Limiting Behavior in Missing Sums of Sumsets

Rauan Kaldybayev
rk19@williams.edu

Chris Yao
chris.yao@yale.edu

Joint work with Aditya Jambhale, Advised by Steven Miller

Young Mathematicians Conference, August 16, 2023

Introduction

Given $A \subseteq \mathbb{Z}$, define its sumset

- $A + A := \{a_1 + a_2 \mid a_1, a_2 \in A\}$.

Sumsets are fundamental objects in number theory.

Introduction

Given $A \subseteq \mathbb{Z}$, define its sumset

- $A + A := \{a_1 + a_2 \mid a_1, a_2 \in A\}$.

Sumsets are fundamental objects in number theory.

- Fermat's Last Theorem: if G_k is the set of k^{th} powers of $\mathbb{Z}_{>0}$, $(G_k + G_k) \cap G_k = \emptyset$ for $k > 2$.
- Goldbach Conjecture: for the set of primes P, $P + P \supseteq \{4, 6, 8, \ldots\}$.

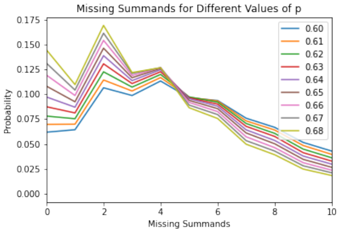
Setting

- Fix $N \geq 0$. Fix $p \in (0, 1)$, and let $q := 1 - p$.
- Select $A \subseteq [0, N]$ by a Bernoulli process: for each $k \in [0, N]$, independently include k in A with probability p.

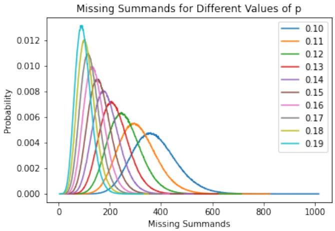
Setting

- Fix $N \geq 0$. Fix $p \in (0, 1)$, and let $q := 1 - p$.
- Select $A \subseteq [0, N]$ by a Bernoulli process: for each $k \in [0, N]$, independently include k in A with probability p.
- Recent research in $|A + A|$ as a random variable.
- Martin and O'Bryant's formative paper [MO] compared $|A + A|$ to $|A - A|$ when $p = 1/2$.

Numerically observed behavior



(a) Large p



(b) Small p

Figure: Point distribution function $\mathbb{P}\left(|(A+A)^c| = m\right)$ for several values of p.

Motivating Questions

- What is the decay rate for the distribution of $|(A+A)^c| := |[0,2N]\backslash(A+A)|$?

Motivating Questions

- What is the decay rate for the distribution of $|(A+A)^c| := |[0,2N]\backslash(A+A)|$?
- What are $\mathbb{E}\left(|A+A|^c\right)$ and $\operatorname{Var}\left(|A+A|^c\right)$?

Motivating Questions

- What is the decay rate for the distribution of $|(A+A)^c| := |[0,2N]\backslash(A+A)|$?
- What are $\mathbb{E}\left(|A+A|^c\right)$ and $\mathrm{Var}\left(|A+A|^c\right)$?
- Why are the "divots" here?

Prior Work

Theorem (Martin and O'Bryant '06)

If $p = \frac{1}{2}$, *then* $\mathbb{E}[|(A+A)^c|] = 10 + O((3/4)^{N/2})$.

Prior Work

Theorem (Martin and O'Bryant '06)

If $p = \frac{1}{2}$, then $\mathbb{E}[|(A+A)^c|] = 10 + O((3/4)^{N/2})$.

Theorem (Lazarev, Miller, and O'Bryant '13 [LMO])

If $p = \frac{1}{2}$, then for $i < j \leq N$ with i, j odd,

$$\mathbb{P}(i \text{ and } j \notin A+A) = \frac{1}{2^{j+1}} F_{q+2}^{r} F_{q+4}^{r'}$$

for q, r, r' depending on i and j, and similar formulations hold for the other 3 parity cases.

Prior Work

- In general, when $p \neq 1/2$, not all sets A are equally likely, which makes the analysis harder.

Prior Work

- In general, when $p \neq 1/2$, not all sets A are equally likely, which makes the analysis harder.
- Chu, King, Luntzlara, Martinez, Miller, Shao, Sun, and Xu [CKLMMSSX] study sumsets for generic p.
- [CKLMMSSX] and [LMO] both use graph-theoretic approaches, particularly the notion of a *condition graph*.

Prior Work

Theorem (King, Martinez, Miller, Sun '19)

For $p \in [0,1]$ and $q := 1-p$,

$$\mathbb{E}[|A+A|] = \sum_{r=0}^{n} p^r q^{n-r} \binom{n}{r} \left(2 \sum_{k=0}^{n-1} \left(1 - \frac{f(k)}{\binom{n}{r}} \right) - \left(1 - \frac{f(n-1)}{\binom{n}{r}} \right) \right),$$

where $n = N+1$ and

$$f(k) = \begin{cases} \sum_{i=\frac{k+1}{2}}^{k+1} 2^{k+1-i} \binom{\frac{k+1}{2}}{i-\frac{k+1}{2}} \binom{n-k-1}{r-i} & \text{for } k \text{ odd} \\ \sum_{i=\frac{k}{2}}^{k} 2^{k-i} \binom{\frac{k}{2}}{i-\frac{k}{2}} \binom{n-k-1}{r-1-i} & \text{for } k \text{ even.} \end{cases}$$

In particular, where the LHS holds for $p > \frac{1}{2}$,

$$2n-1-2\,\frac{1}{1-\sqrt{2q}} - (2q)^{\frac{n-1}{2}} \leq \mathbb{E}[|A+A|] \leq 2n-1-2\,\frac{1-q^{\frac{n-1}{2}}}{1-\sqrt{q}}.$$

Prior Work

Theorem (King, Martinez, Miller, Sun '19)

For $p \in (0,1)$ and $q := 1-p$,

$$\begin{aligned}\mathrm{Var}(|A+A|) &= \sum_{r=0}^{n}\binom{n}{r}p^r q^{n-r} \\ &\quad \times \left(2\sum_{0\le i<j\le 2n-2} 1 - P_r(i,j) + \sum_{0\le i\le 2n-2} 1 - P_r(i)\right) \\ &\quad - \mathbb{E}[|A+A|]^2,\end{aligned}$$

where $n = N+1$,

$$P_r(i) = \mathbb{P}(i \notin A+A \mid |A| = r),$$

and

$$P_r(i,j) = \mathbb{P}(i \text{ and } j \notin A+A \mid |A| = r).$$

The Infinite Case

Setup

- Instead of considering $A \subseteq [0, N]$ for some natural number N, consider $\mathbb{A} \subseteq \mathbb{Z}_{\geq 0}$ chosen randomly via a Bernouli process.
- For any $k \in \mathbb{Z}_{\geq 0}$, include k in $\mathbb{A}$ with probability p.

Setup

- Instead of considering $A \subseteq [0, N]$ for some natural number N, consider $\mathbb{A} \subseteq \mathbb{Z}_{\geq 0}$ chosen randomly via a Bernouli process.
- For any $k \in \mathbb{Z}_{\geq 0}$, include k in $\mathbb{A}$ with probability p.
- With probability 1, $\mathbb{A}$ and $\mathbb{A}^c$ both include infinitely many elements.
- How do $\mathbb{A} + \mathbb{A}$ and $\mathbb{A} - \mathbb{A}$ behave?

Motivation

- In general, sum sets are “almost full” in the middle, and missing elements are only on the fringes. In the “infinite case,” there is only one fringe to worry about.

Motivation

- In general, sum sets are "almost full" in the middle, and missing elements are only on the fringes. In the "infinite case," there is only one fringe to worry about.

- Having "$N = \infty$" allow us to properly use asymptotic notation for the number of missing summands.

Motivation

- In general, sum sets are "almost full" in the middle, and missing elements are only on the fringes. In the "infinite case," there is only one fringe to worry about.
- Having "$N = \infty$" allow us to properly use asymptotic notation for the number of missing summands.
- Studying the "infinite case" will help us understand the "finite case."

Analysis of $\mathbb{A} - \mathbb{A}$

Proposition

With probability 1, $\mathbb{A} - \mathbb{A} = \mathbb{Z}$.

Analysis of $\mathbb{A} - \mathbb{A}$

Proposition

With probability 1, $\mathbb{A} - \mathbb{A} = \mathbb{Z}$.

Proof.

- For $j \in \mathbb{Z}_{\geq 0}$, each pair $(j, 0), (2j + 1, j + 1)$, $(3j + 2, 2j + 2), \dots$ has probability p^2 of occurring.
- Second Borel-Cantelli lemma: with probability 1, infinitely many of these pairs are in $\mathbb{A}$.
- Reverse pairs to get $-j$.

$\square$

Analysis of $\mathbb{A} + \mathbb{A}$

- Unlike $\mathbb{A} - \mathbb{A}$, where there are infinitely many pairs that can lead to j, there are only finitely many for $\mathbb{A} + \mathbb{A}$.
- To check if $n \in \mathbb{A} + \mathbb{A}$, only need to know about the first $n + 1$ elements: $\{0, 1, 2, \ldots, n\}$.
- Focus on the sumset $\mathbb{A} + \mathbb{A}$.

Expected Value

Probability of Missing a Specific Summand

- Define $\mathbb{Y} := |\mathbb{Z}_{\geq 0} \backslash (\mathbb{A} + \mathbb{A})|$, the number of missing summands.

Probability of Missing a Specific Summand

- Define $\mathbb{Y} := |\mathbb{Z}_{\geq 0} \backslash (\mathbb{A} + \mathbb{A})|$, the number of missing summands.
- For each $i \geq 0$, let $\mathbb{X}_i$ be the indicator variable for $i \notin \mathbb{A} + \mathbb{A}$:

$$\mathbb{X}_i := \begin{cases} 1 & i \notin \mathbb{A} + \mathbb{A} \\ 0 & i \in \mathbb{A} + \mathbb{A}. \end{cases}$$

Probability of Missing a Specific Summand

- Define $\mathbb{Y} := |\mathbb{Z}_{\geq 0} \backslash (\mathbb{A} + \mathbb{A})|$, the number of missing summands.
- For each $i \geq 0$, let $\mathbb{X}_i$ be the indicator variable for $i \notin \mathbb{A} + \mathbb{A}$:

$$\mathbb{X}_i := \begin{cases} 1 & i \notin \mathbb{A} + \mathbb{A} \\ 0 & i \in \mathbb{A} + \mathbb{A}. \end{cases}$$

- The upshot is that

$$\mathbb{Y} = \sum_{i=0}^{\infty} \mathbb{X}_i.$$

- To calculate $\mathbb{E}(\mathbb{Y})$, need $\mathbb{E}(\mathbb{X}_i) = \mathbb{P}(i \notin \mathbb{A} + \mathbb{A})$.

Probability of Missing a Specific Summand

Like [LMO], for odd n,

$$\{n \notin \mathbb{A}+\mathbb{A}\} = \{(0 \notin \mathbb{A} \text{ or } n \notin \mathbb{A}) \text{ and } \cdots \text{ and } (\frac{n-1}{2} \notin \mathbb{A} \text{ or } \frac{n+1}{2} \notin \mathbb{A})\}$$

and for even n,

$$\{n \notin \mathbb{A} + \mathbb{A}\} = \{(0 \notin \mathbb{A} \text{ or } n \notin \mathbb{A}) \text{ and } \cdots \text{ and } n/2 \notin \mathbb{A}\}.$$

Probability of Missing a Specific Summand

Like [LMO], for odd n,

$$\{n \notin \mathbb{A}+\mathbb{A}\} = \{(0 \notin \mathbb{A} \text{ or } n \notin \mathbb{A}) \text{ and } \cdots \text{ and } (\frac{n-1}{2} \notin \mathbb{A} \text{ or } \frac{n+1}{2} \notin \mathbb{A})\}$$

and for even n,

$$\{n \notin \mathbb{A}+\mathbb{A}\} = \{(0 \notin \mathbb{A} \text{ or } n \notin \mathbb{A}) \text{ and } \cdots \text{ and } n/2 \notin \mathbb{A}\}.$$

Hence,

$$\mathbb{P}(n \notin \mathbb{A}+\mathbb{A}) = \begin{cases} (1-p^2)^{\frac{n+1}{2}} & n \text{ odd} \\ (1-p)(1-p^2)^{\frac{n}{2}} & n \text{ even.} \end{cases}$$

Calculating $\mathbb{E}(\mathbb{Y})$

- By the Monotone Convergence Theorem,

$$\mathbb{E}(\mathbb{Y}) = \sum_{n=0}^{\infty} \mathbb{E}(\mathbb{X}_n) = \sum_{n \text{ odd}} (1-p^2)^{(n+1)/2} + \sum_{n \text{ even}} (1-p)(1-p^2)^{n/2}.$$

Calculating $\mathbb{E}(\mathbb{Y})$

- By the Monotone Convergence Theorem,

$$\mathbb{E}(\mathbb{Y}) = \sum_{n=0}^{\infty} \mathbb{E}(\mathbb{X}_n) = \sum_{n \text{ odd}} (1-p^2)^{(n+1)/2} + \sum_{n \text{ even}} (1-p)(1-p^2)^{n/2}.$$

Proposition

For $p \in (0,1)$,

$$\mathbb{E}(\mathbb{Y}) = \frac{2}{p^2} - \frac{1}{p} - 1.$$

Higher Moments

A Problem with Dependencies

- To calculate $\mathbb{E}\left(\mathbb{Y}^2\right)$, need $\mathbb{P}\left(i, j \notin \mathbb{A} + \mathbb{A}\right)$.
- Unlike $\mathbb{P}\left(i \notin \mathbb{A} + \mathbb{A}\right)$, $\mathbb{P}\left(i, j \notin \mathbb{A} + \mathbb{A}\right)$ is laden with dependencies.
- Example: $\mathbb{P}\left(0 \notin \mathbb{A} + \mathbb{A}\right) = 1 - p$ and $\mathbb{P}\left(1 \notin \mathbb{A} + \mathbb{A}\right) = 1 - p^2$, but $\mathbb{P}\left(0, 1 \notin \mathbb{A} + \mathbb{A}\right) = 1 - p^2$.
- For higher moments, $\mathbb{E}\left(\mathbb{Y}^k\right)$, even more dependency.

A Workaround

- Instead of an exact expression, we find a bound:

$$\begin{aligned}\mathbb{E}\left(\mathbb{Y}^k\right) &= \sum_{n_1=0}^{\infty} \cdots \sum_{n_k=0}^{\infty} \mathbb{P}(n_1, \dots, n_k \notin \mathbb{A}+\mathbb{A}) \\ &\leq \sum_{n_1=0}^{\infty} \cdots \sum_{n_k=0}^{\infty} \mathbb{P}(\max\{n_1, \dots, n_k\} \notin \mathbb{A}+\mathbb{A}).\end{aligned}$$

- We know the probability of $n \notin \mathbb{A}+\mathbb{A}$:

$$\mathbb{E}\left(\mathbb{Y}^k\right) \leq \sum_{n_1=0}^{\infty} \cdots \sum_{n_k=0}^{\infty} (1-p^2)^{(\max\{n_1,\dots,n_k\}+1)/2}.$$

- Intuitively may not be too much loss; if $\max\{n_1, \dots, n_k\} \notin \mathbb{A}+\mathbb{A}$, many elements are missing from $\mathbb{A}$, so other values are probably also missing from $\mathbb{A}+\mathbb{A}$.

The bound

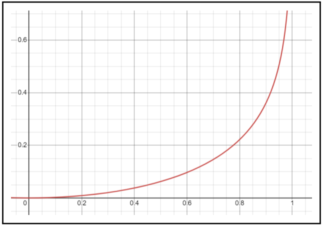
- Evaluating the “almost-geometric” sum yields

$$\mathbb{E}\left(\mathbb{Y}^k\right) \leq \left(1 + \frac{\alpha}{\sqrt{2\pi}}\right)\frac{k!}{\alpha^k},$$

where

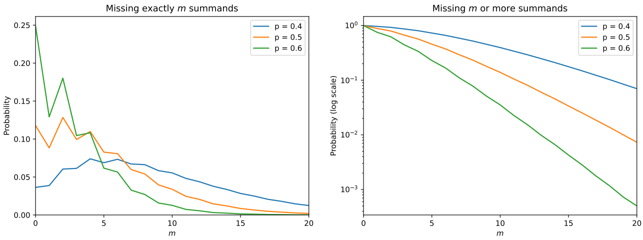
$$\alpha := \log\frac{1}{\sqrt{1-p^2}} = \left|\log\sqrt{1-p^2}\right|.$$

- $O(k!/\alpha^k)$ moments correspond to $f(x) = e^{-\alpha x}$.



The Distribution of $\mathbb{Y}$

Empirical results



- Asymptotic rate of decay seems "approximately exponential."
- $\mathbb{Y}$ appears more likely to be even than odd.
- Question: how does this distribution relate to the "finite case"?

Proving exponential decay

- Since $\mathbb{E}\left(\mathbb{Y}^k\right) = O(k!/\alpha^k)$, Chernoff's inequality yields

$$\mathbb{P}\left(\mathbb{Y} \geq n\right) \;=\; O\left(n\left(1-p^2\right)^{n/2}\right).$$

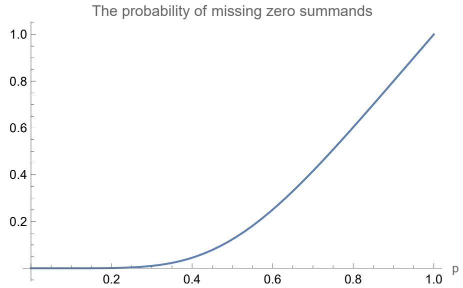
- If $0, \ldots, n/2$ are missing from $\mathbb{A}$, then $0, \ldots, n$ are missing from $\mathbb{A}+\mathbb{A}$. Therefore,

$$\mathbb{P}\left(\mathbb{Y} \geq n\right) \geq (1-p)^{n/2+1}.$$

- Bounded above and below by exponential functions, $\mathbb{P}\left(\mathbb{Y} \geq n\right)$ is "approximately exponential."

Can we determine the details?

- At large, the distribution is exponential.
- Determining the exact details is *very* hard.
- A “simple” question: what is the probability $\mathbb{P}(\mathbb{Y} = 0)$ of missing zero summands, e.g., of having $\mathbb{A} + \mathbb{A} = \mathbb{Z}_{\geq 0}$?



Speculation about $\mathbb{P}(\mathbb{Y} = n)$

- Proposition: $\mathbb{P}(\mathbb{Y} = 0)$ is asymptotically less than every polynomial.
- Conjecture: $\mathbb{P}(\mathbb{Y} = 0) > 0$ for every $p \neq 0$.
- Conjecture: $\mathbb{P}(\mathbb{Y} = 0)$ cannot be an analytic function of p.
- Conjecture: $\mathbb{P}(\mathbb{Y} = n)$ has no closed-form expression in elementary functions.

The Limit of the Finite Case

Review of the finite case

- $A \subseteq [0, N]$ selected at random such that $\mathbb{P}(i \in A) = p$ for all i independently.
- Define $Y := 2N + 1 - |A + A|$ and $X_i := [i \notin A + A]$.
- Object of interest: random variable $Y_{N\to\infty}$,

$$\mathbb{P}(Y_{N\to\infty} = n) := \lim_{N\to\infty} \mathbb{P}(Y = n).$$

- What we will compute: the k-th moment

$$\mathbb{E}\left(Y_{N\to\infty}^k\right) = \lim_{N\to\infty} \mathbb{E}\left(Y^k\right).$$

The k-th moment of Y as a corner sum

- $\mathbb{E}\left(Y^k\right) = \sum_{i_1,\dots,i_k=0}^{2N} \mathbb{E}\left(X_{i_1}\dots X_{i_k}\right)$ is a sum over a k-dimensional hypercube.
- Observation: $A + A$ is “almost full” in the middle.
- Conclusion: To compute $\mathbb{E}\left(Y^k\right)$, we just need to sum over the corners of the hypercube.

Summing over the corners

- Observation: When $j - i > N$, events $i \notin A + A$ and $j \notin A + A$ are independent. Therefore, the corners are independent.
- Result of calculations: the k-th moment of $Y_{N\to\infty}$ is

$$\lim_{N\to\infty} \mathbb{E}\left(Y^k\right) = \sum_{s=0}^{k} \binom{k}{s} \mathbb{E}\left(\mathbb{Y}^s\right) \mathbb{E}\left(\mathbb{Y}^{k-s}\right).$$

The finite case, reduced

- Observation: The moments $\lim_{N\to\infty} \mathbb{E}\left(Y^k\right)$ are the same as those of $\mathbb{Y} + \mathbb{Y}'$. Apply Carleman's condition.

Theorem

The probability distribution of $Y_{N\to\infty}$ is the same as that of $\mathbb{Y} + \mathbb{Y}'$, where $\mathbb{Y}'$ is a copy of $\mathbb{Y}$ independent of it.

- Intuition: Summands can be missing from the left and right fringes, and these are independent for large N.

Future Work

- Use Euler's identity to calculate the even-odd disparity: $\mathbb{P}(\mathbb{Y} \text{ even}) - \mathbb{P}(\mathbb{Y} \text{ odd}) = \mathbb{E}\left(e^{i\pi\mathbb{Y}}\right)$.
- Get tighter bounds on the asymptotic decay rate of $\mathbb{P}(\mathbb{Y} \geq n)$.
- Investigate A^{+k}, the k-th additive power of A, as well as $A^{+\infty} = \{0\} \cup A \cup A^{+2} \ldots$, the set of all possible sums resulting from A.

Acknowledgements

We would like to thank our mentor, Professor Steven J. Miller, and previous years of SMALL for their contributions.

Thanks to our SMALL 2023 faculty, research assistants, and peers for their support.

This presentation was supported by NSF Grants DMS2241623 and DMS2241623. We thank the NSF and Williams College for making SMALL 2023 possible.

Bibliography

- O. Lazarev, S. J. Miller, K. O'Bryant, *Distribution of Missing Sums in Sumsets* (2013), Experimental Mathematics **22**, no. 2, 132–156.
- G. Martin and K. O'Bryant, *Many sets have more sums than differences*, in Additive Combinatorics, CRM Proc. Lecture Notes, vol. 43, Amer. Math. Soc., Providence, RI, 2007, pp. 287–305.
- H. V. Chu, D. King, N. Luntzlara, T. Martinez, S. J. Miller, L. Shao, C. Sun, and V. Xu, *Generalizing the distribution of missing sums in sumsets*, Journal of Number Theory **239** (2022), 402-444