Notes on sum-free sets

pjc, april 2017

Abstract

Some notes from the talk on 19 April.

1 The context

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 $x + y = z$.

Three big theorems in additive combinatorics are:

Theorem 1.1 (van der Waerden) *For a natural number $k > 2$, the set $\mathbb{N}$ cannot be partitioned into finitely many k-AP-free sets.*

Theorem 1.2 (Roth–Szemerédi) *For a natural number $k > 2$, a k-AP-free set has density zero.*

Theorem 1.3 (Schur) *The set $\mathbb{N}$ cannot be partitioned into finitely many sum-free sets.*

At first sight, one would like a theorem to complete the pattern, asserting that a sum-free set has density zero. But this is false, since the set of all odd numbers is sum-free and has density $1/2$. What follows could be motivated as an attempt to find a replacement for the missing fourth theorem.

2 A bijection with Cantor space

The *Cantor space* $\mathcal{C}$ 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 $\mathcal{S}$ of all sum-free subsets of $\mathbb{N}$. Given a sequence $x \in \mathcal{C}$, 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.

For example, suppose that $x = 10110\ldots$.

- The first element of x is 1, so $1 \in S$.
- $2 = 1 + 1$, so $2 \notin S$.
- $3 \notin S + S$; the next element of x is 0, so $3 \notin S$.
- $4 \notin S + S$; the next element of x is 1, so $4 \in S$.
- $5 = 1 + 4$, so $5 \notin S$.
- $6 \notin S + S$; the next element of x is 1, so $6 \in S$.
- $\ldots$

So $S = \{1, 4, 6, \ldots\}$.

3 Baire category

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.

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

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

Theorem 3.1 *The set of sf-universal sets is residual in $\mathcal{S}$.*

Theorem 3.2 (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.

There is a nice application. 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 3.3

- $\mathrm{Cay}(\mathbb{Z}, S)$ *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.

4 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.

Given the measure on $\mathcal{S}$, 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. Figure 1 shows the rather surprising result.

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

Theorem 4.1

- *The probability that a random sum-free set consists entirely of odd numbers is about* 0.218 *(in particular is non-zero).*

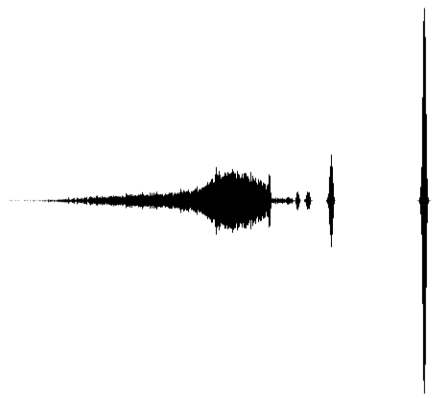

Figure 1: Density of random sum-free sets

- *Conditioned on this, the density of a random sum-free set is almost surely* $1/4$.

Other pieces can also be identified. 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 $\mathcal{S}(n, T)$ denote the set of all sum-free sets which are contained in the union of the congruence classes t mod n for $t \in T$.

Theorem 4.2 *Let T be a sum-free set in $\mathbb{Z}/n\mathbb{Z}$.*

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

In the figure it is possible to see "spectral lines" corresponding to the complete modular sum-free sets $\{2, 3\}$ mod 5 and $\{1, 4\}$ mod 5 (at density $1/5$) and $\{3, 4, 5\}$ mod 8 and $\{1, 4, 7\}$ mod 8 (at density $3/16$).

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

However, this is not the end of the story. 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 4.4 *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 further true that the density spectrum is discrete above 1/6 but has a continuous part below 1/6?

In this connection, consider the following construction of 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 ...

5 Ultimate periodicity

A sequence x is *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 *ultimately periodic* 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 an ultimately periodic sum-free set must itself be ultimately periodic. What about the converse?

Question Is it true that the image of an ultimately periodic sequence is an ultimately periodic sum-free set?

After some investigation, Neil Calkin and I conjectured that the answer is "no". There are some ultimately periodic sequences (the simplest being 01010 repeated) for which no sign of periodicity has been detected despite 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.