Counting sets of integers, no k of which sum to another

Neil J. Calkin Angela C. Taylor[*]
School of Mathematics
Georgia Institute of Technology
Atlanta, GA 30332

[*]Research for the second author was supported by a Georgia Institute of Technology Graduate Research Assistantship.

Proposed Running Head: Counting sets of integers

Address for Proofs:
Neil J. Calkin or Angela C. Taylor
School of Mathematics
Georgia Institute of Technology
Atlanta, GA 30332

Abstract

We show that for every $k \geq 3$ the number of subsets of $\{1, 2, \ldots, n\}$ containing no solution to $x_1 + x_2 + \ldots + x_k = y$, where the x_i need not be distinct, is at most $c2^{\alpha n}$, where $\alpha = (k-1)/k$.

A set S of positive integers is *sum-free* if S contains no x, y and z (not necessarily distinct) such that $x + y = z$. Cameron and Erdös have shown [3] that the number of sum-free sets contained in $\{\frac{1}{3}n, \frac{1}{3}n + 1, \dots, n\}$ is $c2^{\frac{n}{2}}$, and Alon [1], Calkin [2] and Erdös and Granville (personal communication) have independently shown that the number of sum-free sets contained in $\{1, 2, \dots, n\}$ is $o(2^{n(\frac{1}{2}+\epsilon)})$ for every $\epsilon > 0$. Erdös has asked (personal communication) if the number of sets contained in $\{1, 2, \dots, n\}$ without a solution to $x + y + z = t$ is $c2^{\frac{2n}{3}}$. In this paper, we answer this question in the affirmative and show more generally that the number of sets contained in $\{1, 2, \dots, n\}$ with no solution to $x_1 + x_2 + \dots + x_k = y$ (with the x_i not necessarily distinct) is at most $c2^{\alpha n}$, where $\alpha = (k-1)/k$ and $k \geq 3$. (Note that $k = 2$ corresponds to the sum-free case mentioned above. It is interesting that we get a stronger result for $k \geq 3$ than for $k = 2$, and we shall later show where the method used here fails for $k = 2$.) We know this number must be at least $c2^{\alpha n}$, since if a set S has all its elements in $[n - \alpha n, n]$, then the sum of any k elements of S will be greater than n. Hence all $2^{\alpha n}$ subsets of $[n - \alpha n + 1, n]$ will be included in this number.

In what follows, we will define

$$(*)\text{-free to mean having no solution to } \textstyle\sum_{i=1}^{k} x_i = y$$

$$\mathcal{F}_n = \text{ the set of } (*)\text{-free sets in } \{1, 2, \dots, n\}$$

$$f_n = |\mathcal{F}_n|$$

$$g_n = \text{the number of } (*)\text{-free sets in } \{1, 2, \dots, n\} \text{ which contain less than } \epsilon q \text{ elements greater than } n - q$$

$$h_n = \text{the number of } (*)\text{-free sets in } \{1, 2, \dots, n\} \text{ which contain at least } \epsilon q \text{ elements greater than } n - q$$

$$h_{n,l} = \text{the number of } (*)\text{-free sets in } \{1, 2, \dots, n\} \text{ which contain at least } \epsilon q \text{ elements greater than } n - q \text{ and which have least element } l$$

Theorem 1 *Fix $k \geq 3$, and let $\alpha = (k-1)/k$. There exists a constant c such that the number of subsets of $\{1, 2, \dots, n\}$ containing no solution to*

$$\sum_{i=1}^{k} x_i = y$$

is at most $c2^{\alpha n}$.

Proof The proof will be along the following lines: we shall split $\mathcal{F}_n$ into several parts, where each part will be determined by the number of elements each set has in $[n-q+1,n]$ and by the size of its least element l. The reason we consider the size of the least element in a set is that any set which contains many small elements (in relation to n) cannot contain many medium or large elements, and a set with many medium elements cannot contain many large elements. Hence, the $(*)$-free sets of greatest cardinality will be those with a large least element l. Each subset of a $(*)$-free set is clearly $(*)$-free, so most of $\mathcal{F}_n$ will be those sets with many elements in $[n-q+1,n]$ and a large least element l.

But first we must choose ϵ and q in an appropriate way. We will pick d such that $d > \frac{1}{2\alpha-1}$ and then choose ϵ and q such that

$$\binom{q}{\epsilon q}\epsilon q < \frac{1}{2}2^{\alpha q}$$

and such that any set of ϵq elements in $\{1,2,\ldots,q\}$ contains an arithmetic progression of length at least $2d+1$. We are guaranteed the ability to do this by [4].

We shall first consider the sets which have density less than ϵ in the largest q elements of $\{1,2,\ldots,n\}$; that is, they have less than ϵq elements in $[n-q+1,n]$. The number of ways to get less than ϵq elements in $[n-q+1,n]$ is less than

$$\binom{q}{\epsilon q}\epsilon q$$

and this is less than $\frac{1}{2}2^{\alpha q}$, by our choice of ϵ and q. We multiply this by the number of $(*)$-free sets in $\{1,2,\ldots,n-q\}$ and we see that the number g_n of $(*)$-free sets in $\{1,2,\ldots,n\}$ having fewer than ϵq elements in $[n-q+1,n]$ is at most

$$\binom{q}{\epsilon q}\epsilon q f_{n-q} < \frac{1}{2}2^{\alpha q} f_{n-q}.$$

We shall now prove that the number of sets in $\mathcal{F}_n$ having at least ϵq elements in $[n-q+1,n]$ is at most $c2^{\alpha n}$, where c is independent of n, and the result will then follow by induction. First we shall state two lemmas due to Calkin [2].

Lemma 1 *The number of binary sequences of length b without any pairs of 1s at distance exactly $1,3,5,7,\ \ldots,\ 2d-1$, is at most $2^{\frac{d+1}{2d}(b+2d)}$.*

Proof The number of sequences of length $2d$ without pairs of 1s at an odd distance is exactly $2^{d+1}-1$. Thus the number of sequences of length b without pairs of 1s at an odd distance less than $2d$ is at most

$$(2^{d+1}-1)^{\lceil \frac{b}{2d}\rceil} < (2^{d+1})^{\frac{b}{2d}+1} = 2^{\frac{d+1}{2d}(b+2d)}$$

as required.

Lemma 2 *Given an arithmetic progression $b-da, b-(d-1)a, \ldots, b+da$, the number of subsets of $\{1,2,\ldots,b-1\}$ having no pairs x, y such that $x+y$ is an element of the progression, is at most*

$$2^{\frac{d+1}{2d}(b+a(2d+1))}.$$

Proof Write the elements of $\{1,2,\ldots,b-1\}$ in the following a sequences:

$$A_1 = \{1, b-1, 1+a, b-1-a, 1+2a, b-1-2a, \ldots\},$$

$$A_2 = \{2, b-2, 2+a, b-2-a, 2+2a, b-2-2a, \ldots\},$$

$$\vdots$$

$$A_a = \{a, b-a, 2a, b-2a, 3a, b-3a, \ldots\},$$

where each sequence has either $\lceil \frac{b}{a} \rceil$ or $\lfloor \frac{b}{a} \rfloor$ elements, and every element of $\{1,2,\ldots,b\}$ occurs in exactly one such sequence. Then, for any set S which has no pair of elements summing to a member of the arithmetic progression, the characteristic sequence of S is such that when written as a binary sequences in the order given by $A_1, \ldots, A_a$, each of these binary sequences has the property that there are no 1s at distance exactly 1, 3, 5, 7, $\ldots$, $2d-1$. The number of ways of choosing such a set S is thus at most the number of ways of choosing a sequences of length $\frac{b}{a}+1$, without 1s at an odd distance less than $2d$. This is at most

$$2^{\frac{d+1}{2d}(\frac{b}{a}+1+2d)a} = 2^{\frac{d+1}{2d}(b+a(2d+1))}$$

as desired.

Now we shall place an upper bound on h_n.

Lemma 3 *The number h_n of $(*)$-free sets in $\{1,2,\ldots,n\}$ which contain at least ϵq elements greater than $n-q$ is less than $2^{q+1}2^{\alpha n} + 2^{\alpha n}$.*

Proof If a set has $l > \frac{n}{k}$, then the set is clearly $(*)$-free. Then any element of $[l, n]$ can be in the set, hence the number of sets with $l > \frac{n}{k}$ is

$$2^{n-\frac{n}{k}} = 2^{\alpha n}.$$

Now we shall consider the more interesting case where a set has $l \leq \frac{n}{k}$. We have an arithmetic progression $t-da, t-(d-1)a, \ldots, t, t+a, \ldots, t+da$, and least element l in our set S. Let $\mathcal{K}_l$ be the family of sets with least element l. Then $|\mathcal{K}_l|$ is less than the number of subsets of $[1, n]$ with no solution to $x_1 + x_2 + (k-2)l = y$. Now write x_1 as $z_1 + l$ and x_2 as $z_2 + l$. Next we count the number of subsets of $[0, n-l]$ with no solution to

$$z_1 + z_2 = t - da - kl$$

$$z_1 + z_2 = t - (d-1)a - kl$$

$$\vdots$$

$$z_1 + z_2 = t + da - kl$$

An upper bound for this is

$$2^{\frac{d+1}{2d}(t-kl+1+a(2d+1)}2^{(n-l)-(t-kl)+1}$$

(where the first term is obtained as in Lemma 2 and the second term allows all combinations of elements of $[(n-l)-(t-kl), n-l]$ to be chosen)

$$= 2^{\frac{d+1}{2d}(t-kl+1+a(2d+1))}2^{(n-t)+1}2^{(k-1)l}$$

$$= 2^{\frac{d+1}{2d}(n-kl-(n-t-ad)+a(d+1)+1)}2^{(n-t)+1}2^{(k-1)l}$$

$$\leq 2^{\frac{d+1}{2d}(n-kl)+\frac{(d+1)^2}{2d}a+\frac{d+1}{2d}}2^{(n-t)+1}2^{(k-1)l}$$

$$= 2^{\frac{d+1}{2d}(n-kl)+\frac{da}{2}+a+\frac{a}{2d}+\frac{d+1}{2d}}2^{(n-t)+1}2^{(k-1)l}$$

$$\leq 2^{\frac{d+1}{2d}(n-kl)+q}2^{(n-t)+1}2^{(k-1)l}$$

(since $t \in [n-q+1, n]$.) This is the point at which the difference between the cases of $k = 2$ and $k \geq 3$ arises. (We need $2^{\frac{d+1}{2d}n} < 2^{\alpha n}$, but if $k = 2$ this cannot happen since we have $2^{\alpha n} = 2^{\frac{1}{2}}$.) Then, summing over l from 1 to $\frac{n}{k}$, we find the number of $(*)$-free sets with least element $l \leq \frac{n}{k}$ is

$$2^q 2^{\frac{d+1}{2d}n}\frac{1-2^{-\frac{d+1}{2d}(n+k)}}{1-2^{-\frac{d+1}{2d}k}}$$

$$\leq 2^q 2^{\alpha n} 2$$

$$= 2^{q+1}2^{\alpha n}.$$

So we have that $h_n < 2^{q+1}2^{\alpha n} + 2^{\alpha n}$ ∎

Next we shall show that we may choose c independent of n. We know

$$f_n \leq g_n + h_n < \frac{1}{2}2^{\alpha q}f_{n-q} + 2^{q+1}2^{\alpha n} + 2^{\alpha n}$$

so let $c = 2^{q+3}$. Then if $n \leq q$,

$$f_n < c2^{\alpha n}.$$

Assume $f_r < c2^{\alpha r}$ for $r < n$. Then

$$f_n < (\frac{3c}{4} + 1)2^{\alpha n}$$

$$< c2^{\alpha n}$$

as desired ∎

References

[1] Noga Alon. Independent sets in regular graphs and sum-free subsets of finite groups. *Israel J. Math.*, 73:247–256, 1991.

[2] Neil J. Calkin. On the number of sum-free sets. *Bull. London Math. Soc.*, 22:141–144, 1990.

[3] P. J. Cameron and P. Erdös. On the number of sets of integers with various properties. In Richard A. Mollin, editor, *Number Theory*, pages 61–79. Walter de Gruyter, 1990.

[4] E. Szemeredi. On sets of integers containing no k elements in arithmetic progression. *Acta. Arith.*, 27:199–245, 1975.