The base-2 expansion along arithmetic progressions

Lukas Spiegelhofer

March 24, 2021, Departmentseminar Mathematik und Informationstechnologie

The basic question

We write n in base 2:

$$n = \varepsilon_0 2^0 + \varepsilon_1 2^1 + \varepsilon_2 2^2 + \cdots + \varepsilon_\nu 2^\nu,$$

where $\varepsilon_j \in \{0, 1\}$. The vector $(\varepsilon_j)_{j \geq 0}$ is the *binary expansion* of n.

Base ten	Base two
0	0
1	1
2	10
3	11
4	100
5	101
6	110
7	111
8	1000

The basic question

We write n in base 2:

$$n = \varepsilon_0 2^0 + \varepsilon_1 2^1 + \varepsilon_2 2^2 + \cdots + \varepsilon_\nu 2^\nu,$$

where $\varepsilon_j \in \{0, 1\}$. The vector $(\varepsilon_j)_{j \geq 0}$ is the *binary expansion* of n.

Base ten	Base two
0	0
1	1
2	10
3	11
4	100
5	101
6	110
7	111
8	1000

The basic question

We write n in base 2:

$$n = \varepsilon_0 2^0 + \varepsilon_1 2^1 + \varepsilon_2 2^2 + \cdots + \varepsilon_\nu 2^\nu,$$

where $\varepsilon_j \in \{0, 1\}$. The vector $(\varepsilon_j)_{j \geq 0}$ is the *binary expansion* of n.

Base ten	Base two
0	0
1	1
2	10
3	11
4	100
5	101
6	110
7	111
8	1000

$+t$

$+t$

$+t$

Central question:

What happens to the binary expansion of n when a constant t is added?

There is no final answer yet

The (slightly provocative) answer is

"Addition in base 2 is not fully understood".

The appearance of *carries* in the addition $n+t$ causes many cases to be distinguished, and a structural result describing these cases is not available.

$$\begin{array}{rr} & 11101001110110011 \\ + & 10110001001101 \end{array}$$

Questions of this kind have strong connections to computer science and are relevant in cryptography.

There is no final answer yet

The (slightly provocative) answer is

"Addition in base 2 is not fully understood".

The appearance of *carries* in the addition $n+t$ causes many cases to be distinguished, and a structural result describing these cases is not available.

$$\begin{array}{rr} & 11101001110110011 \\ + & 10110001001101 \end{array}$$

Questions of this kind have strong connections to computer science and are relevant in cryptography.

There is no final answer yet

The (slightly provocative) answer is

"Addition in base 2 is not fully understood".

The appearance of *carries* in the addition $n + t$ causes many cases to be distinguished, and a structural result describing these cases is not available.

$$\begin{array}{rr} & 1110100111011001\,1 \\ + & 10110001001101 \end{array}$$

Questions of this kind have strong connections to computer science and are relevant in cryptography.

Addition of 1

The (possibly empty) block of 1s on the right of the binary expansion of n is replaced by 0s, and the 0 to the left of the block is replaced by 1.

$$*\,011\cdots1 \mapsto *100\cdots0 \tag{1}$$

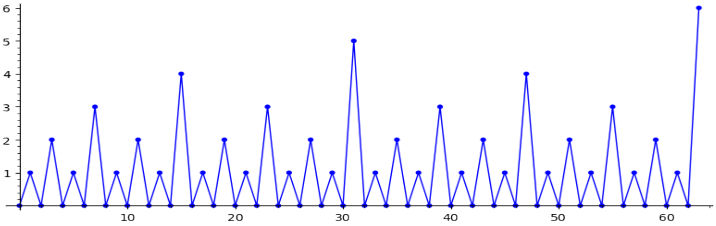


Figure: The number of carries in the addition $n+1$

This is the *ruler sequence* $n \mapsto \nu_2(n+1)$, given by the exponent of two in the prime factorization of $n+1$.

Addition of 1

The (possibly empty) block of 1s on the right of the binary expansion of n is replaced by 0s, and the 0 to the left of the block is replaced by 1.

$$* 011 \cdots 1 \mapsto *100 \cdots 0 \tag{1}$$

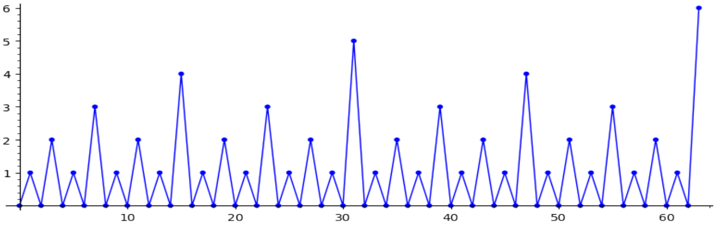
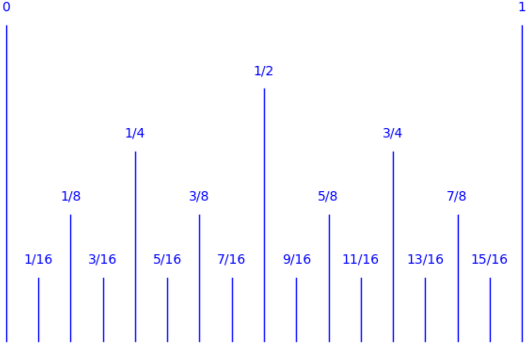


Figure: The number of carries in the addition $n+1$

This is the *ruler sequence* $n \mapsto \nu_2(n+1)$, given by the exponent of two in the prime factorization of $n+1$.

The ruler sequence

The following picture is well known in countries using imperial units.



The case $t \geq 3$

For $t = 3$ we have the following cases:

$$*00 \mapsto *11; \qquad *01^k01 \mapsto *10^k00;$$
$$*01^k10 \mapsto *10^k01; \qquad *01^k11 \mapsto *10^k10.$$

This situation does not get better with growing t. Carries can propagate through many blocks of 1, and many cases occur.

The binary sum-of-digits function

To simplify things, we consider the *binary sum-of-digits function* s. The integer $s(n)$ is the minimal number of powers of 2 needed to write n as their sum.

n	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
$s(n)$	0	1	1	2	1	2	2	3	1	2	2	3	2	3	3	4

The `POPCNT` instruction on modern microprocessors returns the binary sum of digits of an integer $n \in \{0, 2^{64} - 1\}$.

The binary sum-of-digits function

To simplify things, we consider the *binary sum-of-digits function* s. The integer $s(n)$ is the minimal number of powers of 2 needed to write n as their sum.

n	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
$s(n)$	0	1	1	2	1	2	2	3	1	2	2	3	2	3	3	4

The `POPCNT` instruction on modern microprocessors returns the binary sum of digits of an integer $n \in \{0, 2^{64} - 1\}$.

Cusick's conjecture

Our first question is a special case of the main theme "What happens to the binary expansion in the addition $n + t$?".

Does the sum of digits usually increase when a constant is added?

T. W. Cusick conjectured that $c_t > 1/2$, where c_t is the asymptotic density of natural numbers n such that $s(n+t) \geq s(n)$.

This conjecture is (surprisingly!) difficult and open since its introduction in 2011. It derives from the more general conjecture by Tu and Deng, which has its origins in cryptography.

Cusick's conjecture

Our first question is a special case of the main theme "What happens to the binary expansion in the addition $n+t$?".

Does the sum of digits usually increase when a constant is added?

T. W. Cusick conjectured that $c_t > 1/2$, where c_t is the asymptotic density of natural numbers n such that $s(n+t) \geq s(n)$.

This conjecture is (surprisingly!) difficult and open since its introduction in 2011. It derives from the more general conjecture by Tu and Deng, which has its origins in cryptography.

Cusick's conjecture

Our first question is a special case of the main theme "What happens to the binary expansion in the addition $n+t$?".

Does the sum of digits usually increase when a constant is added?

T. W. Cusick conjectured that $c_t > 1/2$, where c_t is the asymptotic density of natural numbers n such that $s(n+t) \ge s(n)$.

This conjecture is (surprisingly!) difficult and open since its introduction in 2011. It derives from the more general conjecture by Tu and Deng, which has its origins in cryptography.

Examples

Setting $t = 1$ we see that

$$\big(s(n+t) - s(n)\big)_n = \big(1, 0, 1, -1, 1, 0, 1, -2, 1, 0, 1, -1, 1, 0, 1, -3, \ldots\big),$$

which is nonnegative in 3 out of 4 cases. That is, $c_1 = 3/4$.

More values:

$$c_3 = 11/16, \quad c_{999} = 37561/2^{16},$$

$$\min_{t \le 2^{30}} c_t = 18169025645289/2^{45} = 0.516\ldots.$$

Examples

Setting $t = 1$ we see that

$$\big(s(n+t) - s(n)\big)_n = \big(1, 0, 1, -1, 1, 0, 1, -2, 1, 0, 1, -1, 1, 0, 1, -3, \ldots\big),$$

which is nonnegative in 3 out of 4 cases. That is, $c_1 = 3/4$.

More values:

$$c_3 = 11/16, \quad c_{999} = 37561/2^{16},$$

$$\min_{t \le 2^{30}} c_t = 18169025645289/2^{45} = 0.516\ldots.$$

Examples

Setting $t = 1$ we see that

$$\big(s(n+t) - s(n)\big)_n = \big(1, 0, 1, -1, 1, 0, 1, -2, 1, 0, 1, -1, 1, 0, 1, -3, \ldots\big),$$

which is nonnegative in 3 out of 4 cases. That is, $c_1 = 3/4$.

More values:

$$c_3 = 11/16, \quad c_{999} = 37561/2^{16},$$
$$\min_{t \le 2^{30}} c_t = 18169025645289/2^{45} = 0.516\ldots.$$

Examples

Setting $t = 1$ we see that

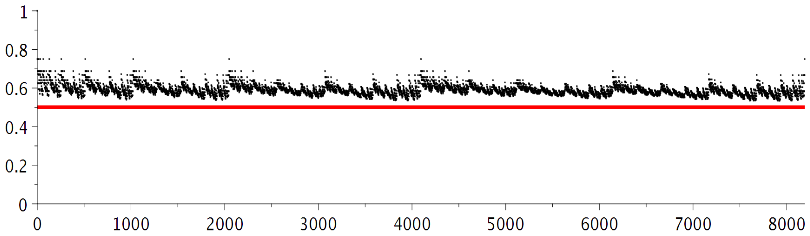
$$\big(s(n+t) - s(n)\big)_n = \big(1, 0, 1, -1, 1, 0, 1, -2, 1, 0, 1, -1, 1, 0, 1, -3, \ldots\big),$$

which is nonnegative in 3 out of 4 cases. That is, $c_1 = 3/4$.

More values:

$$c_3 = 11/16, \quad c_{999} = 37561/2^{16},$$

$$\min_{t \le 2^{30}} c_t = 18169025645289/2^{45} = 0.516\ldots.$$



An “almost solution” to the conjecture

Theorem (S.–Wallner 2021+)

Assume that the positive integer t has at least M blocks of ones in its binary expansion (where M is an absolute, effective constant). Then $c_t > 1/2$.

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

An "almost solution" to the conjecture

Theorem (S.–Wallner 2021+)

Assume that the positive integer t has at least M blocks of ones in its binary expansion (where M is an absolute, effective constant). Then $c_t > 1/2$.

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

SW2021 in a nutshell

Apart from a small set of exceptions $t \in \mathbb{N}$, the following is true.

> *The binary sum of digits, more often than not, (weakly) increases when a constant t is added.*

The difference $s(n+t) - s(n)$ basically gives the number of *carries* that appear in the addition $n + t$ in binary. Moreover, we have

$$s(n+t) - s(n) = s(t) - \nu_2\left(\binom{n+t}{t}\right),$$

where $\nu_2(m) = \max\{k \geq 0 : 2^k \mid m\}$.

SW2021 in a nutshell

Apart from a small set of exceptions $t \in \mathbb{N}$, the following is true.

The binary sum of digits, more often than not, (weakly) increases when a constant t is added.

The difference $s(n+t) - s(n)$ basically gives the number of *carries* that appear in the addition $n+t$ in binary. Moreover, we have

$$s(n+t) - s(n) = s(t) - \nu_2\left(\binom{n+t}{t}\right),$$

where $\nu_2(m) = \max\{k \geq 0 : 2^k \mid m\}$.

SW2021 in a nutshell

Apart from a small set of exceptions $t \in \mathbb{N}$, the following is true.

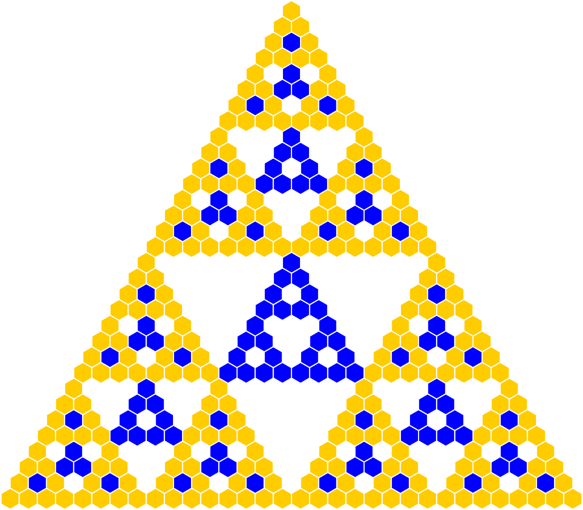
The binary sum of digits, more often than not, (weakly) increases when a constant t is added.

The difference $s(n+t) - s(n)$ basically gives the number of *carries* that appear in the addition $n + t$ in binary. Moreover, we have

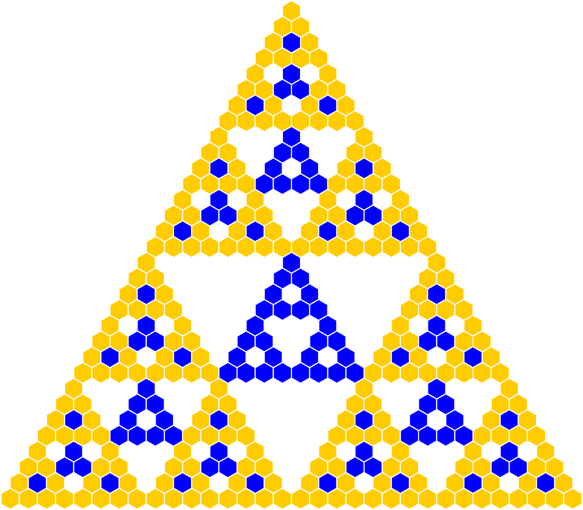
$$s(n+t) - s(n) = s(t) - \nu_2\left(\binom{n+t}{t}\right),$$

where $\nu_2(m) = \max\{k \geq 0 : 2^k \mid m\}$.

The 2-valuation of binomial coefficients



The 2-valuation of binomial coefficients



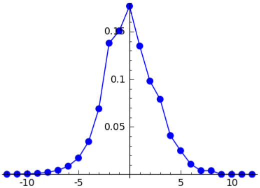
Rule: "Put a discrete Sierpiński triangle of the next color and of maximal size into each triangular hole."

A normal distribution

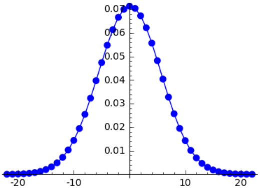
Theorem (S.–Wallner 2021+)

Let $t \geq 0$. The probability mass function δ_t on $\mathbb{Z}$ defined by the differences $s(n+t) - s(n)$ uniformly approaches a Gaussian as the number of blocks of ones in t grows.

In other words, a normal distribution can be found in the number of carries appearing in binary addition.



$t = 999$



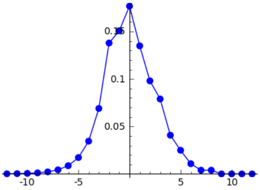
$t = 10^{23} - 1$

A normal distribution

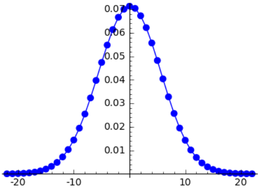
Theorem (S.–Wallner 2021+)

Let $t \geq 0$. The probability mass function δ_t on $\mathbb{Z}$ defined by the differences $s(n+t) - s(n)$ uniformly approaches a Gaussian as the number of blocks of ones in t grows.

In other words, a normal distribution can be found in the number of carries appearing in binary addition.



$t = 999$



$t = 10^{23} - 1$

The Thue–Morse sequence

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

$$\mathsf{tm} = 01101001100101101001011001101001\ldots$$

In many CPUs, the *parity flag* gives the first 2^8 terms of this sequence. The sequence tm 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*

$$\mathsf{tm} = 01101001100101101001011001101001\ldots$$

In many CPUs, the *parity flag* gives the first 2^8 terms of this sequence.
The sequence tm 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*

$$\mathsf{tm} = 01101001100101101001011001101001\ldots$$

In many CPUs, the *parity flag* gives the first 2^8 terms of this sequence. The sequence tm is an *automatic sequence* and as such can be defined via a *uniform morphism on a finite alphabet*: Let us define

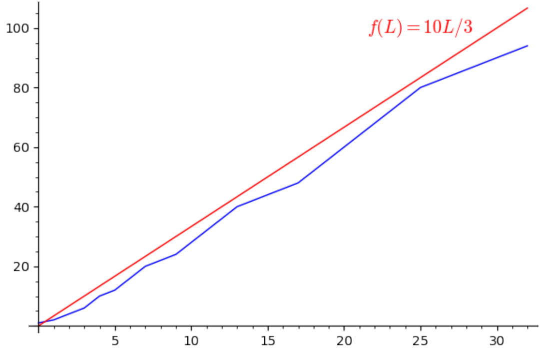
$$\varphi : 0 \mapsto 01, \quad 1 \mapsto 10.$$

Starting with 0, we obtain

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

The factor complexity of tm

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



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

The arithmetic complexity of tm

This situation changes completely when we consider subsequences of tm. First, *arithmetic subsequences*, corresponding to repeated addition of t.

Avgustinovich, Fon-Der-Flaass, and Frid (2003) proved that *every* finite sequence $A \in \{0,1\}^L$ appears as an arithmetic subsequence of tm!

$$\text{“The Thue–Morse sequence has} \left\{ \begin{array}{c} \text{low factor complexity} \\ \text{full arithmetic complexity} \end{array} \right\} \text{”}$$

The arithmetic complexity of tm

This situation changes completely when we consider subsequences of tm. First, *arithmetic subsequences*, corresponding to repeated addition of t.

Avgustinovich, Fon-Der-Flaass, and Frid (2003) proved that *every* finite sequence $A \in \{0,1\}^L$ appears as an arithmetic subsequence of tm!

$$\text{“The Thue–Morse sequence has} \left\{ \begin{array}{c} \text{low factor complexity} \\ \text{full arithmetic complexity} \end{array} \right\}\text{”}$$

Construction of a normal number

Together with Müllner we generalized this result in a quantitative way. Very roughly speaking, every sequence $A \in \{0,1\}^L$ is found with (almost) the same frequency 2^{-L} as a factor of most arithmetic subsequences of tm. This allowed us to prove the following result.

Theorem (Müllner–S. 2017)

Let $1 < c < 3/2$. Then the sequence B defined by

$$n \mapsto \mathrm{tm}\bigl(\lfloor n^c \rfloor\bigr)$$

is normal, meaning that every finite sequence $A \in \{0,1\}^L$ appears in B with asymptotic density 2^{-L}.

Very sparse arithmetic subsequences of tm

We know that arbitrarily long sequences of 0s appear as arithmetic subsequences of tm. However, for "most" arithmetic subsequences A, the number of 0s and 1s will be balanced.

Theorem (S. 2020)

The Thue–Morse sequence has level of distribution 1.

Without taking care of the details, this theorem states the following.

> *For all $\rho > 0$, most arithmetic subsequences A of* tm *having N elements and common difference $\asymp N^\rho$ have about the same number of* 0*s and* 1*s.*

Very sparse arithmetic subsequences of tm

We know that arbitrarily long sequences of 0s appear as arithmetic subsequences of tm. However, for "most" arithmetic subsequences A, the number of 0s and 1s will be balanced.

Theorem (S. 2020)

The Thue–Morse sequence has level of distribution 1.

Without taking care of the details, this theorem states the following.

> *For all $\rho > 0$, most arithmetic subsequences A of* tm *having N elements and common difference $\asymp N^\rho$ have about the same number of 0s and 1s.*

Possible extensions and open problems

Prove that $\mathrm{tm}(\lfloor n^c \rfloor)$ defines a normal sequence for all $c \in (1, 2)$.

Study the sum of digits in different bases: For $\alpha, \beta \in \mathbb{R} \setminus \mathbb{Z}$, the function $n \mapsto \alpha s_2(n) + \beta s_3(n)$ should have level of distribution 1. Such a result can be used for obtaining theorems on prime numbers in different bases.

Prove that there are infinitely many integers n such that $s_2(n) = s_3(n)$.

Prove that the Thue–Morse sequence along n^3 attains 0 and 1 with frequency 1/2 each (n^2: Mauduit–Rivat, Acta Math. 2009)

Possible extensions and open problems

Prove that $\mathrm{tm}(\lfloor n^c \rfloor)$ defines a normal sequence for all $c \in (1, 2)$.

Study the sum of digits in different bases: For $\alpha, \beta \in \mathbb{R} \setminus \mathbb{Z}$, the function $n \mapsto \alpha s_2(n) + \beta s_3(n)$ should have level of distribution 1. Such a result can be used for obtaining theorems on prime numbers in different bases.

Prove that there are infinitely many integers n such that $s_2(n) = s_3(n)$.

Prove that the Thue–Morse sequence along n^3 attains 0 and 1 with frequency 1/2 each (n^2: Mauduit–Rivat, Acta Math. 2009)

Possible extensions and open problems

☕ Prove that $\mathrm{tm}(\lfloor n^c \rfloor)$ defines a normal sequence for all $c \in (1, 2)$.

☕☕ Study the sum of digits in different bases: For $\alpha, \beta \in \mathbb{R} \setminus \mathbb{Z}$, the function $n \mapsto \alpha s_2(n) + \beta s_3(n)$ should have level of distribution 1. Such a result can be used for obtaining theorems on prime numbers in different bases.

☕☕☕ Prove that there are infinitely many integers n such that $s_2(n) = s_3(n)$.

☕☕☕☕ Prove that the Thue–Morse sequence along n^3 attains 0 and 1 with frequency 1/2 each (n^2: Mauduit–Rivat, Acta Math. 2009)

Possible extensions and open problems

Prove that $\mathrm{tm}(\lfloor n^c \rfloor)$ defines a normal sequence for all $c \in (1, 2)$.

Study the sum of digits in different bases: For $\alpha, \beta \in \mathbb{R} \setminus \mathbb{Z}$, the function $n \mapsto \alpha s_2(n) + \beta s_3(n)$ should have level of distribution 1. Such a result can be used for obtaining theorems on prime numbers in different bases.

Prove that there are infinitely many integers n such that $s_2(n) = s_3(n)$.

Prove that the Thue–Morse sequence along n^3 attains 0 and 1 with frequency $1/2$ each (n^2: Mauduit–Rivat, Acta Math. 2009)

Arithmetic subsequences of s

Another generalization of the statement "every finite sequence in $\{0,1\}$ appears as an arithmetic subsequence of tm" is the following.

Theorem (S.–Stoll 2020)

Let $k_1,\ldots,k_L$ be integers. There exists an arithmetic progression $(a_0,\ldots,a_L)$ in $\mathbb{N}$ such that for all $1\le \ell\le L$,

$$s(a_\ell)-s(a_0)=k_\ell.$$

For example,

$$\begin{aligned} s(n+t)-s(n) &= 1,\\ s(n+2t)-s(n) &= 2,\\ s(n+3t)-s(n) &= 3,\\ s(n+4t)-s(n) &= 4,\\ s(n+5t)-s(n) &= -2, \end{aligned}$$

for $n=242$ and $t=387$.

Arithmetic subsequences of s

Another generalization of the statement "every finite sequence in $\{0,1\}$ appears as an arithmetic subsequence of tm" is the following.

Theorem (S.–Stoll 2020)

Let $k_1,\ldots,k_L$ be integers. There exists an arithmetic progression $(a_0,\ldots,a_L)$ in $\mathbb{N}$ such that for all $1\le \ell\le L$,

$$s(a_\ell)-s(a_0)=k_\ell.$$

For example,

$$\begin{aligned} s(n+t)-s(n) &= 1,\\ s(n+2t)-s(n) &= 2,\\ s(n+3t)-s(n) &= 3,\\ s(n+4t)-s(n) &= 4,\\ s(n+5t)-s(n) &= -2, \end{aligned}$$

for $n=242$ and $t=387$.

Possible extensions

Study the asymptotic density of integers n such that

$$\begin{aligned} s(n+t)-s(n) &= k_1 \\ s(n+2t)-s(n) &= k_2 \\ &\dots \\ s(n+Lt)-s(n) &= k_L \end{aligned}$$

and prove multidimensional generalizations of Cusick's conjecture and the limit law.

Possible conjectures involve multidimensional Gaussians and tuples $(s(n+\ell t))_{0\le\ell\le L}$ in certain quadrants, octants,. . . .

Other digital expansions

Mauduit and Rivat proved (in particular) that there exist infinitely many prime numbers p such that $\mathrm{tm}(p) = 0$. (Ann. of Math. 2010)
It has been a long standing question to prove such a result for the *Zeckendorf sum-of-digits function*. More precisely: every nonnegative integer n can be written as a sum of Fibonacci numbers; the minimal number of summands needed is the Zeckendorf sum-of-digits $Z(n)$.

$$83 = 55 + 21 + 5 + 2$$

In a forthcoming paper with Drmota and Müllner we prove the following results.

Theorem (Drmota–Müllner–S. 2021+)

The function Z evaluated on prime numbers is uniformly distributed in residue classes.

Other digital expansions

Mauduit and Rivat proved (in particular) that there exist infinitely many prime numbers p such that $\mathrm{tm}(p) = 0$. (Ann. of Math. 2010)
It has been a long standing question to prove such a result for the *Zeckendorf sum-of-digits function*. More precisely: every nonnegative integer n can be written as a sum of Fibonacci numbers; the minimal number of summands needed is the Zeckendorf sum-of-digits $Z(n)$.

$$83 = 55 + 21 + 5 + 2$$

In a forthcoming paper with Drmota and Müllner we prove the following results.

Theorem (Drmota–Müllner–S. 2021+)

The function Z evaluated on prime numbers is uniformly distributed in residue classes.

Work in preparation, continued

Theorem (DMS 2021+)

If k is greater than some absolute bound (which can be stated explicitly), then there is a prime number p that is the sum of k different Fibonacci numbers.

Generalizations of the method of proof are possible for other numeration systems, e.g. β-expansions, rational base number systems,. . . .

Work in preparation, continued

Theorem (DMS 2021+)

If k is greater than some absolute bound (which can be stated explicitly), then there is a prime number p that is the sum of k different Fibonacci numbers.

Generalizations of the method of proof are possible for other numeration systems, e.g. β-expansions, rational base number systems,. . . .

Thank you!

0 Supported by the Austrian Science Fund (FWF), Projects F55 and MuDeRa (jointly with ANR).