Subsequences of digitally defined functions

Lukas Spiegelhofer

TECHNISCHE UNIVERSITÄT WIEN
Vienna University of Technology

MONTAN UNIVERSITÄT

October 31, 2023, TU Wien

Outline

Section 1

Digital expansions

In the simplest case, a *digital expansion* Φ assigns to each natural number a finite string of *digits* in an injective manner.

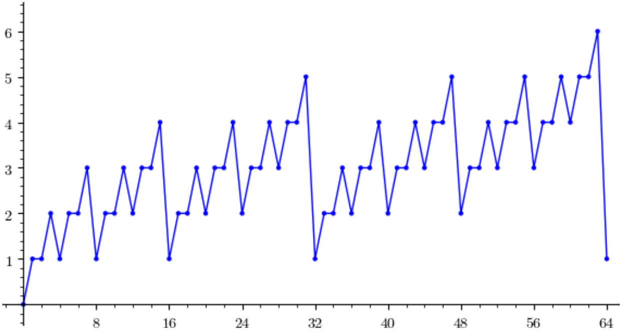
n	$\Phi(n)$	n	$\Phi(n)$	n	$\Phi(n)$
0	0	8	1000	16	10000
1	1	9	1001	17	10001
2	10	10	1010	18	10010
3	11	11	1011	19	10011
4	100	12	1100	20	10100
5	101	13	1101	21	10101
6	110	14	1110	22	10110
7	111	15	1111	23	10111

Figure: The binary expansion

Let $s_2(n)$ be the number of 1s in the binary expansion of n.
The sequence s_2 is a fixed point of the substitution defined by

$$n \mapsto (n, n+1), \quad n \geq 0.$$

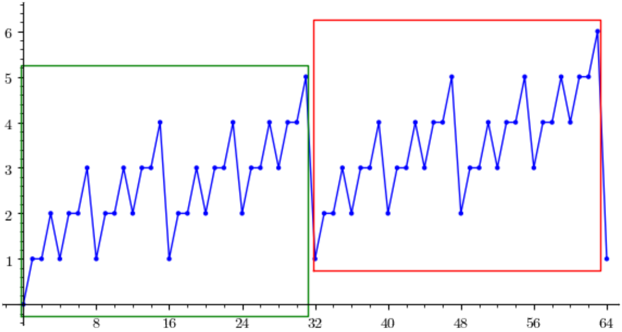
$$s_2 = 011212231223233412232334233434 45\cdots$$



Let $\mathsf{s}_2(n)$ be the number of 1s in the binary expansion of n.
The sequence s_2 is a fixed point of the substitution defined by

$$n \mapsto (n, n+1), \quad n \geq 0.$$

$$\mathsf{s}_2 = 011212231223233412232334233434 45\cdots$$



The Thue–Morse sequence

Let $\mathsf{t}(n) = \mathsf{s}_2(n) \bmod 2$.

$$\mathsf{t} = 01101001 \cdots .$$

This sequence is an *automatic sequence*, and a fixed point of

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

Automatic sequences are given by deterministic finite automata with output (DFAO), where the input is the base-q expansion of integers.

The Thue–Morse sequence

Let $\mathrm{t}(n) = \mathrm{s}_2(n) \bmod 2$.

$$\mathrm{t} = 01101001\cdots.$$

This sequence is an *automatic sequence*, and a fixed point of

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

Automatic sequences are given by deterministic finite automata with output (DFAO), where the input is the base-q expansion of integers.

2-automatic sequences

Each subsequence $n \mapsto \mathrm{t}(An + B)$ is an automatic sequence.

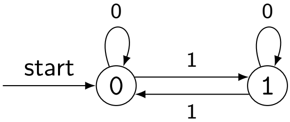


Figure: An automaton for $\mathrm{t}(n)$

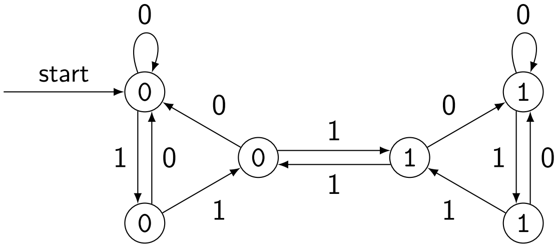


Figure: An automaton for $\mathrm{t}(3n)$

Section 2

Sparse subsequences

Few big steps

General idea: along sparse subsequences, t behaves “randomly”. For example, every finite sequence on $\{0, 1\}$ appears as an arithmetic subsequence of t (Avgustinovich–Fon-Der-Flaass–Frid 2003, Müllner–Spiegelhofer 2017 (“correct” rate of appearance), Konieczny–Müllner 2023+ (general automatic sequences)).

Very sparse arithmetic subsequences of t

The Thue–Morse sequence has level of distribution 1.

Theorem (S. 2020, Compos. Math.)

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 $R > 0$. As $N \to \infty$, the following holds.

> *Most $d \asymp N^R$ have the property that for all a, the number*
>
> $$\#\{0 \le n < N : \mathrm{t}(nd + a) = 0\}$$
>
> *is close to $N/2$.*

Very sparse arithmetic subsequences of t

The Thue–Morse sequence has level of distribution 1.

Theorem (S. 2020, Compos. Math.)

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 $R > 0$. As $N \to \infty$, the following holds.

> *Most $d \asymp N^R$ have the property that for all a, the number*
>
> $$\#\{0 \le n < N : \mathsf{t}(nd + a) = 0\}$$
>
> *is close to $N/2$.*

Drmota–Müllner–S. 2023+

Together with Drmota and Müllner, we proved a theorem on the subsequence indexed by the sequence of primes. Let $\mathsf{z}(n)$ be the smallest k such that n is the number of k Fibonacci numbers.

Theorem (Drmota–Müllner–S., to appear in Mem. Amer. Math. Soc.)

- *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*

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

  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.*

S. 2023+

The recent manuscript "Thue–Morse along the sequence of cubes" contains a proof the following theorem.

Theorem (S. 2023+)

As $N \to \infty$, we have

$$\frac{1}{N}\#\big\{n < N : \mathrm{t}\big(n^3\big) = 0\big\} \to \frac{1}{2}.$$

Informal open questions

- Prove that any finite word ω over $\{0,1\}$ appears in $\mathtt{t}$ about the expected number of times, for most very sparse arithmetic progressions.
- Prove prime number theorems for more general *morphic sequences*, such as

$$\sigma : \left\{ \begin{array}{lllllllll} \mathtt{a} & \mapsto & \mathtt{ae}, & \mathtt{b} & \mapsto & \mathtt{af}, & \mathtt{c} & \mapsto & \mathtt{a} \\ \mathtt{d} & \mapsto & \mathtt{db}, & \mathtt{e} & \mapsto & \mathtt{dc}, & \mathtt{f} & \mapsto & \mathtt{d} \end{array} \right\},$$

$$\pi(\mathtt{a}) = \pi(\mathtt{b}) = \pi(\mathtt{c}) = 0,$$
$$\pi(\mathtt{d}) = \pi(\mathtt{e}) = \pi(\mathtt{f}) = 1$$

The projection under π of the fixed point starting with $\mathtt{a}$ is

$$\mathtt{tr} = 01101001001011001011011010010110110100110 10\cdots.$$

Informal open questions

- Prove that any finite word ω over $\{0,1\}$ appears in $\mathtt{t}$ about the expected number of times, for most very sparse arithmetic progressions.
- Prove prime number theorems for more general *morphic sequences*, such as

$$\sigma : \left\{ \begin{array}{llllllll} \mathtt{a} & \mapsto & \mathtt{ae}, & \mathtt{b} & \mapsto & \mathtt{af}, & \mathtt{c} & \mapsto & \mathtt{a} \\ \mathtt{d} & \mapsto & \mathtt{db}, & \mathtt{e} & \mapsto & \mathtt{dc}, & \mathtt{f} & \mapsto & \mathtt{d} \end{array} \right\},$$

$$\pi(\mathtt{a}) = \pi(\mathtt{b}) = \pi(\mathtt{c}) = \mathtt{0},$$
$$\pi(\mathtt{d}) = \pi(\mathtt{e}) = \pi(\mathtt{f}) = \mathtt{1}$$

The projection under π of the fixed point starting with $\mathtt{a}$ is

$$\mathtt{tr} = \mathtt{0110100100101100101101101001011011010011010}\cdots.$$

More open questions

- Complete the treatment of the *third Gelfond problem (1967/1968)*: consider the sum-of-digits function along arbitrary polynomials f such that $f(\mathbb{N}) \subseteq \mathbb{N}$.
- Prove a level-of-distribution result for arbitrary automatic sequences.

More open questions

- Complete the treatment of the *third Gelfond problem (1967/1968)*: consider the sum-of-digits function along arbitrary polynomials f such that $f(\mathbb{N}) \subseteq \mathbb{N}$.
- Prove a level-of-distribution result for arbitrary automatic sequences.

Section 3

Long arithmetic subsequences

Digital expansions and addition

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

Differences along an arithmetic progression $a + d\mathbb{N}$:

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

For all a, this value is in fact identical to

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

Digital expansions and addition

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

Differences along an arithmetic progression $a + d\mathbb{N}$:

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

For all a, this value is in fact identical to

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

Cusick's conjecture

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

Conjecture (Cusick)

For all $d \geq 0$, we have

$$c_d > 1/2,$$

where

$$\begin{aligned} c_d &= \lim_{N\to\infty} \frac{1}{N} \#\bigl\{n < N : \mathsf{s}_2(n+d) \geq \mathsf{s}_2(n)\bigr\} \\ &= \sum_{j\geq 0} \delta(d,j). \end{aligned}$$

SW2023

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

Theorem (S.–Wallner 2023, Ann. Sc. norm. super. Pisa - Cl. sci.)

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

SW2023, part II

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

Theorem (S.–Wallner 2023)

Set $\kappa(1) = 2$, *and for* $d \ge 1$ *let* $\kappa(2d) = \kappa(d)$, *and*

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

Then

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

for all integers j. *The implied constant is absolute.*

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

SW2023, part II

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

Theorem (S.–Wallner 2023)

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

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

Then

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

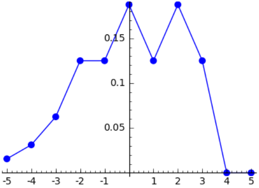
for all integers j. The implied constant is absolute.

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

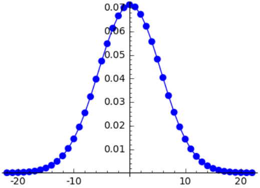
"Your paper reduces my conjecture to what I will call the 'hard cases' [. . .]" (T. W. Cusick, 2021)

"Your paper reduces my conjecture to what I will call the 'hard cases' [...]" (T. W. Cusick, 2021)

Consider graphs of $j \mapsto \delta(d, j)$:



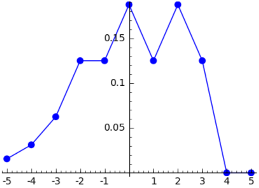
hard



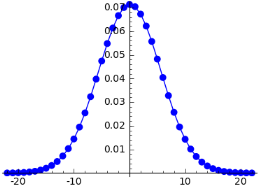
easier

"Your paper reduces my conjecture to what I will call the 'hard cases' [...]" (T. W. Cusick, 2021)

Consider graphs of $j \mapsto \delta(d, j)$:



hard



easier

More structures to be discovered!

Informal open questions

- Reduce the bound M_1 for the number of blocks needed for $c_t > 1/2$ to hold. Goal: $M_1 = 0$.
- Prove a *polynomially perturbed Gaussian law* for the numbers

$$\Theta(n,j) = \#\left\{0 \le k \le n : \nu_2\left(\binom{n}{k}\right) = j\right\},$$

where $\Theta(n,j)$ is compared to

$$p(j)\exp\left(-K(j-E)^2/2\right),$$

and $p = p_n$ is a certain polynomial. This was asked by R. Tichy (2020), and work in progress with B. Sobolewski (Kraków/Leoben);
- Study several differences $s_2(n+t_j) - s_2(n)$ simultaneously, and prove e.g. multidimensional Gaussian laws. For this, consider generalized correlations

$$\sum_n \exp\left(\vartheta s_2(n) + \sum_{j<J} \vartheta_j s_2(n+t_j)\right).$$

Informal open questions

- Reduce the bound M_1 for the number of blocks needed for $c_t > 1/2$ to hold. Goal: $M_1 = 0$.
- Prove a *polynomially perturbed Gaussian law* for the numbers

$$\Theta(n, j) = \#\left\{0 \le k \le n : \nu_2\left(\binom{n}{k}\right) = j\right\},$$

where $\Theta(n, j)$ is compared to

$$p(j)\exp\bigl(-K(j-E)^2/2\bigr),$$

and $p = p_n$ is a certain polynomial. This was asked by R. Tichy (2020), and work in progress with B. Sobolewski (Kraków/Leoben);

- Study several differences $s_2(n + t_j) - s_2(n)$ simultaneously, and prove e.g. multidimensional Gaussian laws. For this, consider generalized correlations

$$\sum_n \exp\left(\vartheta s_2(n) + \sum_{j<J} \vartheta_j s_2(n + t_j)\right).$$

Informal open questions

- Reduce the bound M_1 for the number of blocks needed for $c_t > 1/2$ to hold. Goal: $M_1 = 0$.
- Prove a *polynomially perturbed Gaussian law* for the numbers
$$\Theta(n,j) = \#\left\{0 \le k \le n : \nu_2\left(\binom{n}{k}\right) = j\right\},$$
where $\Theta(n,j)$ is compared to
$$p(j)\exp\bigl(-K(j-E)^2/2\bigr),$$
and $p = p_n$ is a certain polynomial. This was asked by R. Tichy (2020), and work in progress with B. Sobolewski (Kraków/Leoben);
- Study several differences $s_2(n+t_j) - s_2(n)$ simultaneously, and prove e.g. multidimensional Gaussian laws. For this, consider generalized correlations
$$\sum_n \exp\left(\vartheta\, \mathsf{s}_2(n) + \sum_{j<J} \vartheta_j\, \mathsf{s}_2(n+t_j)\right).$$

Section 4

Digital expansions in different bases

“Collisions” of digit sums in different bases

A former folklore conjecture states that the equation

$$s_2(n) = s_3(n)$$

admits infinitely many solutions n in the positive integers.

Theorem (S. 2023, Israel J. Math.)

For all $\delta > 0$ we have

$$\#\{n < N : s_2(n) = s_3(n)\} \gg N^{\frac{\log 3}{\log 4} - \delta}, \qquad (1)$$

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 former folklore conjecture states that the equation

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

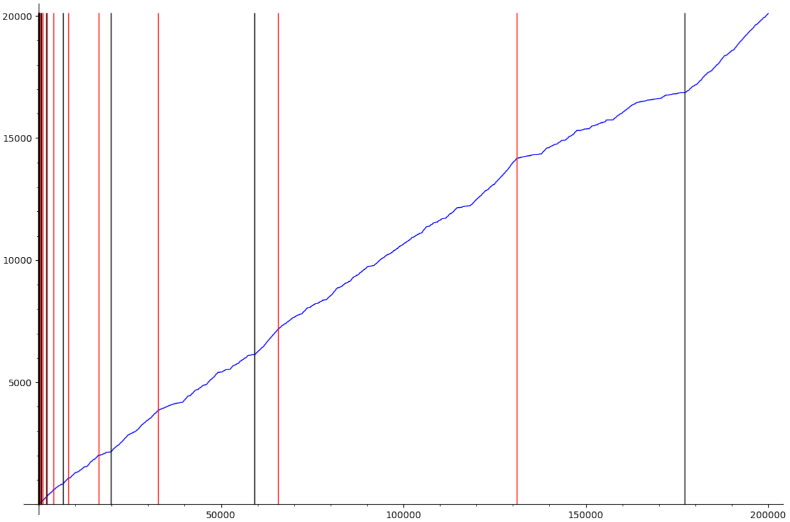
admits infinitely many solutions n in the positive integers.

Theorem (S. 2023, Israel J. Math.)

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 up to N; 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.

Using tail estimates, we see that there can only exist few collisions.

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.

Using tail estimates, we see that there can only exist few collisions.

Connection to the main topic of the talk

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

$$s_3\left(3^\zeta n\right) = s_3(n),$$

which has a value close to $\log(n)/\log(3)$ for most n. Meanwhile, the binary digits of $3^\zeta n$ should be "random", and therefore

$$s_2\left(3^\zeta n\right) \approx \log_4\left(3^\zeta n\right) = \frac{\log(n)}{\log 4} + \zeta\frac{\log 3}{\log 4}.$$

We can therefore adjust ζ so that $s_3\left(3^\zeta n\right)$ and $s_2\left(3^\zeta n\right)$ have many values close to each other.

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

$$s_3\left(3^\zeta n\right) = s_3(n),$$

which has a value close to $\log(n)/\log(3)$ for most n. Meanwhile, the binary digits of $3^\zeta n$ should be "random", and therefore

$$s_2\left(3^\zeta n\right) \approx \log_4\left(3^\zeta n\right) = \frac{\log(n)}{\log 4} + \zeta\frac{\log 3}{\log 4}.$$

We can therefore adjust ζ so that $s_3\left(3^\zeta n\right)$ and $s_2\left(3^\zeta n\right)$ have many values close to each other.

We will look for collisions along $3^\zeta\mathbb{N}$!

Connection to the main topic of the talk

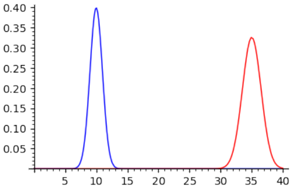
The central idea of the proof is a simple heuristic. We have

$$s_3\left(3^\zeta n\right) = s_3(n),$$

which has a value close to $\log(n)/\log(3)$ for most n. Meanwhile, the binary digits of $3^\zeta n$ should be "random", and therefore

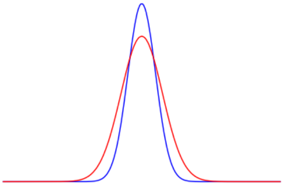
$$s_2\left(3^\zeta n\right) \approx \log_4\left(3^\zeta n\right) = \frac{\log(n)}{\log 4} + \zeta\frac{\log 3}{\log 4}.$$

We can therefore adjust ζ so that $s_3\left(3^\zeta n\right)$ and $s_2\left(3^\zeta n\right)$ have many values close to each other.
We will look for collisions along $3^\zeta\mathbb{N}$!



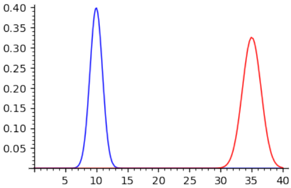
all n

$\rightarrow$

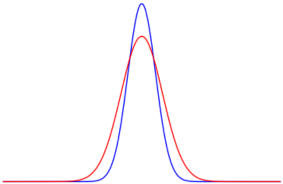


only $n \in \left(3^{\zeta}\mathbb{Z}\right) \cap I$

Transition to a *subsequence of a digitally defined function* enabled us to prove the folklore conjecture.



$\rightarrow$



all n only $n \in (3^\zeta \mathbb{Z}) \cap I$

Transition to a *subsequence of a digitally defined function* enabled us to prove the folklore conjecture.

Open questions

- Work in Progress with M. Drmota: new proof of the theorem, refinements. With J.-M. Deshouillers: digits of $n!$ in base 12.
- Conjecture: There exist constants c and η such that

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

- Prove that there are infinitely many prime numbers p such that $\mathsf{s}_2(p) = \mathsf{s}_3(p)$.

Open questions

- Work in Progress with M. Drmota: new proof of the theorem, refinements. With J.-M. Deshouillers: digits of $n!$ in base 12.
- Conjecture: There exist constants c and η such that

$$\#\{n < N : s_2(n) = s_3(n)\} \sim cN^\eta.$$

- Prove that there are infinitely many prime numbers p such that $s_2(p) = s_3(p)$.

Open questions

- Work in Progress with M. Drmota: new proof of the theorem, refinements. With J.-M. Deshouillers: digits of $n!$ in base 12.
- Conjecture: There exist constants c and η such that

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

- Prove that there are infinitely many prime numbers p such that $\mathsf{s}_2(p) = \mathsf{s}_3(p)$.

THANK YOU!

- M. Drmota, C. Müllner, and L. Spiegelhofer, *Primes as sums of Fibonacci numbers*, 2021.
  Accepted for publication in Mem. Amer. Math. Soc. (2022).
- L. Spiegelhofer, *The level of distribution of the Thue–Morse sequence*, Compos. Math., 156 (2020), pp. 2560–2587.
- L. Spiegelhofer, *Collisions of digit sums in bases 2 and 3*, 2023.
  To appear in Israel J. Math.
- L. Spiegelhofer, *Thue–Morse along the sequence of cubes*, 2023.
  Preprint, http://arxiv.org/abs/2308.09498.
- L. Spiegelhofer and M. Wallner, *The binary digits of $n + t$*, Ann. Sc. Norm. Super. Pisa, Cl. Sci. (5), 24 (2023), pp. 1–31.

Supported by the FWF–ANR joint project ArithRand, and P36137 (FWF).