Sum-free sets

Peter J. Cameron
University of St Andrews

Topological dynamics, functional equations,
infinite combinatorics and probability
LSE, June 2017

Three theorems

A set of natural numbers is k-AP-free if it contains no k-term arithmetic progression, and is sum-free if it contains no solution to $x + y = z$.

Theorem (van der Waerden)

For a natural number $k > 2$, the set $\mathbb{N}$ cannot be partitioned into finitely many k-AP-free sets.

Theorem (Roth–Szemerédi)

For a natural number $k > 2$, a k-AP-free set has density zero.

Theorem (Schur)

The set $\mathbb{N}$ cannot be partitioned into finitely many sum-free sets.

The missing fourth?

At first sight, one would like a theorem to complete the pattern, asserting that a sum-free set has density zero.

However, this is false; the set of all odd numbers is sum-free and has density $1/2$.

What follows can be regarded as an attempt to find a replacement for the missing fourth theorem.

Note that other people have tried to fill this gap in different ways …

A bijection with Cantor space

The *Cantor space* $\mathfrak{C}$ can be represented as the set of all (countable) sequences of zeros and ones.

It carries the structure of a complete metric space (the distance between two sequences is a monotonic decreasing function of the index of the first position where they differ) or as a probability space (corresponding to a countable sequence of independent tosses of a fair coin).

We define a bijection between Cantor space and the set $\mathfrak{S}$ of all sum-free subsets of $\mathbb{N}$. Given a sequence $x \in \mathfrak{C}$, we construct S as follows:

> Consider the natural numbers in turn. When considering n, if n is the sum of two elements already put in S, then of course $n \notin S$. Otherwise, look at the first unused element of x; if it is 1, then put $n \in S$, otherwise, leave n out of S. Delete this element of the sequence and continue.

An example

- **Input:** 101…
- **Output:** 1,4,…
- 1 is not a sum, so take an input bit; it's 1, so $1 \in S$
- $2 = 1 + 1$, so skip
- 3 is not a sum, so take an input bit; it's 0, so $3 \notin S$
- 4 is not a sum, so take an input bit; it's 1, so $4 \in S$
- $5 = 1 + 4$, so skip
- … and so on …

Baire category

$\mathfrak{C}$ is a complete metric space, where the distance between two sequences is a monotone decreasing function of the first position where they disagree.

The notion of "almost all" in a complete metric space is a residual set; a set is residual if it contains a countable intersection of dense open sets.

Thus, residual sets are non-empty (by the Baire Category Theorem); any countable collection of residual sets has non-empty intersection; a residual set meets every non-empty open set; and so on.

Baire category in $\mathfrak{C}$

It is a straightforward exercise to show the following.

- A subset X of $\mathfrak{C}$ is open if and only if it is finitely determined, that is, for any $x \in X$, there is a finite initial segment x_0 of x such that any sequence beginning with x_0 belongs to X.
- A subset X of $\mathfrak{C}$ is dense if and only if it is always reachable, that is, any finite sequence is an initial segment of some element of X.

So a set is residual if and only if it contains a countable intersection of sets which are finitely determined and always reachable.

sf-universal sets

A sum-free set is called sf-universal if everything which is not forbidden actually occurs.
Precisely, S is sf-universal if, for every $A \subseteq \{1, \ldots, n\}$, one of the following occurs:

- there are $i, j \in A$ with $i < j$ and $j - i \in S$;
- there exists N such that $S \cap [N+1, \ldots, N+n] = N + A$,

where $N + A = \{N + a : a \in A\}$.

Properties of sf-universality

Theorem

The set of sf-universal sets is residual in $\mathfrak{S}$.

Theorem (Schoen)

A sf-universal set has density zero.

Thus our “missing fourth theorem” asserts that almost all sum-free sets (in the sense of Baire category) have density zero.

Application to Henson's graph

Henson's graph is the unique countable homogeneous universal triangle-free graph. Henson showed that it has automorphisms acting as cyclic shifts on the vertices, but the analogous K_m-free graphs for $m > 3$ do not.

Let S be an arbitrary subset of $\mathbb{N}$. We define the Cayley graph $\mathrm{Cay}(\mathbb{Z}, S)$ to have vertex set $\mathbb{Z}$, with $x \sim y$ if and only if $|y - x| \in S$. Note that this graph admits the group $\mathbb{Z}$ acting as a shift automorphism on the vertex set.

Theorem

- $\mathrm{Cay}(\mathbb{Z}, S)$ *is triangle-free if and only if* S *is sum-free.*
- $\mathrm{Cay}(\mathbb{Z}, S)$ *is isomorphic to Henson's universal homogeneous triangle-free graph if and only if* S *is sf-universal.*

So Henson's graph has uncountably many non-conjugate shift automorphisms.

Measure

In a probability space, large sets are those which have measure 1, that is, complements of null sets. Just as for Baire category, these have the properties one would expect: the intersection of countably many sets of measure 1 has measure 1; a set of measure 1 intersects every set of positive measure; and so on.

The first surprise is that measure and category give entirely different answers to what a typical set looks like:

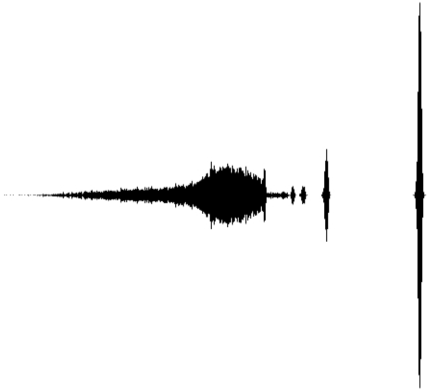
Conjecture

The set of sf-universal sets has measure zero.

Although this is not proved yet (to my knowledge), it is certain that this set does not have measure 1.

Density

Given the measure on $\mathfrak{S}$, and our interest in density, it is natural to ask about the density of a random sum-free set. This can be investigated empirically by computing many large sum-free sets and plotting their density. Here is the rather surprising result.



Sets of odd numbers

The spike on the right corresponds to density $1/4$ and is explained by the following theorem.

Theorem

- *The probability that a random sum-free set consists entirely of odd numbers is about* 0.218 *(in particular is non-zero).*
- *Conditioned on this, the density of a random sum-free set is almost surely* $1/4$.

The intuition is: Suppose we are building a random sum-free set and have so far not seen any even numbers.

- Then the odd numbers will be independent; so the next even number is very likely to be the sum of two odd numbers chosen, and so excluded.
- On the other hand, the next odd number is not a sum of two chosen numbers, so we include it with probability $1/2$.

So the pattern tends to persist.

Other positive pieces

Let $\mathbb{Z}/n\mathbb{Z}$ denote the integers modulo n. We can define the notion of a sum-free set in $\mathbb{Z}/n\mathbb{Z}$ in the obvious way. Such a sum-free set T is said to be complete if, for every $z \in \mathbb{Z}/n\mathbb{Z} \setminus T$, there exist $x, y \in T$ such that $x + y = z$ in $\mathbb{Z}/n\mathbb{Z}$. Now the theorem above extends as follows. Let $\mathfrak{S}(n, T)$ denote the set of all sum-free sets which are contained in the union of the congruence classes $t \bmod n$ for $t \in T$.

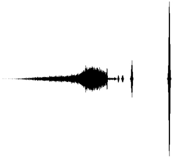
Theorem

Let T be a sum-free set in $\mathbb{Z}/n\mathbb{Z}$.

- *The probability of $\mathfrak{S}(n, T)$ is non-zero if and only if n is complete.*
- *If T is complete then, conditioned on $S \in \mathfrak{S}(n, T)$, the density of S is almost surely $|T|/2n$.*

Discrete or continuous?

The next complete modular sum-free sets are $\{2,3\}$ mod 5 and $\{1,4\}$ mod 5, and $\{3,4,5\}$ mod 8 and $\{1,4,7\}$ mod 8.



These give "spectral lines" at $1/5$ and $3/16$ in the figure, for similar reasons to the odd numbers.

The density spectrum appears to be discrete above $1/6$, and there is some evidence that this is so. However, a recent paper of Haviv and Levy shows the following result.

Theorem

The values of $|T|/2n$ for complete sum-free sets $T \subseteq \mathbb{Z}/n\mathbb{Z}$ are dense in $[0, 1/6]$.

More pieces

Neil Calkin and I showed that the event that 2 is the only even number in S has positive (though rather small) probability. More generally,

Theorem

Let A be a finite set and T a complete sum-free set modulo n. Then the event $A \subseteq S \subseteq A \cup (T \bmod n)$ has positive probability.

Question

Is it true that a random sum-free set almost surely has a density? Is it also true that the density is strictly positive almost surely? Is it even true that a random sum-free set almost surely lies in one of the pieces so far described?

Discrete or continuous? 2

Question

Is it true that the density spectrum is discrete above $1/6$ *but has a continuous part below* $1/6$*?*

The following construction is due to Calkin and Erdős. Let α be an irrational number, and define $S(\alpha)$ to be the set of natural numbers n for which the fractional part of $n\alpha$ lies between $1/3$ and $2/3$. It is easy to see that $S(\alpha)$ is sum-free and has density $1/3$.

However this does not resolve the question, since the event $S \subseteq S(\alpha)$ for some α has probability zero.

However, there might be other examples along these lines …

Counting, 1

> "I count a lot of things that theres no need to count," Cameron said. "Just because that's the way I am. But I count all the things that need to be counted."

Richard Brautigan, *The Hawkline Monster: A Gothic Western*

I will look briefly at counting, and describe two conjectures I made with Paul Erdős (both now proved, the first independently by Ben Green and Sasha Sapozhenko, the second by József Balogh, Hong Liu, Maryam Sharifzadeh, and Andrew Treglown).
These results give a picture of the typical sum-free set which disagrees with both the measure-theoretic and Baire-categoric versions.

Counting, 2

How many sum-free subsets of $\{1, \ldots, n\}$ are there?

- Any set of odd numbers is sum-free; this gives $2^{\lceil n/2 \rceil}$ sets.
- Any set with least element more than $n/2$ is sum-free; this also gives $2^{\lceil n/2 \rceil}$ sets. The intersection with the previous type is small (about $2^{n/4}$).
- We can find a few more sets by allowing numbers a bit smaller than $n/2$ and correspondingly excluding numbers close to n. Careful analysis shows that this gives about $c.2^{n/2}$ sets, where c takes one of two values depending on the parity of n.

The first type as we have seen contribute $0.218\ldots$ to the measure. The rest are invisible both to measure and to Baire category.

Cameron–Erdős, 1

Our first conjecture, now a theorem, asserted that asymptotically this accounts for everything:

Theorem

There are constants c_e and c_0 such that, if $s(n)$ is the number of sum-free subsets of $\{1, \ldots, n\}$, then

$$\frac{s(n)}{2^{n/2}} \to c_e \text{ or } c_o$$

as $n \to \infty$ through even, resp. odd, values.

The constants are approximately 6.0 and 6.8 but have not been evaluated accurately.

Cameron–Erdős, 2

This shows that most sum-free sets are contained in a very small number of maximal ones; e.g. a positive fraction are contained in the odd numbers.

> *How many maximal sum-free subsets of* $\{1, \ldots, n\}$ *are there?*

Erdős and I constructed about $2^{n/4}$ such sets and conjectured that this is the right answer. This was proved by Balogh *et al.*, who showed:

Theorem

There are constants m_0, m_1, m_2, m_3 *such that, if* $s_{\max}(n)$ *is the number of maximal sum-free subsets of* $\{1, \ldots, n\}$*, then*

$$\frac{s_{\max}(n)}{2^{n/4}} \to m_i$$

as $n \to \infty$ *through values congruent to* $i \pmod 4$.

Rationality

An infinite sequence x is rational, or ultimately periodic, if there exist positive integers n and k such that $x_{i+k} = x_i$ for all $i \geq n$. Analogously, a sum-free set S is rational if $(i \in S) \Leftrightarrow (i + k \in S)$ for all $i \geq n$.
It is easy to see that, in our bijection from sequences to sum-free sets, a sequence which maps to a rational sum-free set must itself be rational. What about the converse?

Question

Is it true that the image of a rational sequence is a rational sum-free set?

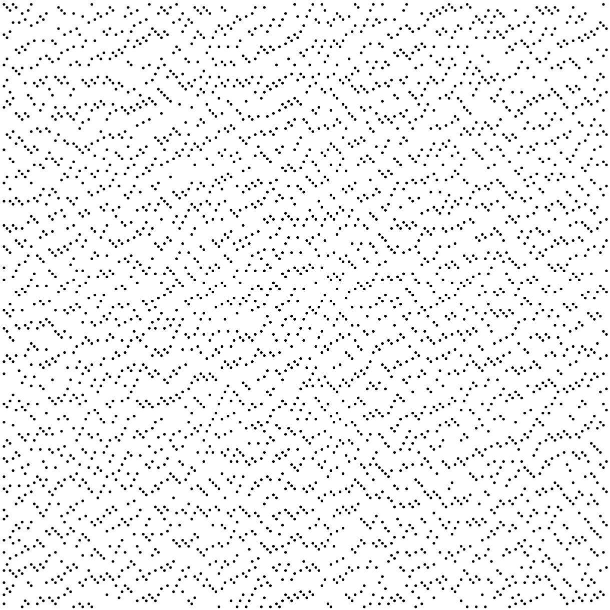
Rationality, 2

After spending some time on this question, Neil Calkin and I conjectured that the answer is “no”. There are some rational sequences (the simplest being 01010 repeated) for which the corresponding output shows no sign of periodicity despite our computing nearly a million terms.

These sets are fascinating, and seem sometimes to exhibit an “almost periodic” structure; they settle into a period, which is then broken and a longer period established, and so on.

The next slide shows the integers in $[1, 40000]$ in the sum-free set generated by 01010 repeated. Some of the structure is clearly visible in the picture.

The first 40000 numbers




THANK YOU