Collisions of digit sums in different bases

Lukas Spiegelhofer

MONTAN UNIVERSITÄT

March 21, 2022, Rencontres de théorie analytique et élémentaire des nombres

The setup

Let $q \geq 2$ be an integer and write

$$n = \delta_0 q^0 + \delta_1 q^1 + \delta_2 q^2 + \cdots + \delta_{L-1} q^{L-1},$$

where $(\delta_j)_{0 \leq j < L} \in \{0, \ldots, q-1\}^L$ and $(L = 0$ or $\delta_{L-1} \neq 0)$. The base-q sum of digits of n is defined by

$$\mathsf{s}_q(n) \coloneqq \sum_{j \in \mathbb{N}} \delta_j.$$

In this talk, we are interested in the pairs

$$\bigl(\mathsf{s}_2(n), \mathsf{s}_3(n)\bigr).$$

In particular, we prove that $\mathsf{s}_2(n) = \mathsf{s}_3(n)$ infinitely often, thereby settling a folklore conjecture.

Theorem

There are infinitely many positive integers n such that $\mathsf{s}_2(n) = \mathsf{s}_3(n)$. *More precisely, for all* $\delta > 0$ *we have*

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

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

Let us call a natural number n such that $\mathsf{s}_2(n) = \mathsf{s}_3(n)$ a *collision* of the binary and ternary sum-of-digits functions.

The starting point for the present paper is an article by Deshouillers, Habsieger, Laishram, and Landreau.

> *"[...] it seems to be unknown whether there are infinitely many integers n for which $s_2(n) = s_3(n)$ or even for which $|s_2(n) - s_3(n)|$ is significantly small."*

They give an answer concerning the second part.

Theorem (DHLL)

For sufficiently large N, we have

$$\#\{n \le N : |s_3(n) - s_2(n)| \le 0.1457205 \log n\} > N^{0.970359}.$$

Note that the difference $s_3(n) - s_2(n)$ is expected to have a value around $C \log n$, where

$$C = \frac{1}{\log 3} - \frac{1}{\log 4} = 0.18889\ldots.$$

The question on the infinitude of collisions is not a new one. M. Drmota received a hand-written letter from A. Hildebrand more than 20 years ago, in which the same question was asked.

We also note a result by de la Bretèche, Stoll, and Tenenbaum (2019): they proved that the sequence

$$n \mapsto \frac{\mathsf{s}_a(n)}{\mathsf{s}_b(n)}$$

is dense in $[0, \infty)$, if a and b are multiplicatively independent bases. In particular, for all $\varepsilon > 0$,

$$\bigl|\mathsf{s}_a(n) - \mathsf{s}_b(n)\bigr| \le \varepsilon \mathsf{s}_b(n) \le 2\varepsilon \log_2(n)$$

for infinitely many integers n.

Powers of 2 and 3

Mixing different bases is a source of difficult problems, cf. conjectures of Furstenberg. ("Expansions of real numbers in multiplicative independent bases look very different".)
We give an idea of the occurring problems by considering the set of all powers of 2 and 3, sorted in ascending order:

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

This is sequence `A006899` in Sloane's OEIS. How are the powers of 2 and 3 arranged?
In between powers of 3, there are one or two powers of 2. More precisely, we take logarithms in base 2.

The powers of 2 and 3 are arranged in the same way as multiples of $\alpha = \log 3/\log 2$ and 1, that is, we are interested in the Sturmian word

$$\bigl(\lfloor (n+1)\alpha\rfloor - \lfloor n\alpha\rfloor\bigr)_{n\ge 0} = (1,2,1,2,1,2,2,1,2,1,\ldots).$$

This sequence gives the number of powers of 2 in between powers of 3. This word can be described by the *continued fraction expansion* of α,

$$\alpha = [1;1,1,2,2,3,1,5,2,23,2,2,1,1,55,1,4,3,1,1,\ldots],$$

which is not even known to be bounded (i.e. whether α is *badly approximable*).

Understanding the ternary expansion of powers of two is a proper extension of the above problem, and even harder to understand. A conjecture of Erdős states that 1, 4, and 256 are the only powers of two that can be written as a sum of pairwise different powers of three.

power of 2	ternary expansion
2^0	11
2^1	2
2^2	1111
2^3	22
2^4	121
2^5	1012
2^6	2101
2^7	11202
2^8	100111100111
2^9	200222
2^{10}	1101221

Erdős: "[. . .] as far as I can see, there is no method at our disposal to attack this conjecture"

- The difficulty in proving our theorem lies in the separation of the values of $\mathsf{s}_2(n)$ and $\mathsf{s}_3(n)$.
- The sum-of-digits functions can be thought of as a sum of i.i.d. random variables, and they concentrate around the values $\log_4 N$ and $\log_3 N$ respectively, where $0 \le n < N$.
- By Hoeffding's inequality, the tails of these distributions decay as least as fast as $\exp(-C(x-\mu)^2/\sigma^2)$, where μ is the expected value, and $\sigma^2 \asymp \log N$ the variance.
- Since the gap $(1/\log 3 - 1/\log 4)\log N$ is as large as $\asymp (\log N)^{1/2}$ standard deviations, we can only expect a number $\ll N^\delta$ of collisions, where $\delta < 1$ is some constant.
- In the light of this argument, we see that our result (the number of collisions is $\gg N^\eta$) cannot be too far from the true number of collisions.

Candidates for $s_2(n) = s_3(n)$

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

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

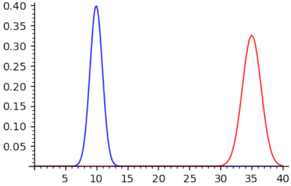
while the binary digits of $3^\zeta n$ should be "random". We therefore expect

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

Let us choose

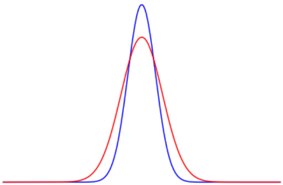
$$\zeta \approx \frac{\log(n)}{\log 3}\left(\frac{\log 4}{\log 3} - 1\right).$$

Then $s_2(3^\zeta n)$ and $s_3(3^\zeta n)$ should concentrate around the same expected value. We will look for collisions in the class $3^\zeta\mathbb{N}$!



all n

$\rightarrow$



only n in a residue class

Sketch of proof

Let us write $f(n) = s_2(n) - s_3(n)$. The search for collisions will consist of three main steps.

1. **"Preparation".** Find a residue class A, and certain shifts $d_j \in \mathbb{Z}$ and correction terms $\xi_j \in \{0, 1\}$ for $-J \le j \le J$, such that

$$f(n + d_j) - f(n) = jm + \xi_j \quad \text{for all } j \in \{-J, -J + 1, \ldots, J\}.$$

2. **"Rarefaction".** We rarefy and truncate the class A, obtaining an arithmetic progression $A' \subseteq A \cap [N, 2N)$, in such a way that $|f(n)| \le Jm$ for most $n \in A'$.
3. **"Selection".** Select only those $n \in A'$ such that $f(n) \in m\mathbb{Z}$.

The common difference m will be smaller than the standard deviations of the sum-of-digits functions, by a small factor (the *fineness*):

$$m \asymp (\log N)^{1/2}/f \quad \text{where} \quad f = (\log\log N)^{1/2+\varepsilon}.$$

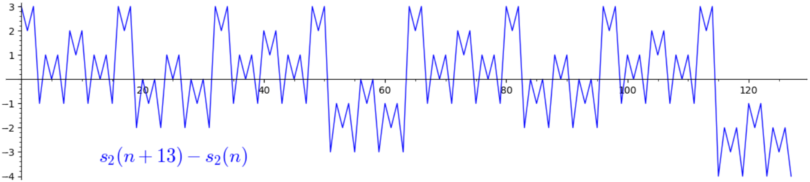
Moreover, we have $J = f^2$, so that $[-Jm, Jm]$ covers many standard deviations of the sum-of-digits functions.

Constant differences of the sum-of-digits function

- Recall the ruler sequence $n \mapsto \nu_2(n+1)$: we have

$$\mathsf{s}_2(n+1) - \mathsf{s}_2(n) = 1 - \nu_2(n+1).$$

  Therefore $\mathsf{s}_2(n+1) - \mathsf{s}_2(n) = m$ for $m \le 1$ and $n \in 2^{1-m} - 1 + 2^{2-m}\mathbb{Z}$, as we had before (in the first talk).
- More generally, $\mathsf{s}_q(n+d) - \mathsf{s}_q(n)$ can be decomposed into constant parts, which are arithmetic progressions modulo q^λ (Bésineau 1972).



By an argument containing an intersection of residue classes,

$$\bigl(a+2^{\nu}\mathbb{N}\bigr)\cap\bigl(K+3^{\beta}\mathbb{N}\bigr)=L+2^{\nu}3^{\beta}\mathbb{N},$$

we arrive at the following technical proposition, capturing the "preparation" step.

Proposition

Let $\eta := \lfloor(\log N)^{3/4}\rfloor$, $\beta := (2J+1)\eta+1$, *and* $2^{\nu-1}\geq 3^{\beta}$. *Set*

$$d_j := \bigl(1^{(j+1+J)\eta}0\bigr)_3.$$

There exists $L\in\{0,\ldots,2^{\nu}3^{\beta}-1\}$ *such that* $L\equiv 9 \bmod 12$, *and* $\xi_j\in\{0,1\}$ *for* $-J\leq j\leq J$ *such that*

$$f(n+d_j)-f(n)=jm+\xi_j \qquad \begin{array}{ll}\textit{for all} & j\in\{-J,\ldots,J\}\\ \textit{and all} & n\in A:=L+2^{\nu}3^{\beta}\mathbb{N}.\end{array}$$

The second step of the proof

- Now the heuristic argument comes into play. We set

$$A' = \left(L + 2^\nu 3^{\beta+\zeta}\mathbb{N}\right) \cap [N, 2N),$$

where

$$\zeta \sim \log(n)\left(1/\log 3 - 1/\log 4\right).$$

- Set $\alpha = \beta + \zeta$ and $b_2 = \lfloor 2^{-\nu} L \rfloor < 3^\alpha$. We consider

$$\mathsf{s}_2\left(b_2 + 3^\alpha k\right) = s_2\left(\left\lfloor k\frac{3^\alpha}{2^{\kappa_2}} + \sigma\right\rfloor\right) + s_2\left(\left(b_2 + 3^\alpha k\right) \bmod 2^{\kappa_2}\right),$$

where $\kappa_2 = \min\{m : 2^m \geq 3^\alpha\}$ and $\sigma = b_2 2^{-\kappa_2} < 1$.

- Each summand satisfies a concentration property similar to the binomial distribution, and analogously for s_3.
- The exponent ζ is chosen in such a way that the concentration of $f(n) = \mathsf{s}_2(n) - \mathsf{s}_3(n)$, where $n \in A'$, happens close to zero!

The content of "rarefaction" step is the following: we have

$$|\mathsf{s}_2(n) - \mathsf{s}_3(n)| \le Jm \asymp (\log N)^{1/2} (\log\log N)^{1/2+\varepsilon}$$

for most $n \in A' = \bigl(L + 2^{\nu} 3^{\beta+\zeta}\mathbb{N}\bigr) \cap [N, 2N)$.

This already gives us a sharpening of the DHLL-result, and of part of the dlBST-result.

Since A' is an arithmetic progression, we may expect that an exponential sum estimate (à la Mauduit–Rivat) yields a statement on $\mathsf{s}_2(n) - \mathsf{s}_3(n) \in m\mathbb{Z}$, where n runs through A'.

The third step: $f(n) \in m\mathbb{Z}$

Proposition

Let

$$P := \#\{n \in A' : f(n) \in m\mathbb{Z}\}.$$

As $N \to \infty$, we have

$$P = \frac{|A'|}{m}\bigl(1 + o(1)\bigr).$$

The residue class $m\mathbb{Z}$ receives the ratio $(\log N)^{-1/2}(\log\log N)^{1/2+\varepsilon}$ of the values of $f(n) = s_2(n) - s_3(n)$ along the finite arithmetic progression A'.

The proof of this statement contains two ingredients:

1. Applying van der Corput's inequality

$$\left|\sum_{n\in I} z_n\right|^2 \le \frac{|I|+R-1}{R} \sum_{-R<r<R} \left(1-\frac{|r|}{R}\right) \sum_{n\in I\cap(I-r)} z_n\overline{z_{n+r}}$$

to

$$\begin{aligned} z_k &:= \mathrm{e}\left(\frac{\ell}{m}\left(\mathrm{s}_2\left(L+2^\nu 3^{\beta+\zeta}k\right)-\mathrm{s}_3\left(L+2^\nu 3^{\beta+\zeta}k\right)\right)\right) \\ &= x\,\mathrm{e}\left(\frac{\ell}{m}\left(\mathrm{s}_2\left(L'+3^{\beta+\zeta}k\right)-\mathrm{s}_3\left(L''+2^\nu k\right)\right)\right), \end{aligned}$$

where $|x|=1$, we may discard all but $\sim \log_2\left(3^\zeta\right)$ binary digits, and all but a small number of ternary digits.

2. By the Chinese remainder theorem, we may replace the sum over arithmetic progressions by full sums. This allows us to prove an upper bound for the exponential sums for the case $\ell/m \notin Z$, and the proposition follows.

Putting together the pieces

The second step consisted in showing that $|f(n)| \leq Jm$ for *most* $n \in A'$. In fact, there will only be $\ll |A'|(\log N)^{-D}$ exceptional $n \in A'$, for each $D > 0$.

By our third step, the number of integers $n \in A'$ such that $f(n) \in m\mathbb{Z}$ is $|A'|(\log N)^{-1/2}(\log\log N)^{1/2+\varepsilon}\,(1 + o(1))$.

Choosing $D > 1/2$, we see that there are many non-exceptional $n \in A'$ such that $f(n) \in m\mathbb{Z}$. That is,

$$\mathsf{s}_2(n) - \mathsf{s}_3(n) \in \{-Jm, (-J+1)m, \ldots, Jm\}$$

has many solutions $n \in A'$.

Now we can apply our preparation (first step). We have

$$f(n+d_j)-f(n)=jm+\xi_j \qquad \begin{array}{ll} \text{for all} & j\in\{-J,\ldots,J\} \\ \text{and all} & n\in A=L+2^\nu 3^\beta\mathbb{N}. \end{array}$$

For each $n\in A'$ such that $f(n)\in\{-Jm,(-J+1)m,\ldots,Jm\}$, we choose $j(n)\in\{-J,\ldots,J\}$ such that

$$f(n+d_{j(n)})\in\{0,1\}.$$

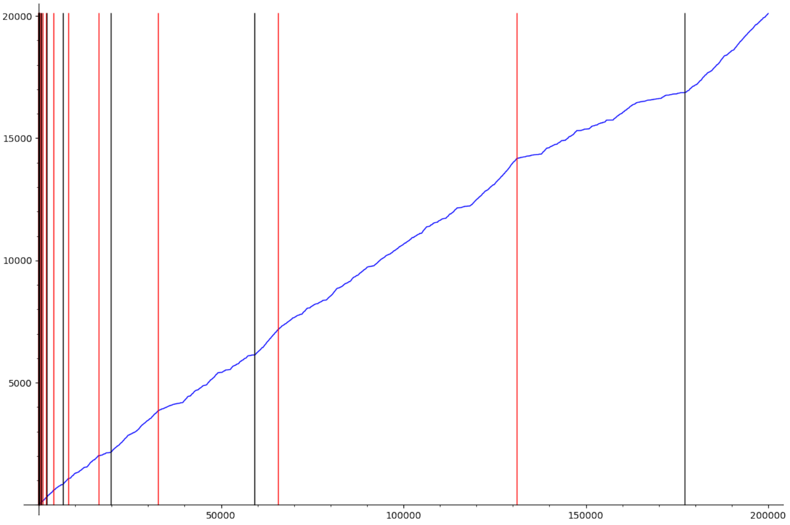
Finally, we use the condition $L\equiv 9 \bmod 12$ in order to eliminate the unpleasant correction term $\xi_j\in\{0,1\}$.

Eliminating the correction term

Set $A = \log 3/\log 4 - \delta$. If the number of solutions of $f(n + d_{j(n)}) = 0$ with $n < N$ is $\gg N^A$, we are done. Otherwise there is a number $\gg N^A$ of solutions $n \in (9 + 12\mathbb{Z}) \cap [N, 2N)$ of $f(n) = 1$.
We have

$$n \equiv 9 \bmod 12 \Leftrightarrow (n \equiv 0 \bmod 3 \text{ and } n \equiv 1 \bmod 4),$$

and therefore $\mathrm{s}_3(n+1) = \mathrm{s}_3(n) + 1$ and $\mathrm{s}_2(n+1) = \mathrm{s}_2$. We obtain $\gg N^A$ solutions of $f(n+1) = 0$.



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

Open problems

1. Find a construction method for collisions.
2. Prove that there is a base $q \geq 2$ and infinitely many prime numbers p such that
$$s_q(p) = s_{q+1}(p).$$
3. Prove or disprove the asymptotic formula
$$\#\{n < N : s_2(n) = s_3(n)\} \sim cN^\eta \tag{2}$$
for some real constants c and η.

4. Prove an asymptotic formula (in k) for the number of solutions of the equation
$$2^{\mu_1} + \cdots + 2^{\mu_k} = 3^{\nu_1} + \cdots + 3^{\nu_k}, \tag{3}$$
and for the numbers
$$\#\{n \in \mathbb{N} : s_2(n) = s_3(n) = k\}$$
(finiteness in the second case was proved by Senge and Straus (1973)).
5. Consider all pairs (q_1, q_2) of multiplicatively independent bases, and arbitrary families $(q_1, \ldots, q_K)$ of pairwise coprime bases ≥ 2.
6. Study collisions of integer-valued *k-regular sequences* in coprime bases, generalizing the sum-of-digits case. For example, are there infinitely many positive integers n such that the number $|n|_{2,11}$ of blocks 11 in binary equals the number $|n|_{3,1}$ of digits 1 in ternary?

Thank you!

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