Research Problem about Power Monoids

Anay Aggarwal

Contents

1 Layout of Final Paper

If I end up writing a paper on this, then this might be what it looks like?

Possible Title: On Sets of Non-Negative Integers Irreducible as Sumsets.

Topic: Studying the number of irreducible sets by considering the metric of the size of the set as well. Connections to irreducible boolean polynomials and dismal arithmetic, as well as finding maximum independent sets for certain families of graphs. Studying more general properties of $\mathcal{P}(\mathbb{N}_0)$ as well.

Layout:

- Abstract
- Introduction
- Pushing past results to the maximum – original open problem, unimodality results.
- Strongly irreducible sets and lower bounding.
- Upper bounding by looking at multiples of $\{0, d\}$.
- Miscellaneous things (plan later as the results stack up).
- Conclusion/Acknowledgements.
- References.
- Appendices with binomial coefficient asymptotics and empirical computations.

2 Solution to Original Open Problem

This is trivial by [1]. Here's a rough sketch: we can first bound

$$\mathbb{E}[X_n] \leq \frac{\sum k\alpha_{n,k}}{\alpha_n} \leq \frac{\sum k\binom{n}{k}}{2^n(1-\exp(-\Omega(n)))} = \frac{n}{2}\frac{1}{1-\exp(-\Omega(n))} = \frac{n}{2}(1+o(1)) = \frac{n}{2}+o(n).$$

And then lower bound

$$\mathbb{E}[X_n] \geq \frac{\sum k\alpha_{n,k}}{\alpha_n} \geq \frac{\sum\left(k\binom{n}{k} - k\cdot 2^n\exp(-\Omega(n))\right)}{\alpha_n}.$$

Because $\sum k\cdot 2^n\exp(-\Omega(n)) = O(n^2 2^n\exp(-\Omega(n))) \ll O(2^n) = \alpha_n$, we get the lower bound is $\frac{n}{2}+o(n)$ as well, so

$$\mathbb{E}[X_n] = \frac{n}{2}+o(n).$$

Remark 2.1. *The reason that [1] doesn't kill all our problems about $\alpha_{n,k}$ is because looking at the subsets of $[n]$ turns out to be a lot easier. In fact, the method employed is to define random variables and generate the subsets randomly, then bound probabilistically. Unfortunately, sampling a fixed number of elements from $[n]$ is a lot more difficult to describe (we'll need something like Reservoir Sampling, but even this turns out to be difficult), so this is a lot harder. Solving problems about $\alpha_{n,k}$ asymptotically will require new thinking, different from [1] and [2]. I'm pessimistic about a recursion for $\alpha_{n,k}$, but perhaps some sort of recursive bound on the $\alpha_{n,k}$ is possible.*

3 Empirical Results

I've written code to compute many things, from $\alpha_{n,k}$ to $\mathbb{E}[X_n]$ to the number of strong irreducible sets (more on that later). For brevity, I do not paste my code here. One can conjecture many things based on empirical

evidence, such as that $f(n) = \frac{\mathbb{E}[X_n]}{n}$ is *decreasing* with limit $\frac{1}{2}$. However, one should be wary of empirical evidence because it is misleading in computation of $|A_n|$ for small values of n.

4 Bijection to Boolean Polynomials and Applications to Open Problems

There's a bijection between sets and boolean polynomials: If $S \subset [n]$,

$$S \leftrightarrow \bigoplus_{0 \le a \le n} \mathbf{1}_{a \in S} x^a.$$

This is mentioned in Shitov's paper [2], cited in [1]. This paper computes $|A_n| = 2^n(1 - o(1))$. As Susie mentions, [1] extends this. The arguments in [2] are quite elementary and seem adaptable. This bijection presents another method of attack. Additionally, with this bijection, the 2005 paper [3] implies that

Theorem 4.1 ([3])**.** *Determining whether or not a set S is irreducible is NP-complete.*

Be careful to note that this does NOT imply that the problem of computing $\alpha_{n,k}$ is NP-hard directly. Additionally, one should note that the open problem involving the set of lengths of $\mathcal{P}(\mathbb{N}_0)$ could benefit from a Shitov-like argument. In particular, to bound $\tau(f)$ one should generalize Lemma 2.2 of [2].

5 Lower Bounding Partial Sums

One can bound partial sums of the $\alpha_{n,k}$ using Shitov's lemmas. With some work one can prove

Theorem 5.1. *For any integer k with $1 \le k \le n$,*

$$\sum_{t \le k} \alpha_{n,t} \ge \left(\sum_{i \le k} \binom{n}{i} \right) - \min_{d \in \mathbb{Z}, 0 \le d \le k} \left(n^{2d+3} 2^{n/2} + n 2^n \exp\left(-\frac{(n+2d-2k)^2}{2n} \right) \right).$$

In particular, if $k \ge n/2 - \sqrt{n}$,

$$\sum_{t \le k} \alpha_{n,t} \ge \Theta(2^n) - \min_{d \in \mathbb{Z}, 0 \le d \le k} \left(n^{2d+3} 2^{n/2} + n 2^n \exp\left(-\frac{(n+2d-2k)^2}{2n} \right) \right),$$

and if $k \le n/2 - \sqrt{n}$,

$$\sum_{t \le k} \alpha_{n,t} \ge \Theta\left(\binom{n}{k} \left(1 - \frac{2k}{n} \right)^{-1} \right) - \min_{d \in \mathbb{Z}, 0 \le d \le k} \left(n^{2d+3} 2^{n/2} + n 2^n \exp\left(-\frac{(n+2d-2k)^2}{2n} \right) \right).$$

Asymptotics on the min term seem to be difficult but make for a very interesting problem. I have a hunch one should take something of the form $d \sim \frac{Cn}{\ln n}$. One should also note that because

$$\alpha_n \mathbb{E}[X_n] = \sum_{k=1}^{n} k \alpha_{n,k} = n\alpha_n - \sum_{k=1}^{n} \sum_{t \le k} \alpha_{n,t},$$

this gives us an upper bound on $\mathbb{E}[X_n]$. In particular, one can show that the min term is $o(2^n)$ without much difficulty (pick $d = \frac{Cn}{\ln n}$ when $k = O(n)$ and $d = O(1)$ otherwise). Hence this gives us the bound $\mathbb{E}[X_n] \le \frac{n}{2} + o(n)$. The reason we do this is that it perhaps allows for tightening of the $o(n)$ term. This tightening may allow us to make further progress on the front of proving unimodality. In terms of direct application to unimodality, this doesn't contribute anything more than [1] as is, we'll need some more smart thinking. In the next section we'll get some heuristics that will allow us to push theorem 4.1 as far as possible. It turns out that this bound is essentially always trivial.

6 Pushing Past Results To Unimodality

As far as I know, there are only two proofs that the atoms are dense in the literature. One is due to Shitov, [2], and the other is due to Geroldinger and Bienvenu, [1]. As we will soon see, a slight optimization of Shitov's argument allows us to get a precise asymptotic on $\alpha_{n,k}$ for a large interval of k. While Geroldinger and Bienvenu's bound is stronger, it seems difficult to transfer their probabilistic argument over. This is because when we fix the number of ones, independence is lost, so simple arguments don't work (without significantly weakening the bound). However, we can slightly modify their argument to give full unimodality with a little trick.

Before we present our results, we need a few lemmas:

Lemma 6.1. $\sum_{t\leq k}\binom{n}{t} \geq \Theta\left(\binom{n}{k}\left(1-\frac{2k}{n}\right)^{-1}\right)$

Proof. I'll write it later. I saw it on math overflow. □

Lemma 6.2. *If $k = \Omega(n)$,*

$$\log\binom{n}{k} = (1+o(1))n\left(\frac{k}{n}\log\frac{n}{k} + \frac{n-k}{n}\log\frac{n}{n-k}\right)$$

Proof. By Stirling's approximation,

$$\binom{n}{k} = \frac{n!}{k!(n-k)!} = (1+o(1))\frac{\sqrt{2\pi n}(n/e)^n}{\sqrt{2\pi k}(k/e)^k\sqrt{2\pi(n-k)}((n-k)/e)^{n-k}}.$$

Hence

$$\binom{n}{k} = (1+o(1))\sqrt{\frac{n}{2\pi k(n-k)}}\left(\frac{n}{k}\right)^k\left(\frac{n}{n-k}\right)^{n-k}.$$

Taking logarithms,

$$\log\binom{n}{k} = O(1) + \log\sqrt{\frac{n}{2\pi k(n-k)}} + k\log\frac{n}{k} + (n-k)\log\frac{n}{n-k}.$$

Because $k = \Omega(n)$, the second term in this sum can be swept under the rug. This implies the desired result. □

We transfer over the following lemma of Shitov from the language of Boolean Polynomials to our context:

Lemma 6.3. *Let $d > 0$. Then there are at most $n^{2d+3}2^{n/2}$ pairs of sets (A, B) such that $\max A + \max B = n$ and $|A + B| \leq |A| + |B| + d$.*

We can use this to prove the following result:

Theorem 6.1. *For $0.11n < k < 0.5n$, we have that $\alpha_{n,k} = \binom{n}{k-1}(1-o(1))$.*

Proof. Assume $k < n/2$. Consider the number of sets A, B with $|A + B| = k$ and $\max A + \max B = n$. Let d be a real that we will choose later. Then by the lemma there are at most $n^{2d+3}2^{n/2}$ pairs (A, B) with $|A| + |B| \geq k - d$. To count the number of such sets with $|A| + |B| < k - d$, fix $\max A = r$. We may then upper bound this count by

$$\sum_{r=1}^{n-1}\sum_{a+b=k-d}\binom{r}{a}\binom{n-r}{b} = \sum_{r=1}^{n-1}\binom{n}{k-d} < n\binom{n}{k-d}.$$

Therefore the number of sets $|C| = k$ with maximum element n that are reducible is at most $n^{2d+3}2^{n/2} + n\binom{n}{k-d}$. This implies that whenever $\log_2 \binom{n}{k} = n(1/2+c)$ for a constant $c > 0$, we have that almost all sets of size k with maximum element n are irreducible. It is easy to see that this implies that $\alpha_{n,k} = \binom{n}{k-1}(1-o(1))$. Now, if $k = \delta n$,

$$\log_2 \binom{n}{k} = (1 + o(1))nH_2(\delta),$$

where $H_2(x) = -x \log_2 x - (1-x)\log_2(1-x)$. We hence require $H_2(\delta) > \frac{1}{2}$. Numerically solving this for $\delta < 0.5$, $\delta \approx 0.11$ is the "turning point", as desired. □

Pushing [1] directly gives the following:

Theorem 6.2. *There are $\Omega(n)$ values of k such that the appropriate unimodality inequality holds, i.e. $\alpha_{n,k} < \alpha_{n,k+1}$ if $k < n/2$ and $\alpha_{n,k} > \alpha_{n,k+1}$ otherwise.*

Proof. Consider only $k < n/2$. The result follows simply by [1] and some calculations. We will sketch the proof and not be strict about our calculations. We will replace any appearance of $k \pm 1$ with k and any appearance of $n \pm 1$ with n. This doesn't matter asymptotically. If $\alpha_{n,k+1} = \binom{n}{k} - \beta_{n,k+1} < \alpha_{n,k} < \binom{n}{k-1}$, then by Pascal's identity $\beta_{n,k+1} > \binom{n-1}{k-1} \sim \binom{n}{k}$. Therefore, if there are $\delta(n)$ values for k for which this inequality is true, we have that (asymptotically)

$$\sum_{t=1}^{\delta(n)} \binom{n}{t} = O(c^n),$$

for a constant $c < 2$. Because of the lower estimate

$$\sum_{t=1}^{\delta(n)} \binom{n}{t} \geq \Theta\left(\binom{n}{\delta(n)}\left(1 - \frac{2\delta(n)}{n}\right)^{-1}\right)$$

due to 6.1, this implies that

$$\binom{n}{\delta(n)} = O(c^n).$$

Set $\gamma(n) := \frac{\delta(n)}{n}$. By 6.2, taking logarithms, we have that

$$\gamma(n)\log\frac{1}{\gamma(n)} + (1-\gamma(n))\log\frac{1}{1-\gamma(n)} < c + \varepsilon$$

for every $\varepsilon > 0$. Because the function $f(x) = -x\log x - (1-x)\log(1-x)$ is continuous on $(0,1)$ and attains a maximum value of $\log 2$ with $x = \frac{1}{2}$, this implies that the inequality is true for all $\gamma \leq \gamma(n) < \frac{1}{2}$ for some constant $\gamma < \frac{1}{2}$. This implies that $\delta(n) = n - \Omega(n)$, as desired. □

Remark 6.1. *One can explicitly upper bound the amount of k values this method gives using the lower bound $c > 1.754$. Numerically, $f(x) = \log(1.754)$ holds for $\kappa \approx 0.2496$ and $1 - \kappa$. Hence one can lower bound:*

$$n - \delta(n) > (1 - 2\kappa)n \approx 0.5008n.$$

Therefore we will need stronger asymptotics if we desired to prove this for more than $0.5008n$ values of k.

Remark 6.2. *$f(x)$ is the binary entropy function.*

We can, however, slightly generalize a probabilistic lemma in [1] to our liking (we lose the strength of the estimate, but the bound is still $o(1)$):

Lemma 6.4. *Let $k = \delta n$ for a constant $\delta = O(1)$. Let r be a constant positive integer and let $b = (b_1, b_2, \ldots, b_r)$ for some $0 < b_1 < b_2 < \cdots < b_r < n$ such that $n - b_r = \Omega(n)$. Let z be some fixed binary sequence of length r. Uniformly at random, choose a binary string $(a_1, a_2, \ldots, a_n)$ with exactly k ones. Let $A(b; z)$ be the random variable denoting the number of indices j such that $a_{j+b_i} = z_i$ for all i. Then for any $\lambda > 0$,*

$$\mathbb{P}(|A(b;z) - \mathbb{E}[A(b;z)]| > \lambda \mathbb{E}[A(b;z)]) = O(n^{-1}).$$

Proof. By Chebyshev's inequality,

$$\mathbb{P}(|A(b;z) - \mathbb{E}[A(b;z)]| > \lambda \mathbb{E}[A(b;z)]) \leq \frac{\mathrm{Var}(A(b;z))}{\lambda^2 \mathbb{E}[A(b;z)]^2}.$$

It is clear that the denominator here is $\Theta(n^2)$. We will show that the numerator is $O(n)$, proving our claim. One can split

$$A(b;z) = \sum_{i=1}^{n-b_r} A_i(b;z),$$

where $A_i(b;z)$ is the indicator random variable for the equality $(a_{i+b_j}) = z_j$ for all j. We hence have

$$\mathrm{Var}(A(b;z)) = \sum_{1 \leq i \leq n-b_r} \mathrm{Var}(A_i(b;z)) + \sum_{1 \leq i_1 < i_2 \leq n-b_r} \mathrm{Cov}(A_{i_1}(b;z), A_{i_2}(b;z)).$$

The former term is $O(n)$, it suffices to show that the latter is as well. This follows because the covariance is 0 whenever the sequences corresponding to indices i_1 and i_2 don't intersect. Since each sequence intersects $O(1)$ other sequences, there are $O(n)$ positive covariance terms, from which the result follows because each term is $O(1)$. □

Now, we can prove the following in a very similar way to [1]:

Theorem 6.3. *Let $k = \delta n$ for some constant $0 < \delta < 1$. Then $\alpha_{n,k} = \binom{n}{k-1}(1 - o(1))$.*

Proof. Let γ be a positive constant that we will specify later. Let r be a constant positive integer that we will specify later. We choose a set $A \subset [n]$ of size k uniformly at random (such that A contains 0). We track the probability that $A = B + C$ for sets B, C with $\min(|B|, |C|) \geq 2$. We do this by splitting into cases:

- Let S_1 be the set of such A with a decomposition $A = B + C$ such that $|B| + |C| \leq \gamma n$.
- Let S_2 be the analogous set for $\min(|B|, |C|) \leq r$.
- Let S_3 be the set for $|B| + |C| > \gamma n$ and $\min(|B|, |C|) > r$.

It suffices to show that there is $o(1)$ probability that $A \in S_1 \cup S_2 \cup S_3$. We will prove this by showing that the probability that $A \in S_i$ is $o(1)$ for all i (the result follows by union bound).

We have $A \in S_1$ happens with probability at most

$$\binom{n}{k-1}^{-1} \sum_{t \leq \gamma n} \sum_{b+c=t} \binom{n}{b-1}\binom{n}{c-1} \lesssim \binom{n}{k-1}^{-1} \cdot n \cdot \binom{2n}{\gamma n},$$

by well-known binomial sums (follows from an easy generating functions argument). We will choose γ so that this term is $o(1)$.

To calculate $\mathbb{P}(A \in S_2)$, let $|B| \leq r$. Let $\omega \in (0,1)$ that we will choose later. If $\max B \geq \omega n$ then $C \subseteq [0, (1-\omega)n]$ so there are at most $n^r 2^{(1-\omega)n}$ possibilities for A. Now assume $\max B \leq \omega n$. Assume

$$B = \{0, b_1, b_1 + b_2, \ldots, b_1 + b_2 + \cdots + b_r\}$$

where $b_1 > 0$ and $b_i \geq 0$ for $i > 1$. Now for all $n_0 \in A$, $n_0 + b_1 \in A$ or some $n_0 - b_i \in A$. Therefore if $f := \mathbb{1}_A$, we have for all n_0 that

$$(f(n_0), f(n_0 + b_1), f(n_0 - b_1), \ldots, f(n_0 - b_r)) \neq (1, 0, 0, \ldots, 0).$$

By our lemma (with $\lambda = 1$), this happens with probability $o(1)$. Choosing ω sufficiently close to 1 so that

$$(1 - \omega)n < \log_2 \binom{n}{k},$$

we get the desired $o(1)$ overall probability for this case. This is possible by our binomial coefficient asymptotics.

For $\mathbb{P}(A \in S_3)$, assume WLOG (by Pigeonhole and symmetry) that $|B| \geq \gamma n/2$ and $|C| \geq r$. Let $D \subseteq C$ have $|D| = r$. Then $\forall n_0 \in B$, $n_0 + D \subseteq A$. Hence the number of $n_0 \leq n$ such that $n_0 + D \subseteq A$ is at least $\frac{\gamma n}{2}$. On the other hand, the expected amount is asymptotically at most $(n - \max D)\delta^r \leq n\delta^r$. Therefore, when $\gamma > 2\delta^r$, this happens with probability $o(1)$.

Finally, we need to choose r and γ such that $\log \binom{n}{k} - \log \binom{2n}{\gamma n} > n\varepsilon$ for some positive $\varepsilon = O(1)$, and $\gamma > 2\delta^r$. The first condition is equivalent to $H(\delta) > 2H(\gamma/2)$, so we choose r small enough that $H(\delta) > 2H(\delta^r)$ and then we may find an appropriate γ. This completes the proof. □

Finally, this gives us the result we want:

Corollary 6.1. *Let $k = \delta n$ for a constant $0 < \delta < 1$. Then the appropriate unimodality inequality holds. In other words, the sequence $\alpha_{n,k}$ is unimodal for almost all k.*

Proof. If $k < n/2$ then $\alpha_{n,k} > \binom{n}{k-1}(1 - O(1)) = \binom{n}{k-2} \geq \alpha_{n,k-1}$. One can say the same thing if $k \geq n/2$. □

Currently, the best lower bound comes from looking at sets A that can be written as $B+C$ with $|B| = 2$. There is a nice recursive argument that gives an asymptotic here. I conjecture that looking at such decompositions with $|B| = 3$ will give a better bound. For $n = 2, 3, \ldots$, the number of sets with such decompositions are $2, 5, 13, 28, 55, 97, 169, 293, \ldots$ (found via code).

7 A Notion of Strong Irreducibility

It has been noted in CrowdMath 2023 that sets that are 2-sparse and contain 1 are irreducible. One can extend this notion more generally:

Definition 7.1. Let $S \subset [n]$ have minimum nonzero element m. Suppose $2m \notin S$ and for all $s \in S_{>m}$ there is a $t \in S_{<s}$ such that $s + t \notin S$. Call such a t a *nuller* of S. Call S *strongly irreducible*.

We have the following lemma:

Lemma 7.1. *All strongly irreducible sets are irreducible.*

Proof. Suppose S is strongly irreducible and $S = A + B$. Clearly $A, B \subset S$ as $0 \in A \cap B$. The minimum element m cannot be written as the sum of two other elements of S, so m is in exactly one of A, B. WLOG suppose $m \in A$. Then $m \notin B$ because $2m \notin S$. Now, suppose for the sake of contradiction that B contains some element $s \neq 0$. Let s be minimal. Let t be a nuller of s. Clearly $t \notin A$ because then $A + B$ contains an element not in S. But because s is minimal, t cannot be in B. Contradiction. □

One should also note that there are irreducible sets that are not strongly irreducible, take $[n] \cup \{2n + 1\}$. Thus

$$\text{Strongly Irreducible} \implies \text{Irreducible}$$

$$\text{Irreducible} \not\Longrightarrow \text{Strongly Irreducible}$$

Hence the name strong irreducibility.

8 Computing the number of strongly irreducible sets

Strongly irreducible sets seem to make up for quite a large portion of irreducible sets empirically. Introduce the following numbers:

Definition 8.1. Let s_n be the number of strongly irreducible subsets of $[n]$. Let $s_{n,k}$ be the number of strongly irreducible subsets of $[n]$ with size k.

Remark 8.1. *Nothing is on OEIS for s_n nor $s_{n,k}$ for fixed $k > 3$.*

Note $s_{n,k} \leq \alpha_{n,k}$, so we're effectively creating a lower bound here. One has that equality holds in this inequality for $k = 3$ and it does not hold for $k > 3$. The proof of this fact is easy so I omit it. This motivates a conjecture:

Conjecture 8.1. *We have that strongly irreducible sets are dense in the set of irreducible sets. In other words,*

$$\lim_{n\to\infty} \frac{s_n}{\alpha_n} = c,$$

where $c = \frac{1}{2}\prod_{k\geq 1}\left(1 - \frac{1}{2^k}\right) \approx 0.144$.

This notion allows us to identify a dense family of atoms, and hence work with them. These bounds are good for when k is small because sparse sets tend to be more likely to be strongly irreducible. They don't work well for large k. We can prove density using the probabilistic method.

Theorem 8.1. *We have that $s_n \gtrsim 2^n \cdot \frac{49}{768}$.*

Proof. Fix $1 \leq m < k < n$ such that $k + m \leq n$. Consider sets S with $r \notin S$ for $0 < r < m$, $m \in S$, $m + r \notin S$ for $0 < r < k - m$, and $k \in S$. Add the additional conditions that $2m \notin S$ and $k + m \in S$. There are $2^n \cdot \frac{1}{2^k} \cdot \frac{1}{4} = 2^{n-k-2}$ such sets. Assign an even probability distribution among such sets. For all $k < i \leq n$, let X_i be the event that $i \in S$ and i has no nuller. For $i \notin \{k + m, 2m\}$ we may compute

$$\mathbb{P}(X_i) = \mathbb{P}(i \in S)\mathbb{P}(i + m \in S)\mathbb{P}(i + k \in S)\prod_{j=k+1}^{i-1}(1 - \mathbb{P}(j \in S, i + j \notin S)) = \frac{1}{8}\prod_{j=k+1}^{i-1}(1 - \mathbb{P}(j \in S, i + j \notin S)).$$

All terms in this product are at most $\frac{3}{4}$ except for potentially one that is 1 if $j = 2m$. Therefore we may bound as follows:

$$\mathbb{P}(X_i) \leq \frac{1}{8}\left(\frac{3}{4}\right)^{i-k-2}.$$

Then note that $\mathbb{P}(X_{2m}) = 0$ and by similar reasoning as above, $\mathbb{P}(X_{k+m}) \leq \frac{1}{4}\left(\frac{3}{4}\right)^{m-1}$. Notice that by Bonferonni's inequality,

$$\mathbb{P}\left(\bigcap_{i=k+1}^{n} X_i^c\right) \geq 1 - \sum_{i=k+1}^{n}\mathbb{P}(X_i) \geq 1 - \left(\frac{1}{4}\left(\frac{3}{4}\right)^{m-1} - \frac{1}{8}\left(\frac{3}{4}\right)^{m-2} + \frac{1}{8}\sum_{i=k+1}^{n}\left(\frac{3}{4}\right)^{i-k-2}\right).$$

This bound comes out to

$$\mathbb{P}\left(\bigcap_{i=k+1}^{n} X_i^c\right) \geq \frac{1}{3} - \frac{3^{m-1}}{4^m} + \frac{3^{m-2}}{2^{2m-1}} + \frac{3^{n-k-1}}{2^{2n-2k-1}}.$$

Hence by the probabilistic method the number of sets is at least

$$2^{n-k-2}\left(\frac{1}{3}-\frac{3^{m-1}}{4^m}+\frac{3^{m-2}}{2^{2m-1}}+\frac{3^{n-k-1}}{2^{2n-2k-1}}\right).$$

We may handle $m=1$ separately because everything is nicer in this case. Let us wait until the end. For $m>1$, the number of such sets is at least

$$\sum_{1<m<k<n,k+m\le n} 2^{n-k-2}\left(\frac{1}{3}-\frac{3^{m-1}}{4^m}+\frac{3^{m-2}}{2^{2m-1}}+\frac{3^{n-k-1}}{2^{2n-2k-1}}\right).$$

Summing over m, we can lower bound this by

$$\sum_{5\le k\le n/2} 2^{n-k-2}\left(\frac{1}{12}+\frac{3^{n-k-1}}{4^{n-k}}+\frac{3^{n-k-1}}{2^{2n-2k-1}}\right)\sim\frac{1}{768}\cdot 2^n.$$

The number of sets with $m=1$ is at least

$$2^n\cdot\sum_{k=3}^{n-2}\frac{1/2+1/8(3/4)^{n-k-2}}{2^{k+1}}\sim\frac{1}{16}\cdot 2^n.$$

Adding implies the desired result. □

We can also give an easy upper bound:

Lemma 8.1. *We have that $s_n\lesssim\frac{3}{16}\cdot 2^n$,*

Proof. If m is the smallest nonzero element, $2m\notin S$ with probability $\frac{1}{2}$. If k is the second smallest nonzero element, $k+m\in S$ with probability $\frac{1}{2}$, and $k<n/2$ with probability $1-o(1)$. Let t be the next smallest element. Then one of $t+m\notin S, t+k\notin S$ with probability $\frac{3}{4}$, and $t<n/2$ with probability $1-o(1)$. This yields the desired. □

Remark 8.2. *Based on the 2-sparse set ideas from Crowdmath, one can lower bound $s_{n,k}\ge\binom{n-k}{k}$ through standard combinatorial arguments. This same bound gives $s_n=\Omega(\varphi^n)$, which is quite weak.*

Remark 8.3. *I should contribute stuff to the OEIS.*

We can easily bound $s_{n,k}$ when k is small using a weaker version of strong irreducibility:

Theorem 8.2. *If $k=o(\sqrt{n})$, then*

$$s_{n,k}=(1-o(1))\binom{n}{k-1}$$

Proof. We tackle this probabilistically, as usual, assigning an even distribution on $\binom{[n]}{k}$. The minimum nonzero element m of a random set from this distribution is $\ll n-o(n)$ with probability $1-o(1)$, and $2m$ is not in the set with probability also $1-o(1)$. Now consider a random set from this distribution with fixed minimum m. It suffices to show that, when $m\ll n-o(n)$, the set is strongly irreducible with probability $1-o(1)$. We prove the stronger result that there is no $s>m$ in the set such that $s+m$ is also in the set with probability $1-o(1)$. In other words, m is a nuller for all elements. Let Z be the random variable denoting the number of occurences of $(s,s+m)$ both in S. By Markov,

$$\mathbb{P}(Z\ge 1)\le\mathbb{E}[Z]\implies\mathbb{P}(Z=0)\ge 1-\mathbb{E}[Z].$$

We can compute $\mathbb{E}[Z]$ by linearity of expectation easily:

$$\mathbb{E}[Z]\le(n-m)\frac{k^2}{(n-m)^2}.$$

This implies the result. □

In order to adapt our argument for lower bounding s_n to lower bounds on $s_{n,k}$ (for larger k), we need to be more careful about computing $\mathbb{P}(X_i)$. Things are not so easy because independence is lost. We define the following:

Definition 8.2. For some $1 \leq k \leq n$, assign an even probability distribution on the sets $S \in \binom{[n]}{k}$ such that $1 \in S, 2 \notin S$. For a randomly chosen set S from this distribution and a $1 \leq i \leq n$, define the event $X_i^{(k)}$ as the event that $i \in S$ and i has no nuller in S.

It is difficult to precisely bound the probabilities of these events without direct computation. However, we can prove the following slightly weak bound:

Lemma 8.2. *Let $\delta = k/n$ and $j > 4$. If $2j - 1 \leq n$,*

$$\mathbb{P}(X_j^{(k)}) \lesssim \frac{\delta - \delta^2 + \delta^3}{(1-\delta)(j-3)},$$

and if $2j - 1 > n$,

$$\mathbb{P}(X_j^{(k)}) \lesssim \frac{\delta - \delta^2 + \delta^3}{(1-\delta)(n-j-2)}.$$

Proof. Suppose $2j - 1 \leq n$ first. Let $Y_{i,j}^{(k)}$ be the indicator variable for the event that $i \in S$ and $i + j \notin S$. Let A be the event that $j \in S$. Let B be the event that $j + 1 \in S$. Then we have that

$$X_j^{(k)} = A \cap B \cap \left(\bigcap_{i=3}^{j-1} (Y_{i,j}^{(k)} = 0) \right).$$

We have that $\mathbb{P}(A) \sim \mathbb{P}(B) \sim \delta$, so by independence,

$$\mathbb{P}(X_j^{(k)}) \sim \delta^2 \mathbb{P}\left(\sum_{i=3}^{j-1} Y_{i,j}^{(k)} = 0 \right).$$

Now, the idea is that rather than using the union bound (which is too weak), we use Chebyshev's inequality. Let $\mu = \mathbb{E}\left[\sum_{i=3}^{j-1} Y_{i,j}^{(k)}\right]$, then

$$\mathbb{P}\left(\sum_{i=3}^{j-1} Y_{i,j}^{(k)} = 0 \right) \leq \mathbb{P}\left(\left| \sum_{i=3}^{j-1} Y_{i,j}^{(k)} - \mu \right| \geq \mu \right) \leq \frac{\operatorname{Var}\left(\sum_{i=3}^{j-1} Y_{i,j}^{(k)}\right)}{\mu^2}.$$

We can easily compute μ by linearity:

$$\mu = \sum_{i=3}^{j-1} \mathbb{E}[Y_{i,j}^{(k)}] \sim (j-3)\delta(1-\delta).$$

To compute the variance, we can write:

$$\operatorname{Var}\left(\sum_{i=3}^{j-1} Y_{i,j}^{(k)} \right) = \sum_{i=3}^{j-1} \operatorname{Var}(Y_{i,j}^{(k)}) + 2 \sum_{3 \leq i_1 < i_2 \leq j-1} \operatorname{Cov}(Y_{i_1,j}^{(k)}, Y_{i_2,j}^{(k)}).$$

We will show that the covariance is negative, i.e. our Y variables are pairwise negatively correlated. To do this, we have to be a bit precise. Suppose t of the elements in $\{3, 4, \ldots, j-1\} \cup \{j+3, j+4, \ldots, 2j-1\}$ are in our random set. Note that t is a random variable here, but we will show that the desired inequality holds for any discrete value of t. We will show that

$$\mathbb{P}(Y_{i_2,j}^{(k)} = 1 \mid Y_{i_1,j}^{(k)} = 1) \leq \mathbb{P}(Y_{i_2,j}^{(k)} = 1).$$

Because the Y variables are indicators, this implies $\mathbb{E}[Y_{i_1,j}^{(k)}Y_{i_2,j}^{(k)}] \le \mathbb{E}[Y_{i_1,j}^{(k)}]\mathbb{E}[Y_{i_2,j}^{(k)}]$, which is desired. Now, note that, for $j > 4$,

$$\mathbb{P}(Y_{i_2,j}^{(k)} = 1 \mid Y_{i_1,j}^{(k)} = 1) = \frac{t-1}{2j-7} \cdot \frac{2j-7-t}{2j-8},$$

$$\mathbb{P}(Y_{i_2,j}^{(k)} = 1) = \frac{t}{2j-6} \cdot \frac{2j-6-t}{2j-7},$$

by first principles of probability. With the substitution $x = 2j-7$, we wish to show that

$$\frac{(t-1)(x-t)}{x-1} \le \frac{t(x-t+1)}{x+1}.$$

This is true because

$$t(x-t+1)(x-1) - (t-1)(x-t)(x+1) = (x-t)^2 + (x-t) + t(t-1) \ge 0.$$

Therefore,

$$\operatorname{Var}\left(\sum_{i=3}^{j-1} Y_{i,j}^{(k)}\right) \le \sum_{i=3}^{j-1} \operatorname{Var}(Y_{i,j}^{(k)}) \sim (j-3)(\delta(1-\delta) - \delta^2(1-\delta)^2).$$

Putting everything together, we may bound

$$\mathbb{P}(X_j^{(k)}) \lesssim \frac{\delta^2(j-3)(\delta(1-\delta) - \delta^2(1-\delta)^2)}{(j-3)^2\delta^2(1-\delta)^2} = \frac{\delta - \delta^2 + \delta^3}{(j-3)(1-\delta)}.$$

When $2j-1 > n$, everything is the same except we replace the upper bound $j-1$ with $n-j$, so that

$$\mathbb{P}(X_j^{(k)}) \lesssim \frac{\delta - \delta^2 + \delta^3}{(1-\delta)(n-j-2)},$$

as desired. The only minor detail is that negative correlation still holds, which is not an issue under the transformation $j-1 \mapsto n-j$, as all calculations are equivalent. □

This gives us the following theorem:

Theorem 8.3. *Let $\delta = k/n$. When*

$$\frac{\delta - \delta^2 + \delta^3}{1-\delta} \ll \frac{1}{\log n},$$

we have that $s_{n,k} = (1-o(1))\binom{n}{k-1}$.

Proof. Assume k is not $O(1)$, this has been resolved earlier. Note that we have $\delta = o(1)$, so with $1-o(1)$ probability the minimum element is $o(n)$. We can effectively assume the minimum is 1 for probability computations, then. Quickly compute $\mathbb{P}(X_3^{(k)}) \sim \delta^2$ and $\mathbb{P}(X_4^{(k)}) \sim \delta^2(1-\delta(1-\delta)) = \delta^2 - \delta^3 + \delta^4$. Then by our typical union-bound calculation,

$$s_{n,k} \ge (1-o(1))\binom{n}{k-1}\left(1 - \sum_{j=3}^{n} \mathbb{P}(X_j^{(k)})\right).$$

Note that

$$\sum_{j=5}^{n} \mathbb{P}(X_j^{(k)}) \lesssim \frac{\delta - \delta^2 + \delta^3}{1-\delta}(2H_n),$$

so the result follows by the fact that $H_n \sim \log n$ and the fact that the inequality in question implies $\delta = o(1)$. □

This reduces to the following wonderful corollary:

Corollary 8.1. *If $k \ll \frac{n}{\sqrt{\log n}}$, $\alpha_{n,k} = (1 - o(1))\binom{n}{k-1}$.*

Computing the probabilities of these events is difficult but still possible via generating functions. We have the following theorem:

Theorem 8.4. *We have that*

$$\mathbb{P}(X_j^{(k)}) = \begin{cases} \binom{n-2}{k-1}^{-1} \sum_{i=0}^{j-3} \binom{j-3}{i}\binom{n-j-i-1}{k-3-2i} & 2j \leq n \\ \binom{n-2}{k-1}^{-1} \sum_{i=0}^{n-j-2} \binom{n-j-2}{i}\binom{j-i-2}{k-3-2i} & 2j > n \end{cases}$$

Proof. Generating Functions. □

Theorem 8.5. *We have that*

$$s_{n,k+1} \geq \binom{n-2}{k-1}\left(1 - S_1 - S_2\right),$$

where

$$S_1 = \sum_{j=3}^{\lfloor n/2 \rfloor} \binom{n-2}{k-1}^{-1} \sum_{i=0}^{j-3} \binom{j-3}{i}\binom{n-j-i-1}{k-3-2i},$$

$$S_2 = \sum_{j=\lfloor n/2 \rfloor+1}^{n} \binom{n-2}{k-1}^{-1} \sum_{i=0}^{n-j-2} \binom{n-j-2}{i}\binom{j-i-2}{k-3-2i}.$$

Proof. Same as the computation of s_n just adapted. □

This is pretty difficult to assess in general, but there are some introductory facts we can prove about $\mathbb{P}(X_j^{(k)})$ to make this more tractable:

Lemma 8.3. *The sequence $(\mathbb{P}(X_j^{(k)}))_{k=0}^n$ is unimodal with peak $n/2$.*

Proof. Annoying. □

This fact, coupled with the following, allows us to bound $\mathbb{P}(X_j^{(k)})$ in a useful manner:

Lemma 8.4. *For any k, $\mathbb{P}(X_j^{(k)}) = \mathbb{P}(X_j^{(n-k-3)})$.*

Proof. Annoying. □

Now, we can upper bound $\mathbb{P}(X_j^{(k)})$. We can use this to get a weaker but more tractable bound on $s_{n,k}$.

Conjecture 8.2. *For all n, $s_{n,k}$ and $\alpha_{n,k}$ are both unimodal.*

Let us derive the constant mentioned in our conjecture. As mentioned above, the question of computing s_n should be attacked probabilistically with a similar scheme as in [1]. The first condition, $2m \notin S$, should approximately restrict us by a factor of $\frac{1}{2}$. It is not too important. Then

$$s_n \approx \frac{1}{2} \cdot 2^n \left(1 - \mathbb{P}\left(\bigcup_{i=1}^{n} X_i\right)\right).$$

This can be made exact easily but let us not worry about that for now. Now, X_i and X_j are not independent (generally), but we do intuitively expect them to be *roughly* independent (small covariance). With this knowledge, one should expect

$$s_n \approx \frac{1}{2} \cdot 2^n \prod_{k \geq 1} \left(1 - \frac{1}{2^k}\right) \approx 0.144 \cdot 2^n.$$

This is promising evidence for the truth of 8.1. Perhaps a graph theoretic scheme like in [1] (which is common among many papers on sumsets, as mentioned) will give us better bounds. Indeed, the Cayley Sum Graph seems like a useful structure to study here. Perhaps one can use the fact that S is strongly irreducible implies $(s+S) \cap S \neq \emptyset$ for all $s \in S$ and a similar graph theoretic approach. I'm not sure how good the upper bound will be here, though. One may also try to use the Lovasz Local Lemma, this seems to be viable. One should also expect the inequality

$$s_{n,k} \geq \binom{n}{k-1} \prod_{t \geq 1} \left(1 + \left(\frac{k}{n}\right)^t\right) = \frac{n}{n+k} \binom{n}{k-1}$$

to hold. This comes from assuming independence of the events that each element has a nuller. We make this a conjecture:

Conjecture 8.3. *For all n, k:*

$$s_{n,k} \geq \frac{n}{n+k} \binom{n}{k-1}.$$

This lower bound would be quite nice to prove, as it would imply unimodality for many k. In particular, if $k = \delta n$,

$$\frac{\binom{n}{k-1}}{\binom{n}{k-2}} \sim \frac{1-\delta}{\delta}.$$

The inequality

$$\frac{1}{1+\delta} > \frac{1-\delta}{\delta}$$

is true when $\delta > \frac{\sqrt{5}-1}{2}$, so this lower bound would imply unimodality for $\left(\frac{3-\sqrt{5}}{2} - o(1)\right) n$ values of k. This is more than $0.38n$ values!

Let us focus on specific cases of strongly irreducible sets. We will count the sets by looking at the set of nullers. A very basic example of this gives us the following lower bound on $\alpha_{n,k}$:

Theorem 8.6. *We have*

$$\alpha_{n,k} \geq \max_d \sum_{a_1 + a_2 + \cdots + a_d = k} \prod_{i=1}^{d} \binom{\lfloor (n - 2d - i)/d \rfloor - a_i}{a_i}.$$

Proof. Consider the strongly irreducible sets with the set of nullers $\{d\}$. Consider the binary sequence of indicator variables for elements of n, whether they are in the set or not. Truncate this sequence to only contain indices over $2d$. Split this sequence into d distinct sequences, based on residue classes modulo d. Then let a_i be the number of ones in the ith class. The result follows by noticing that each class can be anything such that there are no two consecutive ones, and then using a well-known result. □

The above bound is probably not that good, because it is easy to calculate the same way that the number of strongly irreducible subsets of $[n]$ with set of nullers $\{d\}$ is $O((\varphi + \varepsilon)^n)$ for any $\varepsilon > 0$.

Let us consider sets with a more general set of nullers $D = \{d_1, d_2, \ldots, d_t\}$.

Definition 8.3. We say that a set S is *nulled* by a subset $D \subseteq S$ if S is strongly irreducible, and the nuller of every element in S is in D.

We have the following nice theorem:

Theorem 8.7. *Let $D = \{d_1, d_2, \ldots, d_t\}$ with $0 < d_1 < d_2 < \cdots < d_t < n$. Then there are at least*

$$2^{n-2d_t+1} - (n - 3d_t + 1)2^{n-2d_t-t+1}$$

subsets of $[n]$ containing 0 nulled by D.

Proof. We count the number of binary sequences of length $n - 2d_t + 1$ that avoid subsequences

$$(a_{n+d_1}, a_{n+d_2}, \ldots, a_{n+d_t}) = (1, 0, 0, \ldots, 0).$$

We approach this task probabilistically. Let X_D be the number of "bad" subsequences. Then by Markov's inequality,

$$\mathbb{P}(X_D \geq 1) \leq \mathbb{E}[X_D].$$

By linearity of expectation, we can compute $\mathbb{E}[X_D] = \frac{n-3d_t+1}{2^t}$. Therefore, we have the lower bound

$$\mathbb{P}(X_D = 0) \geq 1 - \frac{n - 3d_t + 1}{2^t}.$$

Therefore, there are at least

$$n_D = 2^{n-2d_t+1}\mathbb{P}(X_D = 0) = 2^{n-2d_t+1} - (n - 3d_t + 1)2^{n-2d_t-t+1}$$

strongly irreducible sets nulled by D, as desired. □

Corollary 8.2. *There are at least $2^{n-O(\ln n)}$ strongly irreducible subsets of $[n]$ containing 0. In other words, $s_n \geq 2^{n-O(\ln n)}$.*

Tracking calculations more precisely, we have the following result:

Corollary 8.3. *For any $\varepsilon > 0$,*

$$\frac{s_n}{\alpha_n} \gg n^{-2+\varepsilon}.$$

Of course, this was already known, but it's interesting to see.

Remark 8.4. *Lemma 6.5 of [1] upper bounds $\mathbb{P}(|X_D - (n - 3d_t + 1)2^{-d_t}| > \delta(n - 2d_t + 1))$ by $\exp(-c(n - 2d_t + 1))$ for some positive constant c.*

The wonderful thing about this argument is it is not hard to incorporate k into the mix:

Theorem 8.8. *Let $D = \{d_1, d_2, \ldots, d_t\}$ with $0 < d_1 < d_2 < \cdots < d_t < n$. Then there are at least*

$$\binom{n - 2d_t}{k - d_t}\left(1 - (n - 3d_t + 1)\frac{\binom{n-k-d_t}{k-d_t}}{\binom{n-2d_t}{k-d_t}}\right)$$

subsets of $[n]$ containing 0 nulled by D, of size $k + 1$.

Proof. The proof is the same, this time we need to compute $\mathbb{E}[X_D]$ more carefully, though. We again use linearity of expectation. The probability that some particular (a_{n+d_i}) sequence is $(1, 0, 0, \ldots, 0)$ is

$$\frac{\binom{n-k-d_t}{k-d_t}}{\binom{n-2d_t}{k-d_t}}.$$

Hence we have

$$\mathbb{E}[X_D] = (n - 3d_t + 1)\frac{\binom{n-k-d_t}{k-d_t}}{\binom{n-2d_t}{k-d_t}},$$

and therefore the amount of sets in question is at least

$$\binom{n-2d_t}{k-d_t}\left(1 - (n - 3d_t + 1)\frac{\binom{n-k-d_t}{k-d_t}}{\binom{n-2d_t}{k-d_t}}\right).$$

□

Corollary 8.4. *We have that*

$$\alpha_{n,k} \geq s_{n,k} \geq \max_{1<d_t<k} \binom{n-2d_t}{k-d_t-1}\left(1 - (n - 3d_t + 1)\frac{\binom{n-k-d_t-1}{k-d_t-1}}{\binom{n-2d_t}{k-d_t-1}}\right).$$

This is a bit intractable asymptotically, but I promise that the bound is quite good. We have the following corollary:

Corollary 8.5. *Let k_0 be the minimum positive integer k such that*

$$\frac{(n-k-1)!^2}{(n-2)!(n-2k)!} < \frac{1}{n-3}.$$

For all $k \gg k_0$, $\alpha_{n,k} \geq (1 - o(1))\binom{n-2}{k-2} = \Theta\left(\binom{n}{k-1}\right)$.

In other words, this gives us the bound we want for k large enough.

9 Upper bounds on $\alpha_{n,k}$

To upper bound $\alpha_{n,k}$, we lower bound $\beta_{n,k} = \binom{n}{k-1} - \alpha_{n,k}$. We do this in the same way that the current best bound for β_n has been achieved: By looking at multiples of $\{0, d\}$. We have the following lemma:

Lemma 9.1. *A set S is a multiple of $\{0, d\}$ iff $i \in S \implies i - d \in S$ or $i + d \in S$.*

Proof. Easy. □

Now, biject subsets of $[n]$ with binary strings of length n representing the indicator variables $\mathbb{1}_{i\in S}$ for $1 \leq i \leq n$. Then we have the following:

Definition 9.1. Let $\gamma_{n,k}^{(d)}$ be the number of binary strings a of length n with exactly k ones such that $a_i = 1 \implies a_{i-d} = 1$ or $a_{i+d} = 1$.

Lemma 9.2.

$$\beta_{n,k} \geq \max_d \gamma_{n,k}^{(d)}$$

Proof. Easy. □

Now, one can actually compute $\gamma_{n,k}^{(d)}$ as a coefficient of a bivariate generating function:

Theorem 9.1. *Let $f(a,b)$ be the number of binary strings with exactly a ones and exactly b zeroes such that every 1 is adjacent to another 1. Then*

$$\gamma_{n,k}^{(d)} = \sum_{a_1+a_2+\cdots+a_d=k} \prod_{i=1}^{d} f(a_i, \lfloor (n-i)/d \rfloor - a_i),$$

where the outer summation is over all $(a_1, a_2, \ldots, a_d) \in \mathbb{N}_0^d$ with sum k.

Proof. One can decompose a binary string $\sigma = (\sigma_1, \sigma_2, \ldots, \sigma_n)$ into d disjoint subsequences: for each $1 \leq i \leq d$, set $i_\sigma = (\sigma_{i+d\ell})$ where ℓ ranges over all naturals such that $i + d\ell \in [1, n]$. Suppose the ith subsequence has a_i ones. It is then clear that the number of ways the ith subsequence can exist is $f(a_i, \lfloor (n-i)/d \rfloor - a_i)$, from which the result follows. □

The goal is to now estimate $f(a,b)$. We can do this by first finding the corresponding generating function:

Theorem 9.2. *We have that*

$$F(x,y) = \sum_{a \geq 0} \sum_{b \geq 0} f(a,b) x^a y^b = \frac{xy - x - 1}{x^2 y - xy + x + y - 1}.$$

Proof. One can derive the recursive relation

$$f(a,b) = f(a,b-1) + f(a-2,b-1) + f(a-3,b-1) + f(a-4,b-1) + \cdots + f(0,b-1)$$

for $a \geq 2, b \geq 1$, by considering the last digits of the binary string as $0, 011, 0111$, etc. One may write

$$f(a-1,b) = f(a-1,b-1) + f(a-3,b-1) + f(a-4,b-1) + \cdots$$

which implies

$$f(a,b) = f(a,b-1) + f(a-2,b-1) + f(a-1,b) - f(a-1,b-1),$$

an equality that holds for $a \geq 2$ and $b \geq 1$. One can trivially compute the following base cases:

- $f(a,0) = 1$.
- $f(0,b) = 1$.
- $f(1,b) = 0$ for $b > 0$.

Now write

$$\sum_{a \geq 0} \sum_{b \geq 0} f(a,b) x^a y^b - \sum_{a \geq 2} \sum_{b \geq 1} f(a,b) x^a y^b = -f(1,1)xy - f(0,0) + \sum_{a \geq 0} f(a,0) x^a + \sum_{b \geq 0} (f(1,b) xy^b + f(0,b) y^b)$$

$$= -1 + \frac{1}{1-x} + \frac{1}{1-y} + x.$$

For convenience let

$$G(m,n) = \sum_{a \geq m} \sum_{b \geq n} f(a,b) x^a y^b.$$

Now compute using our recurrence

$$G(2,1) = yG(2,0) + x^2 yG(0,0) + xG(1,1) - xyG(1,0) = G(0,0) + 1 - x - \frac{1}{1-x} - \frac{1}{1-y}.$$

Let us compute the following:

- $G(1,0) = G(0,0) - \frac{1}{1-y}$

- $G(1,1) = G(0,0) - \frac{1}{1-y} - \frac{1}{1-x} + 1$
- $G(2,0) = G(0,0) - \frac{1}{1-y} - x$

Substituting and solving,

$$F(x,y) = G(0,0) = \frac{xy - x - 1}{x^2 y - xy + x + y - 1},$$

as desired. □

The goal is to now somehow extract coefficients from this generating function. The trick to do this is quite clever: Observe the identity

$$-\frac{F(x,y)}{xy - x - 1} = \frac{1}{(1-x)(1-y)} \cdot \frac{1}{1 - \frac{x^2 y}{(1-x)(1-y)}}.$$

From this, we may write by standard generating function identities

$$\begin{aligned}
\frac{F(x,y)}{1 + x - xy} &= \sum_{k \geq 0} \frac{x^{2k}}{(1-x)^{k+1}} \cdot \frac{y^k}{(1-y)^{k+1}} \\
&= \sum_{k \geq 0} \left(\sum_{a \geq 0} \binom{a-k}{k} x^a \right) \left(\sum_{b \geq 0} \binom{b}{k} y^b \right) \\
&= \sum_{a \geq 0} \sum_{b \geq 0} \sum_{k \geq 0} \binom{a-k}{k} \binom{b}{k} x^a y^b.
\end{aligned}$$

In turn, we have that

$$f(a,b) = \sum_{k \geq 0} \left(\binom{a-k}{k} \binom{b}{k} + \mathbb{1}_{a \geq 1} \binom{a-1-k}{k} \binom{b}{k} - \mathbb{1}_{a \geq 1} \mathbb{1}_{b \geq 1} \binom{a-1-k}{k} \binom{b-1}{k} \right).$$

Corollary 9.1. *We have that*

$$\beta_{n,k} \geq \max_d \sum_{a_1 + a_2 + \cdots + a_d = k} \prod_{i=1}^{d} \sum_{\ell \geq 0} \binom{a_i}{\ell} \binom{\lfloor (n-i)/d \rfloor - a_i - \ell}{\ell}$$

and

$$\alpha_{n,k} \leq \binom{n}{k-1} - \max_d \sum_{a_1 + a_2 + \cdots + a_d = k} \prod_{i=1}^{d} \sum_{\ell \geq 0} \binom{a_i}{\ell} \binom{\lfloor (n-i)/d \rfloor - a_i - \ell}{\ell}.$$

Nice! This gives us the following theorem:

Theorem 9.3. *Suppose $k = \delta n$. Let $p, q \in (0,1)$ solve the equations*

$$\frac{p - p^2 + p^3}{(1-p)^2 (1+p)} = \frac{1-\delta}{\delta},$$

$$(1-p)^2 (1-q) = pq.$$

Then

$$\log \beta_{n,k} \geq -n\delta \log(1-p) - n(1-\delta) \log(1-q) - O(\log n).$$

Proof. Set all $a_i \sim \frac{n\delta}{d}$ and apply Petrov's asymptotic for the summation

$$\sum_{k\geq 0} \binom{a}{k}\binom{b-k}{k}.$$

The asymptotics will be put in the appendix. □

Corollary 9.2. *We have $\beta_{n,n/2} \geq \Omega(3^{n/2})$ and hence $\alpha_{n,n/2} \leq \binom{n}{n/2} - \Omega(3^{n/2})$. Additionally,*

$$\alpha_{n,0.6n} \leq \binom{n}{0.6n} - \exp(0.562n),$$

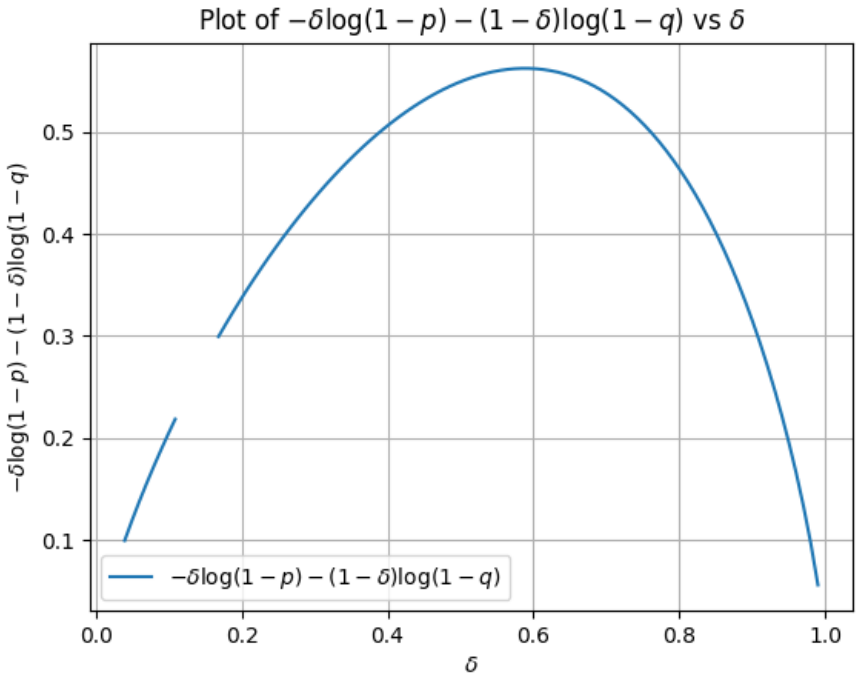
and $\binom{n}{0.6n} \sim \exp(0.67n)$.

Compare this asymptotic with the fact that $\binom{n}{n/2} \sim \frac{2^n}{\sqrt{n}}$.

Corollary 9.3. *When $k = \Omega(n)$,*

$$\alpha_{n,k} \leq \binom{n}{k-1} - \exp(\Omega(n)).$$

One can graph the relationship between δ and the factor $c(\delta)$ such that our bound is $\log \beta_{n,k} \geq c(\delta)n$:



This bound should be quite good, nearly sharp. This is due to a paper by Granville on the number of sunsets.

Remark 9.1. *Fix so that we account for the extra term in $f(a,b)$, and use Pascal. The bound is better!*

We can prove without all this nonsense the following:

Theorem 9.4. *If $k = n - o(\sqrt{n})$ then $\alpha_{n,k} = \binom{n}{k-1}o(1)$.*

Proof. Its almost always a multiple of $\{0,1\}$. □

10 Miscellaneous Things

This work is essentially about what size means for irreducible sumsets, so it is important to consider the question: *What is the distribution of $|A+B|$ for random sets A, B?*. This allows us to get a sense for which k the reducibles "mass towards". We cannot make anything explicit, however, because there is not unique factorization. This problem is studied for $B = A$ and the case where B is correlated with A in [4]. In our case, A and B are independent and hence the problem is much easier. We may obtain explicit results.

There are two main frameworks to consider:

- A, B are random subsets of $[n]$, each described by n indicator variables.
- A, B are chosen randomly over all pairs of sets (A', B') with $A' + B' \subseteq [n]$.

The former is less messy to analyze, but we will see that it is not that interesting. The latter is more closely related to the general focus of this work. We have the following lemma:

Lemma 10.1. *In the former case, $\mathbb{E}[|A+B|] = 2n - \frac{1}{3} - \frac{1}{3\cdot 16^n}$.*

Proof. Note

$$\mathbb{P}(x \in A+B) = 1 - \mathbb{P}(x \notin A, B) \prod_{i+j=x i,j>0} \mathbb{P}((i,j) \notin A \times B) = 1 - \frac{1}{4^x}.$$

Therefore by linearity of expectation,

$$\mathbb{E}[|A+B|] = \sum_{x=0}^{2n} \mathbb{E}[\mathbb{1}_{x\in A+B}] = \sum_{x=0}^{2n} \left(1 - \frac{1}{4^x}\right) = 2n - \frac{1}{3} - \frac{1}{3\cdot 16^n}.$$

□

In other words, $A+B$ tends to be quite large. This is expected, as $|A| + |B| - 1 \leq |A+B|$. This case seems to be studied quite a bit, so we shift our focus to the latter case. We can study this to some extent, and bring results about divisors in from lunar arithmetic stuff.

We may ask a few more miscellaneous questions:

Conjecture 10.1. *Goldbach's conjecture holds: every non-irreducible A can be written as $B + C$ for irreducible B, C.*

The Schirellmann density of such sets is 1, so this is almost always true. To see that it is always true requires more work, though.

It is also interesting to study primes in this monoid. Here primes are different from irreducibles, an element A is prime if $A \mid B + C$ implies $A \mid B$ or $A \mid C$. We write $A \mid B$ meaning $B = A + D$ for some D. We have a few basic results:

Lemma 10.2. *If A is prime, A must be irreducible.*

Proof. Suppose otherwise, that $A = A_1 + A_2$. Let $D = A_1 - \{\max(A_1)\}$. Then

$$A + D = A_1 + (A_2 + D).$$

Clearly $A \nmid A_1, A_2 + D$ by size. We're done if $|A_1| > 2$ or $|A_2| > 2$. The case left to resolve is when $A = \{0, a\} + \{0, b\} = \{0, a, b, a+b\}$. We must show that this is not prime. Note that

$$\{0, a, b, a+b\} + \{0, c\} = \{0, a\} + (\{0, b\} + \{0, c\}) = \{0, a\} + \{0, b, c, b+c\},$$

so taking $c < a$ proves the statement if $a \neq 1$. If $a = 1$ but $b \neq 1$, the same statement with $c \neq a$ does the job. It suffices now to show that $\{0, 1, 2\}$ is not prime. This follows because

$$\{0, 1, 2\} + \{0, 2\} = \{0, 1\} + \{0, 2, 3\}.$$

□

There also exist irreducibles that are not prime:

Example 10.1. $\{0, 2\}$ is a non-prime irreducible, because $\{0, 2\} + \{0, 2, 3\} = \{0, 1\} + \{0, 2, 3, 4\}$ and it can be checked that $\{0, 2\} \nmid \{0, 2, 3, 4\}$.

Example 10.2. $\{0, d, d + 1\}$ is a non-prime irreducible for $d > 1$ because $\{0, d, d + 1\} + \{0, 2\} = \{0, d, d + 2\} + \{0, 1\}$.

It turns out that there are no primes, as shown in [1]... Whoops!

11 Next Steps/Current Work

- Make corollary 7.2 precise by doing the dirty work (whole section more precise in general).
- Finish improving bounds on $s_{n,k}$, order of probabilities and genfunc.
- New notion instead of strong irreducibility to capture more dense sets? More specific families besides the sparse ones captured by strong irreducibility.
- Think about how many times a certain composite set gets hit, i.e. the number of divisors it has. Cite dismal/lunar arithmetic, various papers talk about number of divisors in these systems which transfers over.
- Cite Granville paper for non-improvement of upper bound. Remark that the lower bound c is probably sharp, so upper bound that we do is also probably sharp.
- Ask more general questions about the system. Are there primes (in the strict sense of the definition)? Does Goldbach hold?
- Write everything up all nice.

References

[1] P. Bienvenu and A. Geroldinger. On Algebraic Properties of Power Monoids of Numerical Monoids. *Israel Journal of Mathematics*, 2023.

[2] Yaroslav Shitov. How many boolean polynomials are irreducible? *Int. J. Algebra Comput.*, 24:1183–1190, 2014.

[3] Ki Hang Kim and Fred W. Roush. Factorization of polynomials in one variable over the tropical semiring, 2005.

[4] Hung V. Chu, Dylan King, Noah Luntzlara, Thomas C. Martinez, Steven J. Miller, Lily Shao, Chenyang Sun, and Victor Xu. Generalizing the distribution of missing sums in sumsets, 2021.