Limiting Behavior in Missing Sums of Sumsets

Rauan Kaldybayev
rk19@williams.edu

Joint work with Aditya Jambhale, Chris Yao
aj644@cam.ac.uk, chris.yao@yale.edu

Advised by Steven Miller
sjm1@williams.edu

Workshop on Combinatorial and Additive Number Theory, May 22, 2024

Introduction

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

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

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.

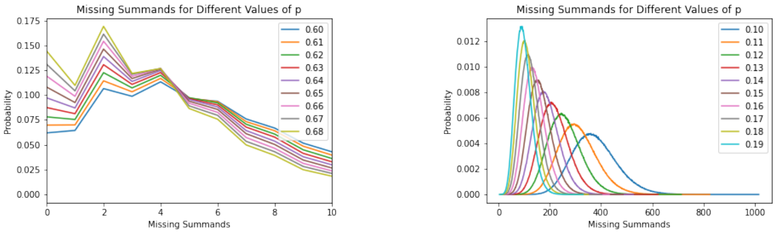
Setting

- 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 seminal paper [MO] compared $|A + A|$ to $|A - A|$ when $p = 1/2$.

Why study sumsets?

- Prove patterns seen from Monte Carlo simulations.
- Might potentially aid other number-theoretic work.

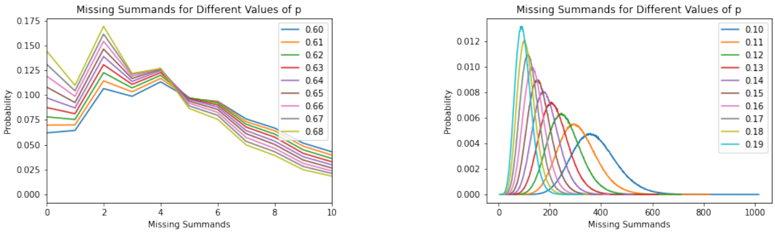
Observed: Divots and Concentration



(a) Large p **(b)** Small p

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

Observed: Divots and Concentration



(a) Large p **(b)** Small p

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

- For large p, missing an even number appears more likely.
- For small p, we see concentration around the mean.

Observed: Exponential Decay

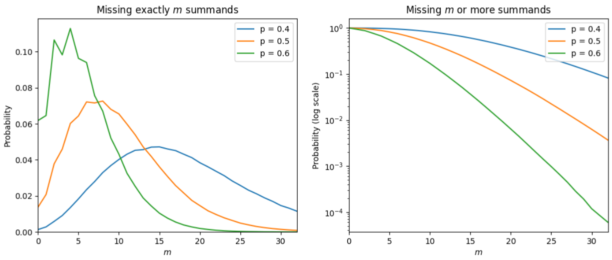


Figure: Point distribution function $\mathbb{P}(|(A+A)^c| = m)$ and cumulative distribution function $\mathbb{P}(|(A+A)^c| \geq m)$ for several values of p, for N very large.

- CDF appears to decay exponentially.

Prior Work: Mean and Variance

Theorem (Martin and O'Bryant '06 [MO])

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

Prior Work: Mean and Variance

Theorem (Martin and O'Bryant '06 [MO])

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: Exponential Decay

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

If $p = \frac{1}{2}$, then

$$m\,(3/4)^{m/2} \ll \mathbb{P}\left(|(A+A)^c| = m\right) \ll (\phi/2)^{m/2} \qquad (1)$$

Prior Work

- When $p \neq 1/2$, not all subsets are equally likely, and previous methods become hard to implement.

Prior Work

- When $p \neq 1/2$, not all subsets are equally likely, and previous methods become hard to implement.
- 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* $q := 1-p$*,*

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

where $n = N+1$*,*

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

and

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

Our Results

- Calculated the mean of $\mathbb{P}\left(|(A+A)^c| = m\right)$ exactly for generic p.
- Calculated the second moment of $\mathbb{P}\left(|(A+A)^c| = m\right)$ to leading order in $1/p$.

Our Results

- Calculated the mean of $\mathbb{P}\left(|(A+A)^c| = m\right)$ exactly for generic p.
- Calculated the second moment of $\mathbb{P}\left(|(A+A)^c| = m\right)$ to leading order in $1/p$.
- Proved concentration in the limit $p \to 0$, thanks to a cancellation of leading terms.

Our Results

- Calculated the mean of $\mathbb{P}\left(|(A+A)^c| = m\right)$ exactly for generic p.
- Calculated the second moment of $\mathbb{P}\left(|(A+A)^c| = m\right)$ to leading order in $1/p$.
- Proved concentration in the limit $p \to 0$, thanks to a cancellation of leading terms.
- Proved exponential bounds for $\mathbb{P}\left(|(A+A)^c| = m\right)$ for generic p.

Our Results

- Calculated the mean of $\mathbb{P}\left(|(A+A)^c| = m\right)$ exactly for generic p.
- Calculated the second moment of $\mathbb{P}\left(|(A+A)^c| = m\right)$ to leading order in $1/p$.
- Proved concentration in the limit $p \to 0$, thanks to a cancellation of leading terms.
- Proved exponential bounds for $\mathbb{P}\left(|(A+A)^c| = m\right)$ for generic p.

This is all in the limit $N \to \infty$.

Our Setup

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 does $\mathbb{A} + \mathbb{A}$ behave?

Motivation for $A \subseteq \mathbb{Z}_{\geq 0}$

- Only one fringe to worry about.

Motivation for $A \subseteq \mathbb{Z}_{\geq 0}$

- Only one fringe to worry about.
- Infinite sums are nice to evaluate.

Motivation for $A \subseteq \mathbb{Z}_{\geq 0}$

- Only one fringe to worry about.
- Infinite sums are nice to evaluate.
- Easy to convert to the original “finite case.”

Motivation for $A \subseteq \mathbb{Z}_{\geq 0}$

- Only one fringe to worry about.
- Infinite sums are nice to evaluate.
- Easy to convert to the original "finite case."
- To check if $n \in \mathbb{A} + \mathbb{A}$, only need to know about the first $n + 1$ elements: $\{0, 1, 2, \ldots, n\}$.

Mean and variance

Probability of Missing a Specific Summand

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

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

- Then

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

- 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,

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

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 $p \in (0,1)$,

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

Probability of Missing Two Specific Summands

- Let $n < m \leq N$.

Probability of Missing Two Specific Summands

- Let $n < m \leq N$.
- Let $l = \lceil \frac{n+1}{m-n} \rceil$ be the “degree of twistedness”.

Probability of Missing Two Specific Summands

- Let $n < m \leq N$.
- Let $l = \lceil \frac{n+1}{m-n} \rceil$ be the "degree of twistedness".

Proposition

If m, n, l are all odd,

$$\mathbb{P}\,(m, n \notin A + A) \;=\; (a_{2l+2})^{\frac{(m+1)-l(m-n)}{2}} (a_{2l})^{\frac{l(m-n)-(n+1)}{2}}.$$

Similar formulas hold for other parities.

Probability of Missing Two Specific Summands

- Let $n < m \leq N$.
- Let $l = \lceil \frac{n+1}{m-n} \rceil$ be the "degree of twistedness".

Proposition

If m, n, l are all odd,

$$\mathbb{P}(m, n \notin A + A) \;=\; (a_{2l+2})^{\frac{(m+1)-l(m-n)}{2}} (a_{2l})^{\frac{l(m-n)-(n+1)}{2}}.$$

Similar formulas hold for other parities.

Here, $a_1 = 1$, $a_2 = 1 - p^2$, and

$$a_k \;=\; (1-p)a_{k-1} + p(1-p)a_{k-2}. \tag{2}$$

$\mathbb{E}\left(\mathbb{Y}^2\right)$ as an infinite sum

Proposition

$$\mathbb{E}\left(\mathbb{Y}^2\right) = -\left(\frac{2}{p^2} - \frac{1}{p} - 1\right) + \\ + 2\sum_{l=1}^{\infty} \frac{a_{2l} + (1-p)a_{l-1} + (1-p)a_l a_{2l} + (1-p)^2 a_l a_{l-1}}{(1-a_{2l+2})(1-a_{2l})}. \tag{3}$$

Here, $a_1 = 1$, $a_2 = 1 - p^2$, and

$$a_k = (1-p)a_{k-1} + p(1-p)a_{k-2}.$$

Asymptotics of the second moment

Proposition

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

$$\mathbb{E}\left(\mathbb{Y}^2\right) = \frac{4}{p^4} + o(p^{-4}).$$

Asymptotics of the second moment

Proposition

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

$$\mathbb{E}\left(\mathbb{Y}^2\right) = \frac{4}{p^4} + o(p^{-4}).$$

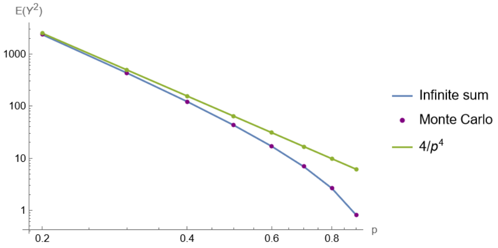


Figure: Exact values and asymptotic estimate for $\mathbb{E}\left(\mathbb{Y}^2\right)$.

Proving concentration

- Recall:

Proposition

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

$$\mathbb{E}\left(\mathbb{Y}^2\right) = \frac{4}{p^4} + o(p^{-4}).$$

Proposition

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

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

Proving concentration

- Therefore, the standard deviation σ,

$$\sigma = \sqrt{\mathbb{E}\left(\mathbb{Y}^2\right) - \mathbb{E}\left(\mathbb{Y}\right)^2} = o(p^{-2}) \qquad (4)$$

grows asymptotically slower than $\mathbb{E}\left(Y\right) \sim 2/p^2$.

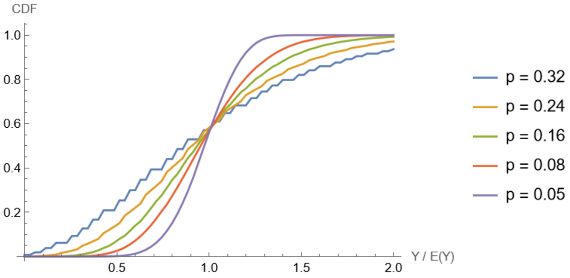


Figure: The cumulative distribution function of Y, normalized by $\mathbb{E}\left(Y\right)$, for $N = 800$ and $p = 0.05, 0.08, 0.16, 0.24, 0.32$. (Monte Carlo simulation.)

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

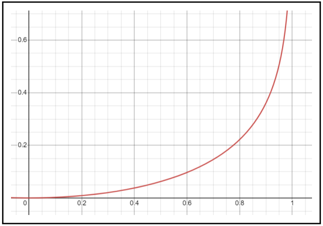
- Evaluating the “almost-geometric” sum yields

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

where

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

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



Exponential Decay

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."

Back to 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

- $\mathbb{E}\left(Y^k\right) = \sum_{i_1,\ldots,i_k=0}^{2N} \mathbb{E}\left(X_{i_1}\ldots X_{i_k}\right)$ 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