ON THE STRUCTURE OF A RANDOM SUM-FREE SET OF POSITIVE INTEGERS

NEIL J. CALKIN

Abstract. Cameron introduced a natural probability measure on the set $\mathbb{S}$ of sum-free sets, and asked which sets of sum-free sets have a positive probability of occurring in this probability measure. He showed that the set of subsets of the odd numbers has a positive probability, and that the set of subsets of any sum-free set corresponding to a complete modular sum-free set also has a positive probability of occurring. In this paper we consider, for every sum-free set S, the representation function $r_S(n)$, and show that if $r_S(n)$ grows sufficiently quickly then the set of subsets of S has positive probability, and conversely, that if $r_S(n)$ has a sub-sequence with suitably slow growth, then the set of subsets of S has probability zero. The results include those of Cameron mentioned above as particular cases.

1. Introduction

Let S be a subset of $\mathbb{N} = \{1, 2, 3, \dots\}$ or $\mathbb{Z}_m = \{0, 1, 2, \dots, m-1\}$; we shall say that S is *sum-free* (with respect to addition or (mod m) addition respectively) if for all $x, y \in S, x + y \notin S$. In order to distinguish between the two cases we shall adopt the following convention: if $S \subset \mathbb{N}$ then we shall refer to S as a *sum-free set*; if $S \subset \mathbb{Z}_m$ then we shall refer to S as a *modular sum-free set*, as a *sum-free set in* $\mathbb{Z}_m$, or as a *sum-free set (mod m)* (according to whether we wish to specify the value of the modulus). We denote by $\mathbb{S}$ the set of all sum-free sets $S \subset \mathbb{N}$.

Given any set $\bar{S} \subset \mathbb{Z}_m$, say $\{s_1, s_2, \dots, s_k\}$ we may easily construct a corresponding set $S \in \mathbb{S}$; indeed let S be the set of all elements of $\mathbb{N}$ which are congruent to an element of $\bar{S}$ (mod m), i. e. the set S is the union of the residue classes $S_i = \{s | s \equiv s_i \pmod{n}\}$. It is clear that if $\bar{S}$ is sum-free (mod m) then the corresponding set $S \subset \mathbb{N}$ is a sum-free set. If m is the least modulus for which S corresponds to a set which is sum-free (mod m), then we say that S is periodic with period m. Clearly S is periodic if and only if there is an m so that for every $n \geq 1$, $n \in S$ if and only if $n + m \in S$. If S differs from a periodic set by only finitely many elements, then we say that S is *ultimately periodic.* Otherwise we say that S is *aperiodic.*

We say that a sum-free set S is *ultimately complete* if there exists n_0 so that for every $n \geq n_0$, if $n \notin S$ then there exist $x, y \in S$ with $x + y = n$. A modular sum-free set $\bar{S}$ is *complete* if for every $n \notin S$ there exist $x, y \in S$ with $x + y = n$.

For $S \in \mathbb{S}$, define $r_S(n)$ to be the number of solutions to the equation $x + y = n$ with $x, y \in S$, $x \leq y$. This function is the *representation function* of Halberstam and Roth [**?**]. Clearly, S is ultimately complete if and only if $r_S(n)$ is positive for all sufficiently large $n \notin S$.

Cameron introduced the following simple bijection between the sets $2^{\mathbb{N}}$ and $\mathbb{S}$. Let σ be an element of $2^{\mathbb{N}}$, say $\sigma_1\sigma_2\sigma_3 \dots$ where $\sigma_i \in \{0, 1\}$ for every i. Construct a sum-free set $\theta(S)$ recursively, by considering each n in turn: if it is not an element of $S+S$, put n in S if the next entry in σ is 1, and discard if the next entry is 0. This bijection is more naturally expressed in terms of tossing coins: for each n, if n is not in $S+S$, toss a coin: if the coin is heads, put n in S: if it is tails, discard it. Note that we do not need to toss a coin (or consider an entry in σ) if $n \in S+S$.

The natural probability measure on $2^{\mathbb{N}}$, tossing a coin infinitely often, together with the bijection θ, thus induce a induces a corresponding probability measure p on $\mathcal{S}$, and it is natural to ask the question: for which subsets $\mathcal{T} \subseteq \mathcal{S}$ can we calculate $p(\mathcal{T})$? We shall restrict ourselves to considering the case where $\mathcal{T} = \mathcal{P}(S)$ is the set of subsets of S.

Firstly, we have the following results, due to Cameron [**?**, **?**]

Proposition 1. *Let S be a sum-free set which is not ultimately complete, say $n_1, n_2, n_3, \ldots \in \mathbb{N} \setminus S$ are such that $r_S(n_i) = 0$, i. e. there are no representations of n_i of the form $x + y = n_i, x, y \in S$ for any i. Then*

$$p(\mathcal{P}(S)) = 0.$$

Proof Let p_n denote the probability that each of the elements less than or equal to n of a random sum-free set T are in S, i. e. $T \cap \{1, 2, ..., n\} \subseteq S \cap \{1, 2, ..., n\}$. Then for each $i, p_{n_{i+1}} \leq \frac{1}{2} p_{n_i}$. Now,

$$p(\mathcal{P}(S)) = \lim_{n\to\infty} p_n \leq \lim_{i\to\infty} 2^{-i} = 0.$$

□

Essentially this proof works because every element of the sequence $\theta^{-1}(T)$ which corresponds to an n_i is constrained to be zero: each of these decreases the value of $p(\mathcal{P}(S))$ by a factor of 2; as there are infinitely many such n_i, $p(\mathcal{P}(S)) = 0$.

Corollary 1. *Let S, T be sum-free sets, and suppose that $S \bigtriangleup T$, the symmetric difference between S and T, is not finite. Then*

$$p(\mathcal{P}(S \cap T)) = 0.$$

Proof $S \bigtriangleup T$ is an infinite set of elements which are not expressible as sums in $S \cap T$. From Proposition 1, $p(\mathcal{P}(S \cap T)) = 0$. □

The first successful approach to the question "for which S is $p(\mathcal{P}(S)) > 0$?" was by Cameron [**?**], who proved:

Theorem 1. *Let $S = \{1, 3, 5, 7, 9, 11, \ldots\}$. Then*

$$0.21759\ldots \leq p(\mathcal{P}(S)) \leq 0.21862\ldots$$

Proof See Cameron [**?**].

Cameron was mainly concerned with the case where S is a periodic sum-free set: in [**?**] he used the Fortuin-Kasteleyn-Ginibre [**?**] inequality to generalise the above result as follows.

Theorem 2. *Let $\bar{S}$ be a complete sum-free set (mod S) and let S be the corresponding sum-free set in $\mathbb{N}$. Then $p(\mathcal{P}(S)) > ((c/2)^{m-k})$ where $k = |\bar{S}|$, and $c = 0.218\ldots$ is the probability that a random sum-free set is contained in* $\{1, 3, 5, 7, 9, \ldots\}$

Proof See Cameron [**?**].

Cameron conjectured, following this result, that with probability 1, a random sum-free set is contained in some modular complete sum-free set. If true, this conjecture would have two related, but distinct consequences:

(i) The only sets S for which $p(\mathcal{P}(S)) > 0$ would be those corresponding to complete sum-free sets (mod S).
(ii) If $\mathcal{T} = \bigcup_S \mathcal{P}(S)$ where the union is taken over all sets of the above form, then

$$p(\{S \mid S \notin \mathcal{T}\}) = 0.$$

In Section 2 we shall show that this conjecture is in fact false.

Cameron [?] also gave a partial converse to Theorem 2 which can be stated as follows:

Theorem 3. *If the sum-free set S in $\mathbb{N}$ is ultimately periodic, ultimately complete, and the corresponding modular sum-free set $\bar{S}$ is not complete, then $p(\mathcal{P}(S)) = 0$.*

We shall prove the following slightly stronger theorem: the proofs use essentially the same ideas, so we shall just prove the latter.

Theorem 4. *If a sum-free set S is ultimately complete, and there exists a finite set B such that $S \setminus B$ is incomplete, then $p(\mathcal{P}(S)) = 0$.*

Proof As usual let p_n denote the probability that a random sum-free set U has all terms $\leq n$ contained in S, i. e.

$$p_n = p(\{U \mid U \cap \{1, 2, \ldots, n\} \subseteq S\}).$$

Let $n_1, n_2, \ldots, n_k, \ldots$ be elements of $\mathbb{N} \setminus S$ such that
(i) $\nexists x, y \in S \setminus B$ such that $x + y = n_i$
(ii) $n_{i+1} > 2n_i$ for each i.
We shall show that $p_{n_{i+1}} \leq p_{n_i}(1 - 2^{-|B|-1})$ so that $p_n \to 0$, proving our result. Let E_n denote the event that

$$U \cap \{1, 2, \ldots, n\} \subseteq S \cap \{1, 2, \ldots, n\},$$

so that $E_i \supseteq E_{i+1} \supseteq E_{i+2} \supseteq \ldots$, and $p_n = p(E_n)$. Assume that E_{n_i} holds. Then if $E_{n_{i+1}}$ holds, either
(i) at least one of $n_{i+1} - b \in S$ for $b \in B$, or
(ii) there is a 0 in the string $\theta^{-1}(U)$ in the position corresponding to n_i.
The probability of (i) is $\leq 1 - 2^{|B|}$, say q; the probability of (ii) is $\leq (1 - q)/2$. Thus

$$p(E_{n_{i+1}} \mid E_{n_i}),$$

the probability of $E_{n_{i+1}}$ given E_{n_i} satisfies

$$\begin{aligned} p(E_{n_{i+1}} \mid E_{n_i}) &\leq q + (1 - q)/2 \\ &= 1/2 + q/2 \\ &= 1/2 + 1/2 - 2^{-|U|}/2 \\ &= 1 - 2^{-|U|-1}. \end{aligned}$$

Therefore $p_{n_{i+1}} \leq p_{n_i}(1 - 2^{-|U|-1})$ as stated. □

These results answer questions about $p(\mathcal{P}(S))$ mainly when S is ultimately periodic: if S is sum-free, ultimately complete, and ultimately periodic, with period m, then $p(\mathcal{P}(S)) > 0$ only if the corresponding set $\bar{S}$ (mod S) is sum-free complete; if S is periodic and ultimately complete then $p(\mathcal{P}(S)) > 0$.

2. The probability that a random sum-free set U is contained in a given sum-free set.

We now take a different approach to the problem: recall the definition of the representation function: $r_S(n)$ denotes the number of representations of n in the form

$$n = x + y, \quad x, y \in S, \quad x \leq y.$$

Observe that $r_S(n) = 0$ if $n \in S$, and that if S is ultimately complete then for some n_0, $r_S(n) > 0$ for all $n > n_0, n \notin S$. Thus, if $p(\mathcal{P}(S)) > 0$ then by Proposition 1 it is necessary that $r_S(n) > 0$ for all $n > n_0, n \notin S$. Further, from Theorems 3 and 4 there are sets for which $r_S(n)$ has a bounded subsequence $r_S(n_k)$, and for which $p(\mathcal{P}(S)) = 0$. What is the relationship between $r_S(n))$ and $p(\mathcal{P}(S))$? We prove that if $r_S(n)$ grows sufficiently quickly then $p(\mathcal{P}(S)) > 0$ (Theorem 5) and that if r_S does not grow sufficiently quickly then $p(\mathcal{P}(S)) = 0$ (Theorem 6).

Theorem 5. *If there exists $c > 1/(2 - \log_2 3)$ and n_0 such that $r_S(n) > c \log_2 n$ for all $n > n_0, n \notin S$ then $p(\mathcal{P}(S)) > 0$.*

Proof It is sufficient to show that there exists a subset $\mathcal{U}$ of $\mathcal{P}(S)$ of positive probability, since $p(\mathcal{P}(S)) \geq p(\mathcal{U}) > 0$. Let

$$S_1 = \{n \mid n \in S, n < n_1\}$$

$$S_2 = \{n \mid n \in S, n \geq n_1\}.$$

where n_1 will be chosen later in a suitable fashion. Further, for any set T define

$$T(n) = \{t \mid t \in T, t \leq n\}.$$

We shall show that the set

$$\mathcal{U} = \{S_1 \cup T | T \in \mathcal{P}(S_2)\}$$

satisfies $p(\mathcal{U}) > 0$. Observe that

$$\mathcal{U} = \{T | S_1 \subseteq T \subseteq S\}$$

so that $\mathcal{U}$ consists of precisely those subsets of S which contain S_1.

Suppose that $k \in \mathbb{N} \setminus S, n_1 < k < n$. How many subsets T of $S_2(n)$ are such that $T \cup S_1$ does not contain x, y such that $x + y = k$? There are $r_S(k)$ pairs $x, y \in S$ such that $x + y = k$; therefore the number of such subsets T is at most

$$(1) \qquad 2^{|S_2(n)| - 2r_S(k)} 3^{r_S(k)} = (\frac{3}{4})^{r_S(k)} 2^{|S_2(n)|}$$

since if $x + y = k$, then x and y cannot both be in T; thus only 3 of the 4 possible cases may occur:

$$x \notin T, \quad y \notin T$$

$$x \in T, \quad y \notin T$$

$$x \notin T, \quad y \in T.$$

Thus for each pair x, y there is a contribution of at most $2^{-2}3$ to the product. (If k is even, and $k/2 \in S$ then the number of subsets is slightly smaller; however, the upper bound given in 1 still holds. Further, if $x < n_1$ for some pair, then $y \notin T$, so the contribution to the product is $2^{-1} < 2^{-2}3$, so the upper bound still holds. If both x and y are less than n_1 then the contribution to the product is zero, and so the number of such subsets is zero.) Thus the number of subsets T of $S_2(n)$ for which there is at least one value of $k \notin S$ such that k must be explicitly excluded (i. e. $\nexists x, y \in T, x + y = k$),$n_1 \leq k \leq n$, is at most

$$\begin{aligned}
\sum_{\substack{n_1 \leq k \leq n \\ k \notin S}} (\frac{3}{4})^{r_S(k)} 2^{|S_2(n)|} &= 2^{|S_2(n)|} \sum_{\substack{n_1 \leq k \leq n \\ k \notin S}} (\frac{3}{4})^{r_S(k)} \\
&< 2^{|S_2(n)|} \sum_{\substack{k \geq n_1 \\ k \notin S}} (\frac{3}{4})^{r_S(k)} \\
&< 2^{|S_2(n)|} \sum_{\substack{k \geq n_1 \\ k \notin S}} (\frac{3}{4})^{c \log_2 k} \\
&< 2^{|S_2(n)|} \sum_{k \geq n_1} (\frac{3}{4})^{c \log_2 k}
\end{aligned}$$

This sum converges if $c > \frac{1}{\log_2 4 - \log_2 3}$. Certainly $c = 3$ will suffice.

Suppose now that $T \subseteq S_2(n)$ is such that each $k \notin S$, $n_1 \leq k \leq n$ is represented as a sum $x+y = k$, $x, y \in S_1 \cup T$. Then the length of the binary sequence generating $S_1 \cup T$ up to n is a constant (depending upon n and n_1, but independent of T). This length is at most

$$n_1 + |S_2(n)|.$$

Thus the number of sequences of length $n_1+|S_2(n)|$ generating a subset of $S_1 \cup S_2(n)$ is at least

$$2^{|S_2(n)|}(1 - \sum_{k \geq n_1} (\frac{3}{4})^{c \log_2 k}).$$

We thus have the probability that a sequence σ generates a subset of $S_1 \cup S_2(n)$ is at least

$$\begin{aligned}
&2^{-n_1 - |S_2(n)|} 2^{|S_2(n)|} (1 - \sum_{k \geq n_1} (\frac{3}{4})^{c \log_2 k}) \\
&= 2^{-n_1} (1 - \sum_{k \geq n_1} (\frac{3}{4})^{c \log_2 k}).
\end{aligned}$$

Observe that this quantity is independent of n.

We shall now show that we may choose n_1 in such a way that this quantity is positive: indeed, since

$$\sum_{k \geq n_1} (\frac{3}{4})^{c \log_2 k}$$

converges, we may choose n_1 sufficiently large that

$$\sum_{k \geq n_1} (\frac{3}{4})^{c \log_2 k} < \frac{1}{2}.$$

Then the probability that a sequence of length n_1+l generates a subset of $S_1 \cup S_2(n)$ is greater than $2^{-(n_1+1)}$. Consequently

$$p(\mathcal{P}(S)) \geq 2^{-(n_1+1)} > 0.$$

as required. □

As an immediate corollary we have

Corollary 2. *If $\bar{S}$ is modular sum-free complete, mod(m), then $p(\mathcal{P}(S)) > 0$.*

Proof. Indeed,

$$r_S(n) > (\frac{1}{2m} - \varepsilon)n$$

for all n sufficiently large, $n \notin S$. □

We note, however, that the numerical bounds obtained by Cameron [**?**] are better than those obtained by this method of proof.

Observe that Theorem 4 demonstrates that a certain class of sum-free sets for which $r_S(n)$ contains a bounded subsequence satisfy $p(\mathcal{P}(S)) = 0$. It is natural to ask then whether the existence of such a bounded subsequence is sufficient to ensure that $p(\mathcal{P}(S)) = 0$. This turns out to be true; in fact the following, significantly stronger statement is true.

Theorem 6. *Let c_k be such that $\sum 2^{-c_k}$ diverges. Then if S is a sum-free set such that there is a subsequence $\{n_k\}$ of $\mathbb{N} \setminus S$ such that $r_S(n_k) = c_k$ and $n_{k+1} > 2n_k$, then $p(\mathcal{P}(S)) = 0$.*

Proof. Let E_k denote the event that n_k is excluded by smaller elements of a random sum-free set U, (i. e. that $\exists x, y \in U, x + y = n_k$). For any Boolean function B of $E_1, E_2, \dots, E_{k-1}$ we have

$$p(E_k | B(E_1, E_2, \dots, E_{k-1})) < 1 - 2^{-c_k}$$

since n_k will certainly not be excluded if, for every pair $x, y \in S$, $x < y$, $x + y = n_k$, the element y is missing from U. As each of these elements is larger than n_{k-1}, the probability that each of these is missing is 2^{-c_k}.

Now let F_k be the event that

$$U \cap \{1, 2, \dots, n_k\} \subseteq S \cap \{1, 2, \dots, n_k\}$$

so that

$$p(\mathcal{P}(S)) = \lim_{k\to\infty} p(F_k)$$

Clearly F_{k+1} implies F_k, so that

$$p(F_k) = p(F_k|F_{k-1})p(F_{k-1}|F_{k-2}) \dots p(F_2|F_1)p(F_1)$$

Furthermore,

$$p(F_{k+1}|F_k) < 1 - 2^{-c_{k+1}}\frac{1}{2} = 1 - 2^{-c_{k+1}-1}$$

We thus have

$$p(F_k) \le \prod_{i=1}^{k}(1 - 2^{-c_i-1})$$

Since $\sum 2^{-c_i}$ diverges, we have

$$\lim_{k\to\infty} p(F_k) \le \prod_{i=2}^{\infty}(1 - 2^{-c_i-1}) = 0$$

Thus, as claimed, $p(\mathcal{P}(S)) = 0$. □

In order for $\sum 2^{-c_i}$ to diverge, it is sufficient that $c_k < \log_2 k$. Unfortunately there remains a gap between Theorem 5 and Theorem 6; this can be seen in the following corollary.

Corollary 3. *Let S be a sum-free set for which $r_S(n) < \log_2 \log_2 n$ for all sufficiently large n. Then $p(\mathcal{P}(S)) = 0$.*

Proof. Indeed, such a set clearly contains a subsequence $\{n_k\}$ for which

$$r_S(n_k) < c + \log_2 k$$

for k sufficiently large, for which $n_{k+1} > 2n_k$. Thus $p(\mathcal{P}(S)) = 0$. □

Essentially the gap that we have is that if $\sum 2^{-r_S(n)}$ converges, then the corresponding probability is positive; if the sum diverges, this is not enough by itself to show that the probability is zero; we require that the sum of a relatively thin subsequence also diverges. In order to prove any stronger results, to close this gap, it would probably be necessary to consider the dependencies of various random variables in great detail.

We shall now show that Cameron's conjecture is false; indeed, the following set is a counterexample: let

$$S_0 = \{1, 4, 10, 12, 17, 19, 26, 32, 35\}$$

be the set of integers s, $1 \le s \le 35$ which are congruent to elements of the smallest asymmetric complete sum-free set, and let S_i, $i \ge 1$ be defined by

$$S_i = \{-s + 32(i+1) \mid s \in S_0\}.$$

Let $S = S_0 \cup S_1 \cup S_2 \cup \ldots$; it is easily seen that S is sum-free. Clearly $r_S(n)$ grows linearly for $n \notin S$, so by Theorem 5, $p(\mathcal{P}(S)) > 0$. We have seen thus that this set S is a counterexample to Cameron's "main conjecture"; it does not however imply that there is no similar result possible.

3. The density of a random sum-free set

In [?], Cameron proved a result for modular complete sum-free sets similar in nature to the strong law of large numbers, namely the following:

Theorem 7. *If $\bar{S}$ is a complete sum-free set mod(m), then, conditioned upon $U \subseteq S$, U almost surely has density $|\bar{S}|/2m$, i. e. half the density of S.*

We can extend the scope of this Theorem to include the sum-free sets shown in Theorem 5 to have $p(\mathcal{P}(S)) > 0$.

Theorem 8. *Let S be a sum-free set such that $r_S(n) > c \log_2 n$ for all $n \ge n_0$, $n \notin S$, where $c > 3/(2\log_2 4 - 2\log_2 3)$, and suppose that S has asymptotic density d. Then, conditioned upon $U \subseteq S$, a random sum-free set U almost surely has density equal to $\frac{d}{2}$.*

Proof Let $S = \{s_1, s_2, \ldots\}$. Define the random variable X_i by

$$X_i = \begin{cases} 1 & \text{if } s_i \in U \\ 0 & \text{if } s_i \notin U \end{cases}$$

and define

$$Y_n = \sum_{i=1}^{n} X_i. \tag{2}$$

We shall show that $Y_n/n \to 1/2$ almost surely. For this we require the following Lemma:

Lemma 1. *For any set S such that $p(\mathcal{P}(S)) > 0$ the following are true:*
(i) $p(U \cap \{1, \dots, n\} \subseteq S) \leq p(U \subseteq S) + k_1 n^{-1/2}$
(ii) $|p(X_n = 1|U \subseteq S) - 1/2| = O(n^{-1/2})$
(iii) For any $(\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n) \in \{0,1\}^n$ $p(X_1 = \varepsilon_1, X_2 = \varepsilon_2, \dots, X_n = \varepsilon_n|E) \leq c_3 2^{-n}$ *for some* c_3*.*

Proof (i) Let E be the event that $U \subseteq S$, and let E_n be the event that $U \cap \{1, 2, \dots, n\} \subset S$. Then $\bigcap_n E_n = E$ and $E_n \supseteq E_{n+1}$, so we have

$$p(E_n) - p(E) = \sum_{i=n}^{\infty} (p(E_i) - p(E_{i+1}))$$

$$= \sum_{i=n}^{\infty} p(E_i \text{ and } i+1 \notin S \text{ and } i+1 \in U).$$

We shall estimate the value q_i of $p(E_i \text{ and } i+1 \notin S \text{ and } i+1 \in U)$ as $i \to \infty$. Clearly $q_i = 0$ if $i+1 \in S$, so we shall assume that $i \notin S$. Then, since S is ultimately complete, if i is sufficiently large we may assume that there exist $r_S(i+1)$ pairs $x, y \in S$ such that $x + y = i + 1$. Since U is sum-free, and contains $i+1$, for each such pair at most one of x and y is contained in U. Let $|S \cap \{1, 2, \dots, i\}| = r$. Then each of the possible events in E_i has probability at most 2^{-r}, since for a set U in E_i, $\theta^{-1}(U)$ contains an entry corresponding to each element of S (and quite possibly entries corresponding to non-elements of S). Of the 2^r such sequences, at most $(3/4)^{r_S(i+1)} 2^r$ satisfy the condition that at least one of $x, i+1-x$ is missing from U. Thus the ith term of the summation, q_i, satisfies

$$q_i \leq (\frac{3}{4})^{r_S(i+1)}.$$

Since

$$\sum_{i=n, i+1 \notin S}^{\infty} (3/4)^{r_S(i+1)} = O(n^{-1/2})$$

we have
(i) $p(U \cap \{1, 2, \dots, n\} \subseteq S) \leq p(U \subseteq S) + k_1 n^{-1/2}$
as claimed.
(ii) Clearly, if the event $E_{s_n - 1}$ holds then s_n is not the sum of two smaller numbers $x, y \in U$, since $U \cap \{1, 2, \dots, s_n - 1\}$ is sum-free, being contained in S. Thus

$$p(E_{s_n-1} \text{ and } s_n \in U) = \frac{1}{2} p(E_{s_n-1}).$$

Also

$$p(E) \leq p(E_{s_n-1}) \leq p(E) + k_1 n^{-1/2},$$

so

$$\begin{aligned} p(E \text{ and } (s_n \in U)) &= p(E) - p(E\text{and}(s_n \notin U)) \\ &\geq p(E) - p(E_{s_n-1} \text{ and } (s_n \notin U)) \\ &= p(E) - 1/2 p(E_{s_n-1}). \end{aligned}$$

Thus

$$|p(E \text{ and } (X_n = 1)) - \frac{1}{2} p(E)| = O(n^{-1/2}).$$

(iii) The probability that we must estimate is

$$\begin{gathered}p(E \text{ and } (X_1 = \varepsilon_1) \text{ and } (X_2 = \varepsilon_2) \text{ and } \dots \text{ and } (X_n = \varepsilon_n))/p(E) \\ \le p((X_1 = \varepsilon_1) \text{ and } (X_2 = \varepsilon_2) \text{ and } \dots \text{ and } (X_n = \varepsilon_n) \\ \text{and all other numbers } < s_n \text{ are missing})/p(E) \\ \le 2^{-n}/p(E).\end{gathered}$$

□

We shall now complete the proof of Theorem 8. Let $U_n = |Y_n - n/2|$. Then $U_{n+1} = U_n + 1/2$ if either $U_n = 0$ or $X_{n+1} - 1/2$ has the same sign as $Y_n - n/2$, while $U_{n+1} = U_n - 1/2$ otherwise. Now so

$$E(U_{n+1}|E) = E(U_n|E) + O(n^{-1/2}) + O(n^{-1/2}).$$

Summing we obtain

$$E(U_n|E) = O(n^{1/2}).$$

Thus

$$\begin{aligned} p(|\tfrac{Y_n}{n} - \tfrac{1}{2}| > \varepsilon) &= p(U_n > \varepsilon n) \\ &= O(n^{-1/2}/\varepsilon). \end{aligned}$$

Now

$$p(|\frac{Y_{n^3}}{n^3} - \frac{1}{2}| > \varepsilon) = O(n^{-3/2}/\varepsilon).$$

Therefore

$$p(\frac{Y_{n^3}}{n^3} \not\to \frac{1}{2}) \le \lim_{m\to\infty} \sum_{n\ge m} O(n^{-3/2}/\varepsilon) = 0.$$

Thus $\frac{Y_{n^3}}{n^3} \to \frac{1}{2}$ almost always. Now, if p is such that $n^3 \le p \le (n+1)^3$ then since $Y_{n^3} \le Y_p \le Y_{(n+1)^3}$ we have

$$\frac{Y_{n^3}}{(n+1)^3} \le \frac{Y_p}{p} \le \frac{Y_{(n+1)^3}}{n^3}$$

Since $\frac{n^3}{(n+1)^3} \to 1$ this implies that $\frac{Y_p}{p} \to \frac{1}{2}$ almost always. □

School of Mathematics, Georgia Institute of Technology, Atlanta, Ga 30332-0160
E-mail address: calkin@math.gatech.edu