INTEGERS: ELECTRONIC JOURNAL OF COMBINATORIAL NUMBER THEORY 5(2) (2005), #A03

DIFFERENCE DENSITY AND APERIODIC SUM-FREE SETS

Neil J. Calkin
Department of Mathematical Sciences, Clemson University, Clemson, SC 29634, USA
calkin@clemson.edu

Steven R. Finch
Clay Mathematics Institute, One Bow Street, Cambridge, MA 02138, USA
steven.finch@inria.fr

Timothy B. Flowers
Department of Mathematical Sciences, Clemson University, Clemson, SC 29634, USA
tflower@clemson.edu

Received: 3/1/04, Accepted: 7/16/04, Published: 9/1/05

Abstract

Cameron has introduced a natural one-to-one correspondence between infinite binary sequences and sets of positive integers with the property that no two elements add up to a third. He observed that, if a sum-free set is ultimately periodic, so is the corresponding binary sequence, and asked if the converse also holds. We introduce the concept of difference density and show how this can be used to test specific sets. These tests produce further evidence of a positive nature that certain sets are, in fact, not ultimately periodic.

1. Introduction

A set S of positive integers is called sum-free if for all $x, y \in S$, $x + y \notin S$. There is a natural bijection between the set of sum-free sets and the set of infinite binary strings. Cameron showed that an ultimately periodic sum-free set always yields an ultimately periodic binary sequence. Subsequent work by Calkin and Finch suggest the converse is not true. We begin by looking at binary strings with specific structure and prove that the corresponding sum-free sets will always be periodic. Next, we define a difference density function on all sum-free sets. The results of this function provide computational evidence that aperiodic sum-free sets could have a corresponding binary string which is periodic. Finally, we present several theoretical and computational open questions related to these results.

2. Definitions

Call $S \subset \mathbb{N}$ *sum-free* if and only if $\forall\ x, y \in S,\ x+y \notin S$ (where x, y need not be distinct). Equivalently, S is sum-free if and only if $S \cap (S+S) = \emptyset$. If $\exists\ p \in \mathbb{N}$ such that for all $n \in \mathbb{N}$, $n \in S$ if and only if $n+p \in S$, then S is called *periodic* (or *purely periodic*). The minimal such p is referred to as the *period* of the sum-free set.

Recall that the symmetric difference between two sets S and T is defined by

$$S \triangle T := (S \cup T) \setminus (S \cap T).$$

If there exists a purely periodic sum-free set T such that $|S \triangle T| < \infty$, then S is called *ultimately periodic*. If p is the period of T, then p is called the *ultimate period* of S. For an ultimately periodic sum-free set S, let $n' = \max(S \triangle T)$ and $n_0 = \left\lceil \frac{n'}{p} \right\rceil p$. Then define the following:

$$\begin{aligned}
&S_* = S \cap \{1, 2, \ldots, n_0\},\\
&S_{\text{per}} := S \setminus S_* = S \cap \{n_0+1, n_0+2, \ldots\},\\
&\overline{S} := S_{\text{per}} \pmod p := \ \{q_1, q_2, \ldots, q_r\} \subseteq \{1, 2, \ldots, p-1\}.
\end{aligned}$$

We refer to S_* as the *preperiod subset* of S. Naturally, a sum-free set which is neither purely periodic nor ultimately periodic (i.e. $S_* = S$) is called *aperiodic*.

Let σ be a (one-way infinite) binary string,

$$\sigma = (\sigma_1 \sigma_2 \sigma_3 \ldots) \in 2^{\mathbb{N}}\ .$$

If there exists some $p \in \mathbb{N}$ such that $\sigma_i = \sigma_{i+p}\ \forall\ i \in \mathbb{N}$, then σ is called *periodic* (or *purely periodic*). The number p is referred to as the *period* of the sequence. Similarly if there exist $p,\ i_0 \in \mathbb{N}$ such that $\sigma_i = \sigma_{i+p}\ \forall\ i \geq i_0$, then σ is called *ultimately periodic*. p is referred to as the *ultimate period*.

3. A Bijection

Cameron [3] observed a natural bijection between sum-free sets and binary sequences. We can use a binary string σ as a *decision sequence* to determine which elements of $\mathbb{N}$ are to be included in our corresponding sum-free set. Each number in $\mathbb{N}$ is tested in order for inclusion in the set. If a number is a sum of numbers already in the set, it is automatically excluded.

Otherwise, the number is included in the set if and only if the next unused σ_i is a 1. We use this process to define the map $\theta : 2^{\mathbb{N}} \longrightarrow \{S \mid S \text{ is sum-free}\}$. Then $\theta(\sigma)$ is a sum-free set.

Now, for a sum-free set S, construct the sequence $\tau = (\tau_1\tau_2\tau_3\ldots)$ by

$$\tau_i = \begin{cases} 1 & \text{if } i \in S \\ * & \text{if } i \in S+S \\ 0 & \text{otherwise.} \end{cases}$$

Then, remove all of the $*$'s to get the decision sequence σ which corresponds to S. This process is θ^{-1}. (See [1] for a more formal definition of θ and θ^{-1}).

For example,

$$\begin{aligned} \theta(100100100100\ldots) &= \{1,5,9,13,\ldots\} \\ \theta(110011100111\ldots) &= \{1,3,8,10,12,19,21,23,\ldots\} \\ \theta(100110101011\ldots) &= \{1,5,7,11,15,19,21,\ldots\} \end{aligned}$$

4. The Relationship Between the Periodicities of θ and $\sigma(\theta)$

Since we now have this correspondence between sum-free sets and binary decision sequences, we want to further explore how their periodicity properties relate. Cameron observed the following:

Lemma 1. *[3] If $S = \theta(\sigma)$ is ultimately periodic then its decision sequence σ is also ultimately periodic.*

It is natural to ask at this point if the converse of Lemma 1 is true. We pose it as a problem.

Problem 2. *[3] Is the following statement true: A sum-free set $S = \theta(\sigma)$ is ultimately periodic if and only if its decision sequence σ is ultimately periodic?*

Several sets have been proposed as possible counterexamples to this statement. However, the ultimate structure of these sets is not known and it has not been shown with certainty that they are aperiodic.

Problem 3. *Suppose a sum-free set S appears to be a counterexample to Problem 2. How would one prove that S is aperiodic?*

First, we need a method for looking at a finite number of terms of a sum-free set and using these terms to conclude the set is ultimately periodic.

Lemma 4. *[1, 2] Let the decision sequence of a sum-free set S be $\sigma = \theta^{-1}(S)$ with σ ultimately periodic. Suppose that for some number p we can compute a finite number of terms of S such that S begins with an apparent preperiod subset and then three consecutive cycles of period p (we will not require that this p be the minimal period of the set). Further, suppose that*

$$S_* = \{s \in S \mid s < p\},$$
$$S_1 = \{s \in S \mid p < s < 2p\},$$
$$S_2 = \{s \in S \mid 2p < s < 3p\},$$
$$S_3 = \{s \in S \mid 3p < s < 4p\}.$$

Finally, suppose that the first element of each of the three periodic cycles corresponds to a 1 in the same position of a periodic part of σ (again, not requiring this to be for the minimal period of σ). Then S is ultimately periodic.

There are 51 distinct (nonzero) purely periodic decision sequences with period at most five. Of these, all but three quickly show an ultimate periodicity which Lemma 4 ensures will continue. For example,

$$\theta(\dot{0}01\dot{1}) = \{n \geq 1 \mid \quad n \equiv 3, 4 \pmod 7\}$$
$$\theta(\dot{1}100\dot{1}) = \{1, 3, 8, 10\} \cup \{n \geq 12 \mid \quad n \equiv 1, 8, 10 \pmod{11}\} .$$

The remaining three purely periodic decision sequences with period five are $(\dot{0}100\dot{1})$, $(\dot{0}101\dot{0})$, and $(\dot{1}001\dot{0})$. Despite calculating the corresponding sum-free sets for all numbers up to 50 million (over four million elements in each set) we still have been unable to determine any ultimate period. (More details on how we searched for an ultimate period are given in section 6)

5. Structure of Certain Sum-Free Sets

One approach to try to understand Cameron's problem is to study the structure of periodic binary strings and how they relate specifically to the structure of the set. The hope is that something can be found in these structures which would explain why a sum-free set, like $\theta(\dot{0}100\dot{1})$ for example, would appear to be aperiodic. The following propositions describe the structure of some sets with a decision sequence containing only one or two 1's.

Proposition 5. *Let σ be a purely periodic decision sequence, $\sigma = (\dot{\sigma}_1\sigma_2 \dots \dot{\sigma}_p)$. Suppose that exactly one of $\sigma_1, \sigma_2, \dots \sigma_p$ (call it σ_j) is 1. Then $\theta(\sigma)$ is purely periodic with period $p+1$ and $\theta(\sigma) = \{n \in \mathbb{N} \mid n \equiv j \bmod (p+1)\}$.*

Proof. Let $S = \theta(\sigma)$. Since $\sigma_1 = \sigma_2 = \dots = \sigma_{j-1} = 0$, j is the first element of S and the sum $j + j$ will not be considered for inclusion in the set. Think of j as an element of the additive group $\mathbb{Z}/(p+1)\mathbb{Z}$ and define $j' \equiv j + j \bmod (p+1)$. Notice that $j' \not\equiv j \pmod{p+1}$.

The decision sequence now has $p-1$ consecutive 0's to run through, causing it to exclude the next p possible numbers from the set, because the decision sequence will not be consulted on j' (or, possibly, $j'+p+1$ if $j \geq \frac{p+1}{2}$). Thus, the second element of S will be $j+p+1$, the third will be $j+2(p+1)$, etc. Thus $\theta(\sigma) = \{n \in \mathbb{N} \mid n \equiv j \bmod (p+1)\}$, which is a purely periodic sum-free set. □

The proofs of next two propositions follow the same idea as the proof of Proposition 5.

Proposition 6. *Let σ be a purely periodic decision sequence of period p with exactly two 1's. Let i be the position of the first 1 and let j be the position of the second. Suppose that $j-i \neq p/2$ (otherwise, σ can be viewed as purely periodic with period $p/2$ and $\theta(\sigma)$ is purely periodic by Proposition 5). Then $\theta(\sigma)$ is a purely periodic sum-free set if and only if one of the following two conditions hold: (i) $j < 2i$ and $p \geq 2j-i-2$, or (ii) $j \geq 2i$ and $p \geq 2j-i$. In particular, in case (i), $\theta(\sigma) = \{\ n \geq 1 \ \mid \ n \equiv i,\ j \pmod{p+3}\ \}$, and in case (ii), $\theta(\sigma) = \{\ n \geq 1 \ \mid \ n \equiv i,\ j+1 \pmod{p+3}\ \}$.*

Proposition 7. *Let σ be a purely periodic decision sequence of period $p \geq 3$, with 1's in the first and last positions and 0's everywhere else. Then $\theta(\sigma)$ is an ultimately periodic sum-free set and $\theta(\sigma) = \{1\} \cup \{n > 1 : n \equiv p+1, p+3 \pmod{p+4}\}$.*

This is by no means an exhaustive look at this approach. Unfortunately these propositions cannot account for all binary strings with two 1's which yield ultimately periodic sets.

6. Difference Density

We now introduce a different approach and look specifically at some sets which are thought to be counterexamples for Problem 2. To better understand the behavior of these sets, we consider the differences between elements of a set. We want to measure how frequently a certain number occurs as a difference between two elements of a set and then compare this to the number of elements to obtain a difference density. For a sum-free set S, define

$$d_n(m) := \frac{\#\{x \in S \mid x - m \in S,\ x \leq n\}}{\#\{x \in S \mid x \leq n\}}.$$

If we look at $d_n(m)$ as $n \longrightarrow \infty$ and this limit exists, then define

$$d(m) := \lim_{n \to \infty} d_n(m).$$

If the limit exists, then clearly $0 \leq d(m) \leq 1$. When we apply these functions to purely or ultimately periodic sum-free sets, they behave in a predictable way as described in the following proposition.

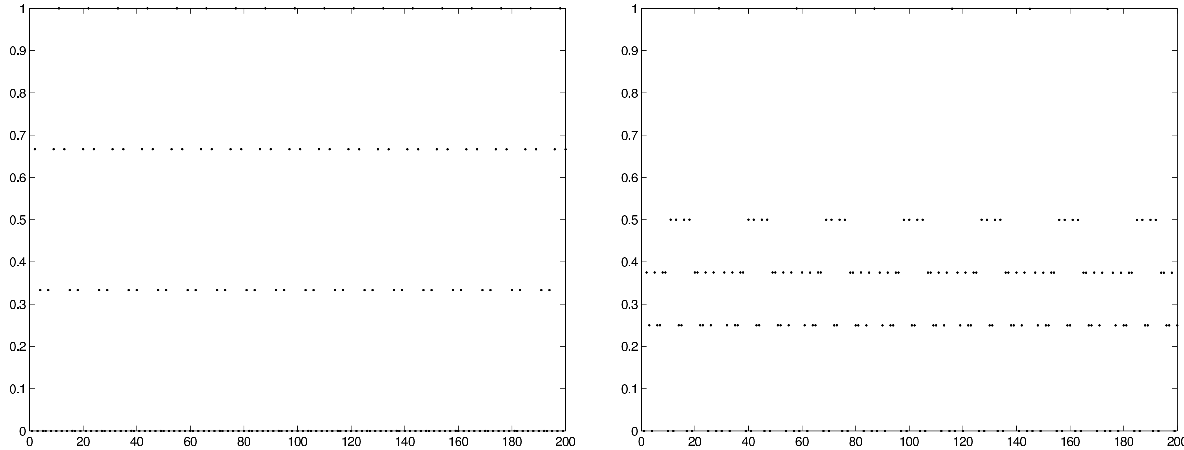

Figure 1: Sample plots of difference densities for periodic sum-free sets.

Proposition 8. *If S is an ultimately periodic sum-free set with ultimate period p and $\overline{S} = \{q_1, q_2, \ldots, q_r\}$, then $d(m)$ exists $\forall\ m \in \mathbb{N}$ and is determined by the following:*

(i) If $p \mid m$, then $d(m) = 1$.

(ii) If $p \not| \, m$ and m cannot be written as $q_{i_1} - q_{i_2} (\text{mod } p)$, then $d(m) = 0$.

(iii) If $p \not| \, m$ and m can be written as $q_{i_1} - q_{i_2} (\text{mod } p)$, then $d(m) = \frac{\gamma}{r}$, where γ is the number of ways to write m in this manner.

This proposition provides a method for searching for a period of a sum-free set: Compute difference density values along with the set and, when a value of $d_n(m)$ approaches 1, that m value is a candidate for the period. Then Lemma 4 can be applied to verify that the set is periodic.

Another consequence of Proposition 8 is that we have a characterization of the appearance of graphs of difference density values for all periodic sets. The graph will have points at 1 for every m which is a multiple of the period, with all other points at 0 or a rational number between. This is demonstrated in the sample plots in Figure 1. Make special note of the points at 1 in each plot. While difference density plots for periodic sum-free sets can be more complicated than these examples, Proposition 8 requires that they all share this distinctive feature.

For comparison, we consider difference density graphs for the sets $\theta(\dot{0}100\dot{1})$, $\theta(\dot{0}101\dot{0})$, and $\theta(\dot{1}001\dot{0})$. Recall that these three sets are potential counterexamples to Problem 2. Plots of $d_n(m)$ for $n = 5 \times 10^7$ and selected m values are included for $\theta(\dot{0}100\dot{1})$ (Figure 2) and for $\theta(\dot{0}101\dot{0})$ and $\theta(\dot{1}001\dot{0})$ (Figure 3). We computed $d_n(m)$ values for these sets for all m up to 1.25×10^7. The plots in all ranges for m look the same as the samples we have included.

We emphasize that these plots never have a point close to 1. In particular, notice that

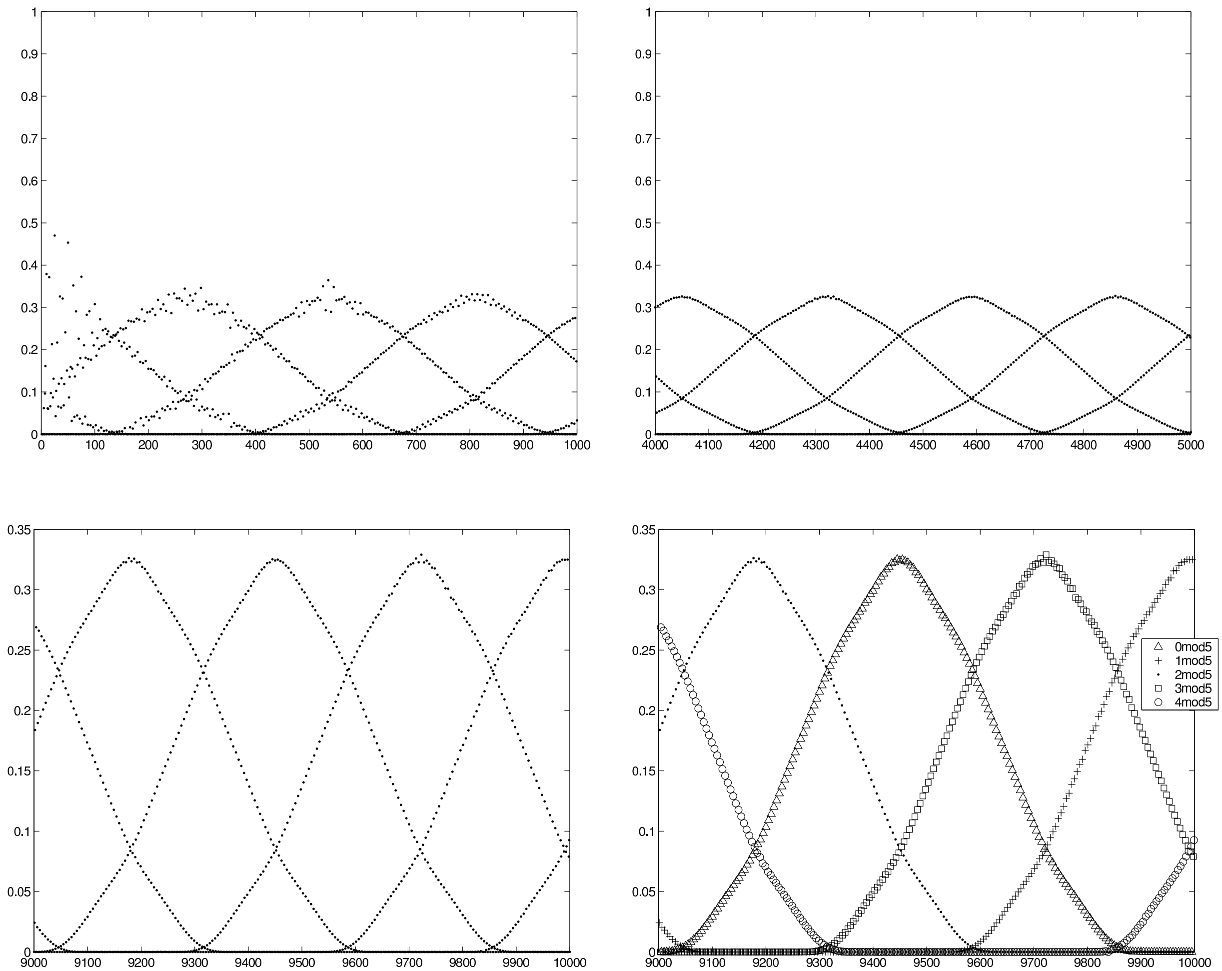

Figure 2: Plots of difference densities for $\theta(\dot{0}100\dot{1})$.

the maximum $d_n(m)$ value is usually around 0.35 and always less than 0.5 . If any of these three sets were ultimately periodic, there should be a point approaching 1 at every multiple of the period. This provides strong evidence that the sets are aperiodic. However, since we cannot show $d(m) < 1$ for all m, we cannot prove any of $\theta(\dot{0}100\dot{1})$, $\theta(\dot{0}101\dot{0})$, or $\theta(\dot{1}001\dot{0})$ to be aperiodic.

We can make another observation from these difference density graphs which suggests the sets are aperiodic. As shown in Figure 2, the $d_n(m)$ values for $\theta(\dot{0}100\dot{1})$ can be viewed as five copies of an overlapping pattern, each with the same shape. These five pieces correspond exactly with viewing the $d_n(m)$ values by the congruence classes of $m \bmod 5$. A similar observation can be made for $\theta(\dot{0}101\dot{0})$, with the exception that the m values should be split into congruence classes mod 7. However, we have not found a comparable way to view the difference densities of $\theta(\dot{1}001\dot{0})$.

All of the evidence given above that the sets $\theta(\dot{0}100\dot{1})$, $\theta(\dot{0}101\dot{0})$, and $\theta(\dot{1}001\dot{0})$ are ape-

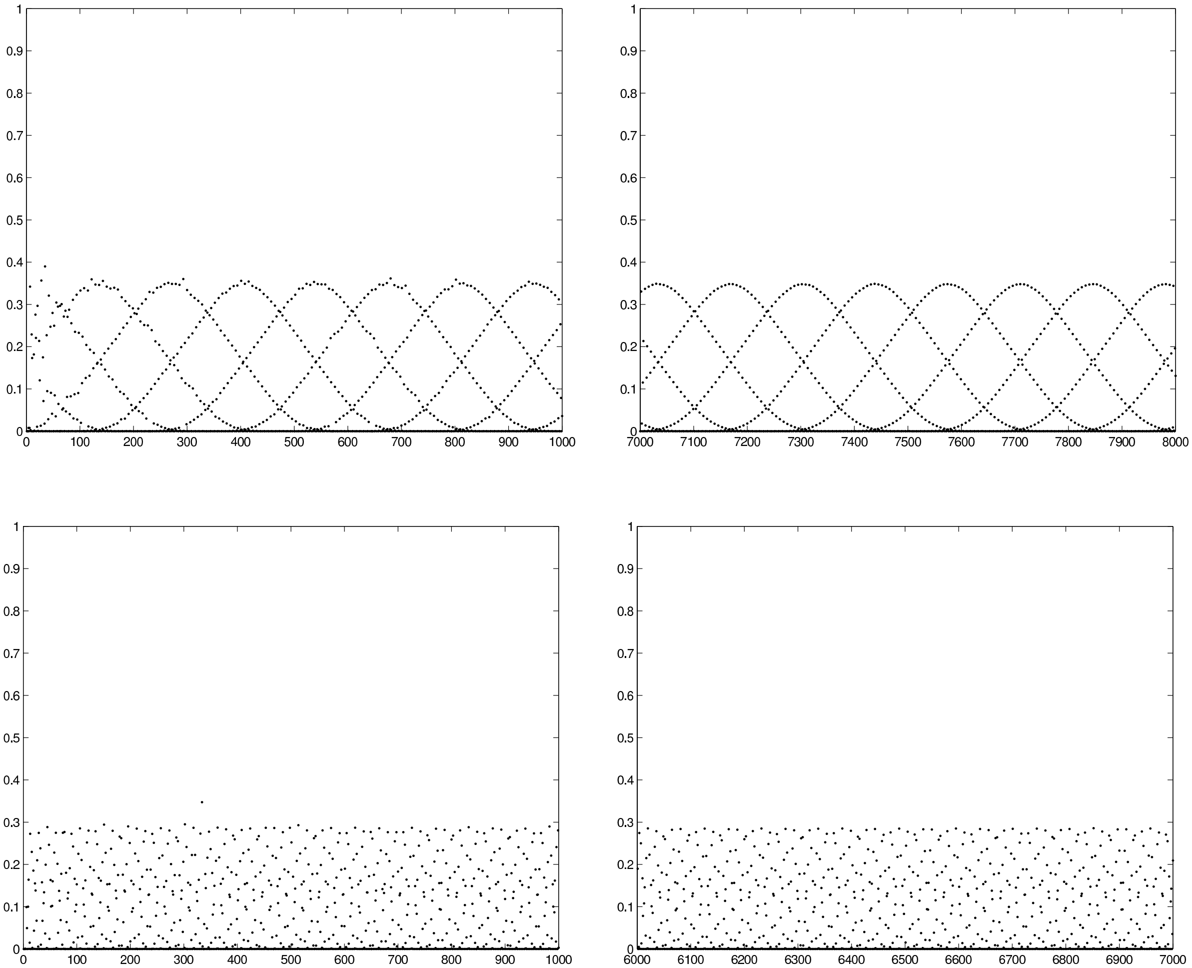

Figure 3: Plots of difference densities for $\theta(\dot{0}101\dot{0})$ (top) and $\theta(\dot{1}001\dot{0})$ (bottom).

riodic is based on finite calculations of infinite sets. Thus, let us consider for a moment the possibility that one or more of the sets is ultimately periodic. This would imply that we have not calculated enough terms of the set to be able to find the ultimate period. To demonstrate why this does not seem to be the case, we need to look at an example of a difference density plot of an ultimately periodic set which only considers elements of the set occuring before the ultimate period takes effect.

Recall that an ultimately periodic sum-free set S can be written as $S = S_* \cup S_{\text{per}}$, where S_* is the preperiod subset. There are examples of ultimately periodic sets where S_* is quite large. For example, the sum-free set $\theta(\dot{0}11001\dot{1})$ has ultimate period $p = 10710$ and there are around 89000 elements in S_* [1]. We computed several million terms of $S = \theta(\dot{0}11001\dot{1})$ and separated S into S_* and S_{per}. We then did two calculations of $d_n(m)$ values, once using only S_* and then again using all of S. In Figure 4, selected plots of difference densities for this set are shown. When $m = p$ in the plot on the left, we observe $d_n(m)$ is about 0.75 . This value is repeated on other intervals when $m = 2p,\ 3p$, etc. In the plot on the right using all of S,

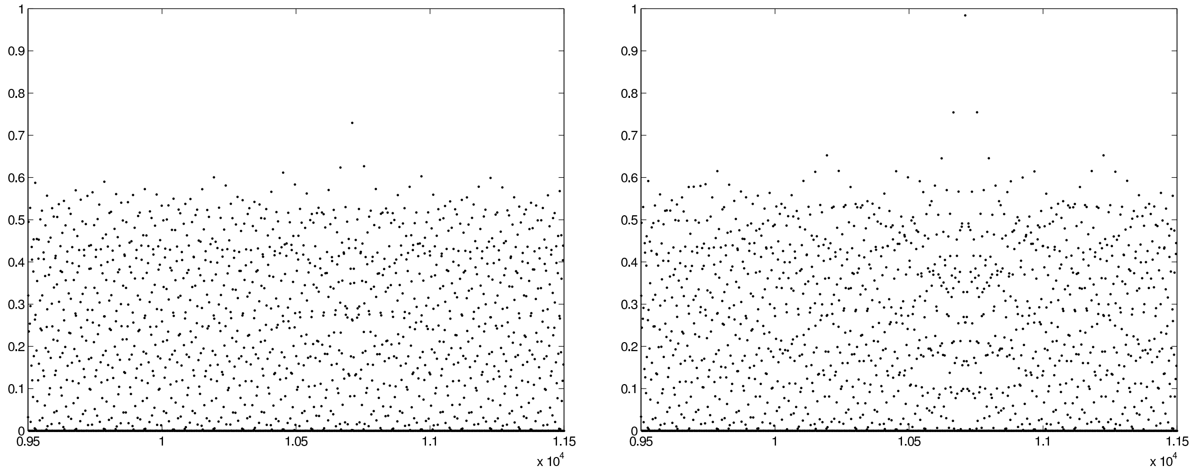

Figure 4: Plots of difference densities for $\theta(\dot{0}11001\dot{1})$ using only elements of S_* (left) and all n up to 10 million (right).

when $m = p$ we observe $d_n(m)$ approaches 1 as expected. Thus, if one of $\theta(\dot{0}100\dot{1})$, $\theta(\dot{0}101\dot{0})$, or $\theta(\dot{1}001\dot{0})$ was ultimately periodic, we might expect to see some evidence of the period in the $d_n(m)$ values. Since the largest $d_n(m)$ value we found always occurs before $m = 400$ and is not repeated, we have further evidence which suggests the three sets are aperiodic.

7. Open Questions

The main open questions are Problems 2 and 3. We have given 3 specific candidates for counterexamples in the preceding sections. We have also computed all the sum-free sets corresponding to purely periodic binary strings of period at least six and no more than 11. Of these we have a total of 1012 sets for which we have not found an ultimate period. However, we have not done extensive computations for these sets.

We submit the following additional open questions:

1. What insights can we gain from difference density graphs of sets with binary period 6-11? How many of these difference densities will exhibit a pattern in congruence classes as seen for $\theta(\dot{0}100\dot{1})$ and $\theta(\dot{0}101\dot{0})$?

2. Can the elements of the sets $\theta(\dot{0}100\dot{1})$ or $\theta(\dot{0}101\dot{0})$ be written in some kind of closed form? Do the patterns in the difference densities give any insight on how this might be done?

3. What property of $\theta(\dot{1}001\dot{0})$ keeps its difference density graph from having a pattern modulo some number? Is this set the exception or is the congruence class pattern rare?

4. Can Propositions 6 and 7 be extended to account for all binary strings with exactly two 1's which yield purely or ultimately periodic sets? Is there a way to use these to predict which binary strings will yield potentially aperiodic sets before computing the sets?

5. Suppose for some arbitrary decision sequence σ, the density of 1's in σ exists. Does $d(m)$ exist for such a σ?

References

[1] Neil J. Calkin, and Steven R. Finch, Conditions on Periodicity for Sum-Free Sets, *Experimental Mathematics* **5** (2) (1996), 131–137.

[2] Neil J. Calkin, Sum-free sets and measure spaces, PhD thesis, University of Waterloo, 1988.

[3] P. J. Cameron, Portrait of a Typical Sum-Free Set, pp. 13-42 in *Surveys in Combinatorics*, 1987 (edited by C. Whitehead), London Math. Soc. Lecture Notes **123**, Cambridge U. Press, Cambridge, 1987.

[4] Richard K. Guy, *Unsolved Problems in Number Theory*, Second Edition, Springer-Verlag, New York, 1994.