathtt{T} = (\mathsf{s}_2(n) \bmod 2)_{n\geq 0} = 01101001100101101001011001101001\cdots$$

The sequence $\mathtt{T}$ is an *automatic sequence* and as such can be defined via a *uniform morphism on a finite alphabet*: Let us define

$$\varphi : 0 \mapsto 01, \quad 1 \mapsto 10.$$

Starting with 0, we obtain

$$0 \mapsto 01 \mapsto 0110 \mapsto 01101001\cdots$$

The Thue–Morse sequence

The parity of the number of ones in the binary expansion yields the *Thue–Morse sequence*

$$\mathtt{T} = (\mathsf{s}_2(n) \bmod 2)_{n\geq 0} = \mathtt{01101001100101101001011001101001}\cdots$$

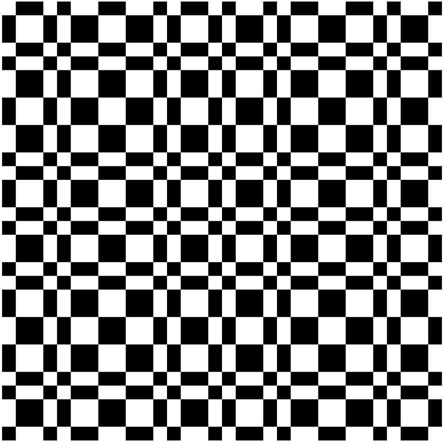
The sequence $\mathtt{T}$ is an *automatic sequence* and as such can be defined via a *uniform morphism on a finite alphabet*: Let us define

$$\varphi : \mathtt{0} \mapsto \mathtt{01}, \quad \mathtt{1} \mapsto \mathtt{10}.$$

Starting with $\mathtt{0}$, we obtain

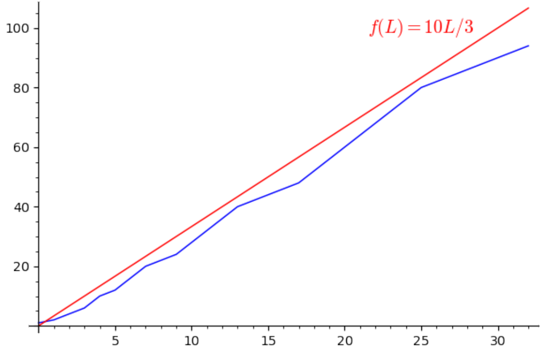
$$\mathtt{0} \mapsto \mathtt{01} \mapsto \mathtt{0110} \mapsto \mathtt{01101001}\cdots$$

In the Thue–Morse sequence, each symbol 0, 1 appears with asymptotic frequency $1/2$. It is built from the two blocks 01 and 10!



The factor complexity of $\mathtt{T}$

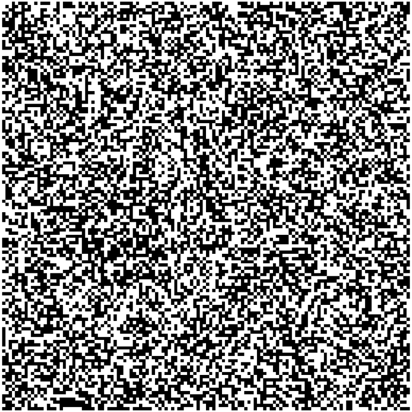
There are only very few words over $\{0,1\}$ appearing as *factors* (contiguous finite subsequences) of $\mathtt{T}$: the number of factors of length L appearing in $\mathtt{T}$ is bounded by CL with an absolute constant C.



L	$p(L)$	2^L
0	1	1
1	2	2
2	4	4
3	6	8
4	10	16
5	12	32
6	16	64
7	20	128
8	22	256
9	24	512
10	28	1024

Sparse arithmetic subsequences of T

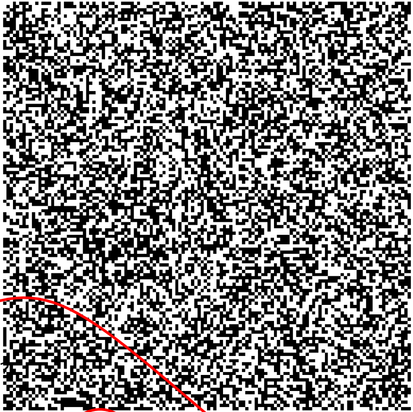
This situation changes completely when we consider *arithmetic subsequences* of T instead.



$N = 128 \times 128$ terms, common difference $N^R = 3^{21}$

Sparse arithmetic subsequences of $\mathtt{T}$

This situation changes completely when we consider *arithmetic subsequences* of $\mathtt{T}$ instead.



does every pattern occur?

$N = 128 \times 128$ terms, common difference $N^R = 3^{21}$

Does every pattern occur?

Avgustinovich, Fon-Der-Flaass, and Frid (2003): *every* finite word $\omega \in \{0, 1\}^L$ appears as an arithmetic subsequence of $\mathtt{T}$.

Müllner–Spiegelhofer (Israel J. Math. 2017):
Let $\omega \in \{0, 1\}^L$ and $0 < \varepsilon < 2$. As $N \to \infty$, the following holds.

For most $d \in [N^{2-\varepsilon}, 2N^{2-\varepsilon})$, the number of times that ω appears as a subword of $(\mathtt{T}(nd + a))_{n<N}$ is close to the expected value $N/2^L$.

We used this in order to construct certain *normal sequences*: For $1 < c < 3/2$, the sequence $n \mapsto \mathtt{T}\big(\lfloor n^c \rfloor\big)$ is normal. Every block $\omega \in \{0, 1\}^L$ appears with asymptotic frequency 2^{-L} in this sequence.

Before that, Drmota, Mauduit, and Rivat (JEMS) proved that $n \mapsto \mathtt{T}(n^2)$ is a normal sequence.

Does every pattern occur?

Avgustinovich, Fon-Der-Flaass, and Frid (2003): *every* finite word $\omega \in \{0,1\}^L$ appears as an arithmetic subsequence of $\mathtt{T}$.

Müllner–Spiegelhofer (Israel J. Math. 2017):
Let $\omega \in \{0,1\}^L$ and $0 < \varepsilon < 2$. As $N \to \infty$, the following holds.

For most $d \in [N^{2-\varepsilon}, 2N^{2-\varepsilon})$, the number of times that ω appears as a subword of $(\mathtt{T}(nd+a))_{n<N}$ is close to the expected value $N/2^L$.

We used this in order to construct certain *normal sequences*: For $1 < c < 3/2$, the sequence $n \mapsto \mathtt{T}(\lfloor n^c \rfloor)$ is normal. Every block $\omega \in \{0,1\}^L$ appears with asymptotic frequency 2^{-L} in this sequence.

Before that, Drmota, Mauduit, and Rivat (JEMS) proved that $n \mapsto \mathtt{T}(n^2)$ is a normal sequence.

Does every pattern occur?

Avgustinovich, Fon-Der-Flaass, and Frid (2003): *every* finite word $\omega \in \{0,1\}^L$ appears as an arithmetic subsequence of $\mathtt{T}$.

Müllner–Spiegelhofer (Israel J. Math. 2017):
Let $\omega \in \{0,1\}^L$ and $0 < \varepsilon < 2$. As $N \to \infty$, the following holds.

For most $d \in [N^{2-\varepsilon}, 2N^{2-\varepsilon})$, the number of times that ω appears as a subword of $(\mathtt{T}(nd+a))_{n<N}$ is close to the expected value $N/2^L$.

We used this in order to construct certain *normal sequences*: For $1 < c < 3/2$, the sequence $n \mapsto \mathtt{T}\big(\lfloor n^c \rfloor\big)$ is normal. Every block $\omega \in \{0,1\}^L$ appears with asymptotic frequency 2^{-L} in this sequence.

Before that, Drmota, Mauduit, and Rivat (JEMS) proved that $n \mapsto \mathtt{T}(n^2)$ is a normal sequence.

Does every pattern occur?

Avgustinovich, Fon-Der-Flaass, and Frid (2003): *every* finite word $\omega \in \{0,1\}^L$ appears as an arithmetic subsequence of $\mathtt{T}$.

Müllner–Spiegelhofer (Israel J. Math. 2017):
Let $\omega \in \{0,1\}^L$ and $0 < \varepsilon < 2$. As $N \to \infty$, the following holds.

For most $d \in [N^{2-\varepsilon}, 2N^{2-\varepsilon})$, the number of times that ω appears as a subword of $(\mathtt{T}(nd+a))_{n<N}$ is close to the expected value $N/2^L$.

We used this in order to construct certain *normal sequences*: For $1 < c < 3/2$, the sequence $n \mapsto \mathtt{T}\bigl(\lfloor n^c \rfloor\bigr)$ is normal. Every block $\omega \in \{0,1\}^L$ appears with asymptotic frequency 2^{-L} in this sequence.

Before that, Drmota, Mauduit, and Rivat (JEMS) proved that $n \mapsto \mathtt{T}(n^2)$ is a normal sequence.

Very sparse arithmetic subsequences of $\mathtt{T}$

Theorem (S. 2020, Compos. Math.)

The Thue–Morse sequence has level of distribution 1*. More precisely, for all* $\varepsilon > 0$ *we have*

$$\sum_{1\le d\le D} \max_{\substack{y,z\ge 0\\ z-y\le x}} \max_{0\le a<d} \left| \sum_{\substack{y\le n<z\\ n\equiv a \bmod d}} (-1)^{\mathsf{s}_2(n)} \right| \le Cx^{1-\eta}$$

for some C *and* $\eta > 0$ *depending on* ε*, where* $D = x^{1-\varepsilon}$*.*

In more relaxed language: let $\rho > 0$. As $N \to \infty$, the following holds.

For most $d \in [N^\rho, 2N^\rho)$*, the number of times that* 0 *appears in*

$$\big(\mathtt{T}(nd + a)\big)_{0\le n<N}$$

is close to $N/2$*.*

Very sparse arithmetic subsequences of $\mathtt{T}$

Theorem (S. 2020, Compos. Math.)

The Thue–Morse sequence has level of distribution 1*. More precisely, for all* $\varepsilon > 0$ *we have*

$$\sum_{1\le d\le D} \max_{\substack{y,z\ge 0\\ z-y\le x}} \max_{0\le a<d} \left| \sum_{\substack{y\le n<z\\ n\equiv a \bmod d}} (-1)^{\mathsf{s}_2(n)} \right| \le C x^{1-\eta}$$

for some C *and* $\eta > 0$ *depending on* ε*, where* $D = x^{1-\varepsilon}$*.*

In more relaxed language: let $\rho > 0$. As $N \to \infty$, the following holds.

For most $d \in [N^\rho, 2N^\rho)$*, the number of times that* 0 *appears in*

$$\big(\mathtt{T}(nd+a)\big)_{0\le n<N}$$

is close to $N/2$*.*

Primes in arithmetic progressions

Remark. The level of distribution is an important concept in analytic number theory. The *Bombieri–Vinogradov theorem* states that the prime numbers have level of distribution (at least) $1/2$. This corresponds to progressions $(nd + a)_{0 \le n < N}$, where $d \le N^{1-\varepsilon}$.

Section 2

Digital expansions of prime numbers, Sarnak's conjecture

Mauduit and Rivat (2010, Ann. of Math.) proved the following.
Let $q \geq 2$, $m \geq 1$, and a be integers such that $\gcd(m, q-1) = 1$.
As p runs through the set of prime numbers, the expression $\mathsf{s}_q(p)$ hits each residue class modulo m with asymptotic frequency $1/m$.

The level of distribution-paper opens up a new path towards problems of this kind.

Theorem (Drmota–Müllner–S., submitted)

- *The sequence* $n \mapsto \exp(2\pi i \vartheta \mathsf{z}(n))$ *has level of distribution* 1.
- *For* $m \geq 1$ *and* $a \in \mathbb{Z}$, *we have*

$$\{p < x : p \text{ prime}, \mathsf{z}(p) \equiv a \bmod m\} \sim \frac{\pi(x)}{m}$$

  as $x \to \infty$.
- *For* k *large enough, there exists a prime number* p *that is the sum of exactly* k *different, non-consecutive Fibonacci numbers.*

Mauduit and Rivat (2010, Ann. of Math.) proved the following.
Let $q \geq 2$, $m \geq 1$, and a be integers such that $\gcd(m, q-1) = 1$.
As p runs through the set of prime numbers, the expression $\mathsf{s}_q(p)$ hits each residue class modulo m with asymptotic frequency $1/m$.

The level of distribution-paper opens up a new path towards problems of this kind.

Theorem (Drmota–Müllner–S., submitted)

- *The sequence* $n \mapsto \exp(2\pi i \vartheta \mathsf{z}(n))$ *has level of distribution* 1.
- *For* $m \geq 1$ *and* $a \in \mathbb{Z}$, *we have*

$$\{p < x : p \text{ prime}, \mathsf{z}(p) \equiv a \bmod m\} \sim \frac{\pi(x)}{m}$$

  as $x \to \infty$.
- *For* k *large enough, there exists a prime number* p *that is the sum of exactly* k *different, non-consecutive Fibonacci numbers.*

Sarnak's conjecture

The Möbius function μ is defined by

$$\mu(n) := \begin{cases} 0, & \text{if } n \text{ is divisible by a square;} \\ (-1)^m, & \text{if } n \text{ has } m \text{ prime factors.} \end{cases}$$

$$\mu = (1, -1, -1, 0, -1, 1, -1, 0, 0, 1, -1, \ldots)$$

Sarnak's conjecture intuitively states that μ behaves randomly. Sarnak formulated this in a precise sense using the language of dynamical systems. At the core of this conjecture we find the condition

$$\sum_{0 \le n < N} f(n)\mu(n) = o(N),$$

in which case the function f is said to satisfy a *Möbius randomness principle* (MRP).

Sarnak's conjecture

The Möbius function μ is defined by

$$\mu(n) := \begin{cases} 0, & \text{if } n \text{ is divisible by a square;} \\ (-1)^m, & \text{if } n \text{ has } m \text{ prime factors.} \end{cases}$$

$$\mu = (1, -1, -1, 0, -1, 1, -1, 0, 0, 1, -1, \ldots)$$

Sarnak's conjecture intuitively states that μ behaves randomly. Sarnak formulated this in a precise sense using the language of dynamical systems. At the core of this conjecture we find the condition

$$\sum_{0 \le n < N} f(n)\mu(n) = o(N),$$

in which case the function f is said to satisfy a *Möbius randomness principle* (MRP).

Sarnak's conjecture

The Möbius function μ is defined by

$$\mu(n) := \begin{cases} 0, & \text{if } n \text{ is divisible by a square;} \\ (-1)^m, & \text{if } n \text{ has } m \text{ prime factors.} \end{cases}$$

$$\mu = (1, -1, -1, 0, -1, 1, -1, 0, 0, 1, -1, \ldots)$$

Sarnak's conjecture intuitively states that μ behaves randomly. Sarnak formulated this in a precise sense using the language of dynamical systems. At the core of this conjecture we find the condition

$$\sum_{0 \le n < N} f(n)\mu(n) = o(N),$$

in which case the function f is said to satisfy a *Möbius randomness principle* (MRP).

Sarnak's conjecture

The Möbius function μ is defined by

$$\mu(n) := \begin{cases} 0, & \text{if } n \text{ is divisible by a square;} \\ (-1)^m, & \text{if } n \text{ has } m \text{ prime factors.} \end{cases}$$

$$\mu = (1, -1, -1, 0, -1, 1, -1, 0, 0, 1, -1, \ldots)$$

Sarnak's conjecture intuitively states that μ behaves randomly. Sarnak formulated this in a precise sense using the language of dynamical systems. At the core of this conjecture we find the condition

$$\sum_{0 \le n < N} f(n)\mu(n) = o(N),$$

in which case the function f is said to satisfy a *Möbius randomness principle* (MRP).

Automatic and morphic sequences

Sequences for which an MRP is expected to hold include *automatic* and *morphic* sequences.
The sequence $\mathtt{T}$ can be defined by

$$0 \mapsto 01, \quad 1 \mapsto 10.$$

It is an automatic sequence.

The sequence $\mathrm{z}(n) \bmod 2$ is given by the following substitution σ together with the coding π:

$$\begin{array}{lllll} \sigma: & \mathrm{a} \mapsto \mathrm{ab}, & \mathrm{b} \mapsto \mathrm{c}, & \mathrm{c} \mapsto \mathrm{cd}, & \mathrm{d} \mapsto \mathrm{a} \\ \pi: & \mathrm{a} \mapsto 0, & \mathrm{b} \mapsto 1, & \mathrm{c} \mapsto 1, & \mathrm{d} \mapsto 0 \end{array}$$

and we consider the fixed point starting with a. The sequence $(\mathrm{z}(n) \bmod 2)_{n \geq 0} = \pi\bigl(\sigma^\infty(\mathrm{a})\bigr)$ is a morphic sequence.

Automatic and morphic sequences

Sequences for which an MRP is expected to hold include *automatic* and *morphic* sequences.
The sequence $\mathtt{T}$ can be defined by

$$\mathtt{0} \mapsto \mathtt{01}, \quad \mathtt{1} \mapsto \mathtt{10}.$$

It is an automatic sequence.

The sequence $\mathsf{z}(n)$ mod 2 is given by the following substitution σ together with the coding π:

$$\begin{array}{lllll} \sigma: & \mathtt{a} \mapsto \mathtt{ab}, & \mathtt{b} \mapsto \mathtt{c}, & \mathtt{c} \mapsto \mathtt{cd}, & \mathtt{d} \mapsto \mathtt{a} \\ \pi: & \mathtt{a} \mapsto 0, & \mathtt{b} \mapsto 1, & \mathtt{c} \mapsto 1, & \mathtt{d} \mapsto 0 \end{array}$$

and we consider the fixed point starting with $\mathtt{a}$. The sequence $(\mathsf{z}(n) \bmod 2)_{n\ge 0} = \pi\bigl(\sigma^\infty(\mathtt{a})\bigr)$ is a morphic sequence.

More cases of Sarnak's conjecture

Müllner proved that all automatic sequences satisfy an MRP.

The major new goal is to "prove an MRP for all morphic sequences".

We plan to use the "level of distribution"-method, as applied in [S2020,DMS2022+], to other morphic sequences defined by numeration systems as well, and thus prove more cases of Sarnak's conjecture.

Also, it would be interesting to prove that (certain) automatic sequences have level of distribution equal to 1.

More cases of Sarnak's conjecture

Müllner proved that all automatic sequences satisfy an MRP.

The major new goal is to "prove an MRP for all morphic sequences".

We plan to use the "level of distribution"-method, as applied in [S2020,DMS2022+], to other morphic sequences defined by numeration systems as well, and thus prove more cases of Sarnak's conjecture.

Also, it would be interesting to prove that (certain) automatic sequences have level of distribution equal to 1.

More cases of Sarnak's conjecture

Müllner proved that all automatic sequences satisfy an MRP.

The major new goal is to "prove an MRP for all morphic sequences".

We plan to use the "level of distribution"-method, as applied in [S2020,DMS2022+], to other morphic sequences defined by numeration systems as well, and thus prove more cases of Sarnak's conjecture.

Also, it would be interesting to prove that (certain) automatic sequences have level of distribution equal to 1.

More cases of Sarnak's conjecture

Müllner proved that all automatic sequences satisfy an MRP.

The major new goal is to "prove an MRP for all morphic sequences".

We plan to use the "level of distribution"-method, as applied in [S2020,DMS2022+], to other morphic sequences defined by numeration systems as well, and thus prove more cases of Sarnak's conjecture.

Also, it would be interesting to prove that (certain) automatic sequences have level of distribution equal to 1.

Section 3

Digital expansions in different bases

"Collisions" of digit sums in different bases

A folklore conjecture states that the equation

$$\mathsf{s}_2(n) = \mathsf{s}_3(n)$$

admits infinitely many solutions n in the positive integers.

We proved this conjecture (positive referee report, Israel J. Math.).

Theorem (S. 2022+)

For all $\delta > 0$ we have

$$\#\{n < N : \mathsf{s}_2(n) = \mathsf{s}_3(n)\} \gg N^{\frac{\log 3}{\log 4} - \delta}, \tag{1}$$

where the implied constant may depend on δ.
Note that $\log 3 / \log 4 = 0.792\ldots$.

“Collisions” of digit sums in different bases

A folklore conjecture states that the equation

$$\mathsf{s}_2(n) = \mathsf{s}_3(n)$$

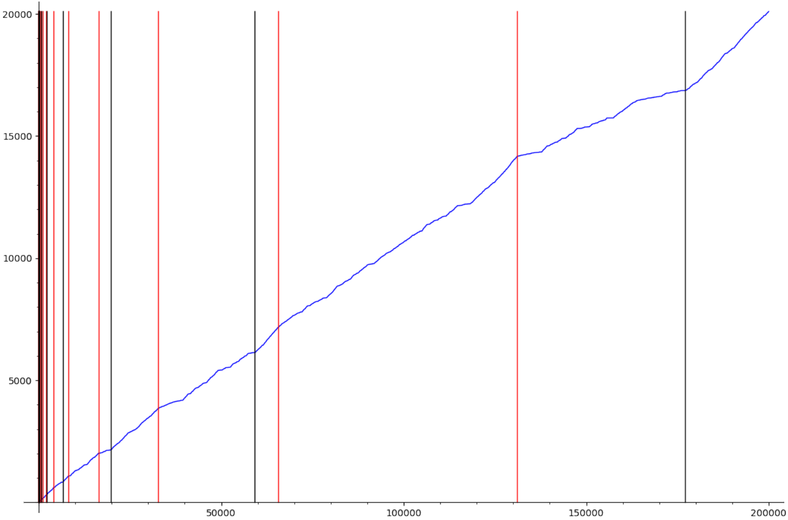
admits infinitely many solutions n in the positive integers.
We proved this conjecture (positive referee report, Israel J. Math.).

Theorem (S. 2022+)

For all $\delta > 0$ we have

$$\#\{n < N : \mathsf{s}_2(n) = \mathsf{s}_3(n)\} \gg N^{\frac{\log 3}{\log 4} - \delta}, \tag{1}$$

where the implied constant may depend on δ.
Note that $\log 3 / \log 4 = 0.792\ldots$.



blue: number of collisions; red: powers of 2; black: powers of 3

Even the arrangement of powers of 2 and 3 is somewhat cryptic.

$$(a_n)_{n\geq 0} = (1, 2, 3, 4, 8, 9, 16, 27, 32, 64, 81, 128, 243, 256, 512, 729, 1024, \ldots)$$

This amounts to understanding the *continued fraction expansion*

$$\frac{\log 3}{\log 2} = [1; 1, 1, 2, 2, 3, 1, 5, 2, 23, 2, 2, 1, 1, 55, 1, 4, 3, 1, 1, \ldots],$$

which is unknown!

This topic has connections to dynamical systems (*Furstenberg's conjectures* on joint digital expansions in different bases), Diophantine approximation (estimates for the irrationality exponent of $\log_2 3$), and Mahler's 3/2-problem (can we have $\{x(3/2)^n\} < 1/2$ for all $n \geq 0$?).

Even the arrangement of powers of 2 and 3 is somewhat cryptic.

$$(a_n)_{n\geq 0} = (1, 2, 3, 4, 8, 9, 16, 27, 32, 64, 81, 128, 243, 256, 512, 729, 1024, \ldots)$$

This amounts to understanding the *continued fraction expansion*

$$\frac{\log 3}{\log 2} = [1; 1, 1, 2, 2, 3, 1, 5, 2, 23, 2, 2, 1, 1, 55, 1, 4, 3, 1, 1, \ldots],$$

which is unknown!

This topic has connections to dynamical systems (*Furstenberg's conjectures* on joint digital expansions in different bases), Diophantine approximation (estimates for the irrationality exponent of $\log_2 3$), and Mahler's 3/2-problem (can we have $\{x(3/2)^n\} < 1/2$ for all $n \geq 0$?).

Even the arrangement of powers of 2 and 3 is somewhat cryptic.

$$(a_n)_{n\geq 0} = (1, 2, 3, 4, 8, 9, 16, 27, 32, 64, 81, 128, 243, 256, 512, 729, 1024, \ldots)$$

This amounts to understanding the *continued fraction expansion*

$$\frac{\log 3}{\log 2} = [1; 1, 1, 2, 2, 3, 1, 5, 2, 23, 2, 2, 1, 1, 55, 1, 4, 3, 1, 1, \ldots],$$

which is unknown!

This topic has connections to dynamical systems (*Furstenberg's conjectures* on joint digital expansions in different bases), Diophantine approximation (estimates for the irrationality exponent of $\log_2 3$), and Mahler's 3/2-problem (can we have $\{x(3/2)^n\} < 1/2$ for all $n \geq 0$?).

A remark on the separation of sum-of-digits functions

The values of $\mathsf{s}_2(n)$ and $\mathsf{s}_3(n)$, as $n < N$ concentrate around $\log_4(N)$ and $\log_3(N)$ respectively. The standard deviations are small compared to the difference of expected values!

By Hoeffding's inequality on i.i.d. random variables there is only a number $\ll N^\alpha$ of collisions $n < N$, where $\alpha < 1$. Our result therefore cannot be too far from the actual number of collisions.

Conjecture

There exist constants c and η such that

$$\#\big\{n < N : \mathsf{s}_2(n) = \mathsf{s}_3(n)\big\} \sim cN^\eta.$$

A remark on the separation of sum-of-digits functions

The values of $\mathsf{s}_2(n)$ and $\mathsf{s}_3(n)$, as $n < N$ concentrate around $\log_4(N)$ and $\log_3(N)$ respectively. The standard deviations are small compared to the difference of expected values!

By Hoeffding's inequality on i.i.d. random variables there is only a number $\ll N^\alpha$ of collisions $n < N$, where $\alpha < 1$. Our result therefore cannot be too far from the actual number of collisions.

Conjecture

There exist constants c and η such that

$$\#\{n < N : \mathsf{s}_2(n) = \mathsf{s}_3(n)\} \sim cN^\eta.$$

A remark on the separation of sum-of-digits functions

The values of $\mathsf{s}_2(n)$ and $\mathsf{s}_3(n)$, as $n < N$ concentrate around $\log_4(N)$ and $\log_3(N)$ respectively. The standard deviations are small compared to the difference of expected values!

By Hoeffding's inequality on i.i.d. random variables there is only a number $\ll N^\alpha$ of collisions $n < N$, where $\alpha < 1$. Our result therefore cannot be too far from the actual number of collisions.

Conjecture

There exist constants c and η such that

$$\#\{n < N : \mathsf{s}_2(n) = \mathsf{s}_3(n)\} \sim cN^\eta.$$

Connection to the main topic of the talk

The central idea of the proof is a simple heuristic. We have

$$\mathsf{s}_3\bigl(3^\zeta n\bigr) = \mathsf{s}_3(n),$$

while the binary digits of $3^\zeta n$ should be "random". We therefore expect

$$\mathsf{s}_2\bigl(3^\zeta n\bigr) \approx \log_4\bigl(3^\zeta n\bigr) = \zeta\frac{\log 3}{\log 4} + \frac{\log(n)}{\log 4}.$$

Let us choose

$$\zeta \approx \frac{\log(n)}{\log 3}\left(\frac{\log 4}{\log 3} - 1\right).$$

Then $\mathsf{s}_2(3^\zeta n)$ and $\mathsf{s}_3(3^\zeta n)$ should concentrate around the same expected value. We will look for collisions along $3^\zeta\mathbb{N}$!

Connection to the main topic of the talk

The central idea of the proof is a simple heuristic. We have

$$\mathsf{s}_3\bigl(3^\zeta n\bigr) = \mathsf{s}_3(n),$$

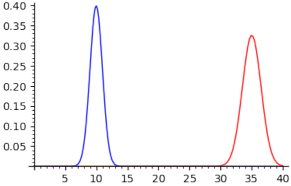
while the binary digits of $3^\zeta n$ should be “random”. We therefore expect

$$\mathsf{s}_2\bigl(3^\zeta n\bigr) \approx \log_4\bigl(3^\zeta n\bigr) = \zeta\frac{\log 3}{\log 4} + \frac{\log(n)}{\log 4}.$$

Let us choose

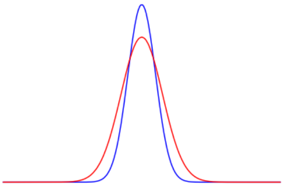
$$\zeta \approx \frac{\log(n)}{\log 3}\left(\frac{\log 4}{\log 3} - 1\right).$$

Then $\mathsf{s}_2(3^\zeta n)$ and $\mathsf{s}_3(3^\zeta n)$ should concentrate around the same expected value. We will look for collisions along $3^\zeta\mathbb{N}$!



all n

$\rightarrow$



only n in a residue class

Section 4

Long arithmetic subsequences — correlations

In the following, let us assume that $d \ge 1$ and $a \ge 0$. We are concerned with the behaviour of s_2 along the arithmetic progression $(nd + a)_{n \ge 0}$.
In other words,

how does the sum of digits of an integer change when a constant d is added repeatedly?

Let us define

$$\delta(j, d, a) := \lim_{N \to \infty} \frac{1}{N} \#\bigl\{n < N : \mathsf{s}_2((n+1)d + a) - \mathsf{s}_2(nd + a) = j\bigr\}.$$

This value is in fact identical to

$$\delta(j, d) := \lim_{N \to \infty} \frac{1}{N} \#\bigl\{n < N : \mathsf{s}_2(n + d) - \mathsf{s}_2(n) = j\bigr\},$$

for all a.

In the following, let us assume that $d \geq 1$ and $a \geq 0$. We are concerned with the behaviour of s_2 along the arithmetic progression $(nd + a)_{n \geq 0}$. In other words,

how does the sum of digits of an integer change when a constant d is added repeatedly?

Let us define

$$\delta(j, d, a) := \lim_{N \to \infty} \frac{1}{N} \#\bigl\{n < N : \mathsf{s}_2((n+1)d + a) - \mathsf{s}_2(nd + a) = j\bigr\}.$$

This value is in fact identical to

$$\delta(j, d) := \lim_{N \to \infty} \frac{1}{N} \#\bigl\{n < N : \mathsf{s}_2(n + d) - \mathsf{s}_2(n) = j\bigr\},$$

for all a.

In the following, let us assume that $d \ge 1$ and $a \ge 0$. We are concerned with the behaviour of s_2 along the arithmetic progression $(nd+a)_{n\ge 0}$. In other words,

> *how does the sum of digits of an integer change when a constant d is added repeatedly?*

Let us define

$$\delta(j,d,a) \coloneqq \lim_{N\to\infty} \frac{1}{N} \#\bigl\{n < N : \mathsf{s}_2((n+1)d+a) - \mathsf{s}_2(nd+a) = j\bigr\}.$$

This value is in fact identical to

$$\delta(j,d) \coloneqq \lim_{N\to\infty} \frac{1}{N} \#\bigl\{n < N : \mathsf{s}_2(n+d) - \mathsf{s}_2(n) = j\bigr\},$$

for all a.

Remark. The *(auto)correlation*

$$\gamma_d := \lim_{N\to\infty} \frac{1}{N} \sum_{n<N} (-1)^{\mathsf{s}_2(n+d)-\mathsf{s}_2(n)}$$

can be computed by a recurrence, starting from $\gamma_1 = -1/3$, and has connections to harmonic analysis (Mahler) and symbolic dynamical systems.

"Fourier coefficients of the spectral measure associated to the Thue–Morse dynamical system".

Remark. The *(auto)correlation*

$$\gamma_d := \lim_{N\to\infty} \frac{1}{N} \sum_{n<N} (-1)^{\mathsf{s}_2(n+d)-\mathsf{s}_2(n)}$$

can be computed by a recurrence, starting from $\gamma_1 = -1/3$, and has connections to harmonic analysis (Mahler) and symbolic dynamical systems.

“Fourier coefficients of the spectral measure associated to the Thue–Morse dynamical system”.

Cusick's conjecture

When traversing an infinite arithmetic subsequence of s_2, how often does the value stay constant or increase? This is the subject of *Cusick's conjecture*.

Conjecture (Cusick)

For all $d \geq 0$, we have

$$c_d := \qquad \delta(0, d) \qquad + \delta(1, d) + \delta(2, d) + \cdots > 1/2,$$

where

$$\delta(j, d) = \lim_{N \to \infty} \frac{1}{N} \#\{n < N : \mathsf{s}_2(n + d) - \mathsf{s}_2(n) = j\}.$$

Cusick's conjecture

When traversing an infinite arithmetic subsequence of s_2, how often does the value stay constant or increase? This is the subject of *Cusick's conjecture*.

Conjecture (Cusick)

For all $d \geq 0$, we have

$$c_d := \underbrace{\delta(0,d)}_{\text{“stays constant”}} + \underbrace{\delta(1,d) + \delta(2,d) + \cdots}_{\text{“increases”}} > 1/2,$$

where

$$\delta(j,d) = \lim_{N \to \infty} \frac{1}{N} \#\{n < N : \mathsf{s}_2(n+d) - \mathsf{s}_2(n) = j\}.$$

SW2021

Let $M = M(d)$ be the number of blocks of 1s in the binary expansion of d.

Theorem (S.–Wallner 2021, Annali SNS)

Set $\kappa_2(1) = 1$, and for $d \geq 1$ let $\kappa_2(2d) = \kappa_2(d)$, and

$$\kappa_2(2d+1) = \frac{\kappa_2(d) + \kappa_2(d+1)}{2} + 1.$$

If M is larger than some absolute, effective constant M_0, we have

$$\delta(j,d) = \frac{1}{\sqrt{2\pi\kappa_2(d)}} \exp\left(-\frac{j^2}{2\kappa_2(d)}\right) + \mathcal{O}\left(\frac{(\log M)^4}{M}\right)$$

for all integers j. The implied constant is absolute.

"The sum of digits along arithmetic progressions changes according to a normal distribution."

SW2021

Let $M = M(d)$ be the number of blocks of 1s in the binary expansion of d.

Theorem (S.–Wallner 2021, Annali SNS)

Set $\kappa_2(1) = 1$, and for $d \geq 1$ let $\kappa_2(2d) = \kappa_2(d)$, and

$$\kappa_2(2d+1) = \frac{\kappa_2(d) + \kappa_2(d+1)}{2} + 1.$$

If M is larger than some absolute, effective constant M_0, we have

$$\delta(j,d) = \frac{1}{\sqrt{2\pi\kappa_2(d)}} \exp\left(-\frac{j^2}{2\kappa_2(d)}\right) + \mathcal{O}\left(\frac{(\log M)^4}{M}\right)$$

for all integers j. The implied constant is absolute.

"The sum of digits along arithmetic progressions changes according to a normal distribution."

SW2020, part II

Again, let $M = M(d)$ be the number of blocks of 1s in d.

Theorem (S.–Wallner 2021, Annali SNS)

Let $d \geq 1$. If $M(d)$ is larger than some absolute, effective constant M_1, then $c_d > 1/2$.

Cusick: "Your paper reduces my conjecture to what I will call the 'hard cases' [...]". $\longrightarrow$ more work to do! ☕

SW2020, part II

Again, let $M = M(d)$ be the number of blocks of 1s in d.

Theorem (S.–Wallner 2021, Annali SNS)

Let $d \geq 1$. If $M(d)$ is larger than some absolute, effective constant M_1, then $c_d > 1/2$.

Cusick: "Your paper reduces my conjecture to what I will call the 'hard cases' [...]". $\longrightarrow$ more work to do!

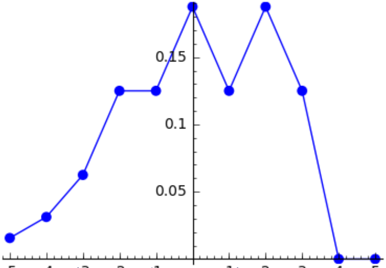
SW2020, part II

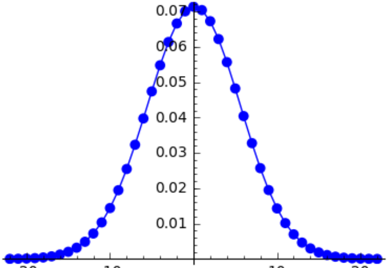
Again, let $M = M(d)$ be the number of blocks of 1s in d.

Theorem (S.–Wallner 2021, Annali SNS)

Let $d \geq 1$. If $M(d)$ is larger than some absolute, effective constant M_1, then $c_d > 1/2$.

Cusick: “Your paper reduces my conjecture to what I will call the ‘hard cases’ [...]”. $\longrightarrow$ more work to do!





Thank you!

[0] Supported by the Austrian Science Fund (FWF), Project ArithRand (jointly with ANR).