Distribution of Missing Sums in Correlated Sumsets

Dylan King
Wake Forest University
kingdal6@wfu.edu

Joint work with Chenyang Sun, Thomas Martinez, Advised by Steven Miller

Shenandoah Undergraduate Mathematics and Statistics Conference, September 21, 2019
Introduction

Given $A \subseteq \{0, \ldots, n - 1\}$, with $|A|$ its size, define its sumset

$A + A = \{a_1 + a_2 \mid a_1, a_2 \in A\} \subseteq \{0, \ldots, 2n - 2\}$. 
Introduction

Given $A \subseteq \{0, \ldots, n - 1\}$, with $|A|$ its size, define its sumset

- $A + A = \{a_1 + a_2 | a_1, a_2 \in A\} \subseteq \{0, \ldots, 2n - 2\}$.

- Sumsets are fundamental objects in number theory
- Fermat’s Last Theorem, $(N_n + N_n) \cap N_n = \emptyset$ for $n \geq 3$ if $N_n$ is the set of $n^{th}$ powers of $\mathbb{N}$
- Goldbach Conjecture: for the set of primes $P$, $P + P \supseteq 2\mathbb{N} \setminus \{0, 2\}$
Recent research in $|A + A|$ as a random variable

Set $\mathbb{P}(i \in A) = p$, where $p \in [0, 1]$ and $q := 1 - p$.

Martin and O’Bryant’s formative paper [MO] compared $|A + A|$ to $|A - A|$.
Motivating Questions

- What is $\mathbb{E}[|A + A|]$?
- What is $\text{Var}(|A + A|)$?
Prior Work

Theorem (Martin and O’Bryant ’06)

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

- Can we compute the same for generic $p$?
- Problem: not all sets $A$ are equally likely...
Results

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

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

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

\[
2n - 1 - 2 \frac{1}{1 - \sqrt{2q}} - (2q)^{\frac{n-1}{2}} \leq \mathbb{E}[|A + A|] \leq 2n - 1 - 2 \frac{1 - q^{\frac{n-1}{2}}}{1 - \sqrt{q}}
\]
How to compute expected value?

- Natural case previously studied by [MO] in 2006: set \( \rho = \frac{1}{2} \).
- Every subset of \( \{0, \ldots, n-1\} \) has equal probability of occurring.

\[
\mathbb{E}[|A + A|] = \frac{\sum_{A \subset \{0, \ldots, n-1\}} |A + A|}{2^n} = \sum_{i=0}^{2n-2} P(i \in A + A) = \sum_{i=0}^{2n-2} \left(1 - P(i \notin A + A)\right)
\]
For other $p$ this fails

For $p \neq \frac{1}{2}$, not every subset of $\{0, \ldots, n-1\}$ has equal probability of occurring.

$$
\mathbb{E}[|A + A|] = \frac{\sum_{A \subseteq \{0, \ldots, n-1\}} |A + A|}{2^n} \\
= \sum_{i=0}^{2n-2} \mathbb{P}(i \in A + A) \\
= \sum_{i=0}^{2n-2} \left(1 - \mathbb{P}(i \notin A + A)\right)
$$
For any $p$, all subsets of $\{0, ..., n - 1\}$ with equal cardinality have the same probability of occurring.

$$
\mathbb{E}[|A + A|] = \sum_{r=0}^{n} \mathbb{P}(|A| = r) \sum_{i=0}^{2n-2} \mathbb{P}(i \in A + A \mid |A| = r)
$$

$$
= \sum_{r=0}^{n} \binom{n}{r} p^r q^{n-r} \sum_{i=0}^{2n-2} \left(1 - \mathbb{P}(i \not\in A + A \mid |A| = r)\right)
$$
A Graph Theoretic Solution

- \( G = (V, E), \ V = \{0, \ldots, n - 1\} \)
- Edge \((k_1, k_2)\) if \(k_1 + k_2 = i\)
- \(A\) corresponds to a subset of these vertices
- A vertex cover of missing elements corresponds to \(i \not\in A + A\)
- This graph is a collection of disjoint edges and isolated vertices
Vertex Cover Definition

A vertex cover $V'$ of an undirected graph $G = (V, E)$ is a subset of $V$ such that $uv \in E \Rightarrow u \in V' \lor v \in V'$. 
A Graph Theoretic Solution

- Since $|A| = r$, we are looking for the number of $r$-vertex vertex covers

\[
\ell \text{ disjoint edges} \begin{cases} 

n - 2\ell \text{ isolated vertices} 
\end{cases}
\]
Computing $\mathbb{E}[|A+A|]$ 

Lemma (King, Martinez, Miller, Sun ’19)

\[
\mathbb{P}[k \not\in S+S \mid |S|=r] = \begin{cases} 
\sum_{i=\frac{k+1}{2}}^{k+1} 2^{k+1-i} \left(\frac{k+1}{2}\right) \left(\frac{n-k-1}{r-i}\right) \\
\sum_{i=\frac{k}{2}}^{k} 2^{k-i} \left(\frac{k}{2}\right) \left(\frac{n-k-1}{r-1-i}\right)
\end{cases} 
\]

for $k$ odd

for $k$ even
Implementing Combinatorics

\[ \mathbb{P}[k \not\in S + S \mid |S| = r] = \frac{\text{ways to place } r \text{ vertices and get cover}}{\text{ways to choose } r \text{ vertices from } n} \]

\[ = \frac{\sum_{i=\frac{k+1}{2}}^{k+1} 2^{k+1-i} \binom{k+1}{i} \binom{n-k-i}{r-i}}{\binom{n}{r}} \]

\[ r = r - i + i = r - i + \frac{k+1}{2} + i - \frac{k+1}{2} \]

\[ \frac{k+1}{2} \] disjoint edges

\[ n - k - 1 \] isolated vertices
Demo Slide

\[
\frac{k+1}{2} \text{ disjoint edges} \quad \begin{cases} 
\cdot & \cdot \\
\cdot & \cdot \\
\cdot & \cdot 
\end{cases} \quad n - k - 1 \text{ isolated vertices}
\]
Implementing Combinatorics

\[ \mathbb{P}[k \not\in S + S \mid |S| = r] = \frac{\text{ways to place } r \text{ vertices and get cover}}{\text{ways to choose } r \text{ vertices from } n} \]

\[ = \sum_{i=\frac{k+1}{2}}^{k+1} 2^{k+1-i} \binom{k+1}{i-k+1} \binom{n-k-1}{r-i} \binom{n}{r} \]

\[ r = r - i + i \]
\[ = r - i + \frac{k+1}{2} + i - \frac{k+1}{2} \]

\[ \frac{k+1}{2} \text{ disjoint edges} \]

\[ n - k - 1 \text{ isolated vertices} \]
Implementing Combinatorics

\[ \mathbb{P}[k \notin S + S \mid |S| = r] = \frac{\text{ways to place r vertices and get cover}}{\text{ways to choose r vertices from n}} \]

\[ = \sum_{i=\frac{k+1}{2}}^{k+1} 2^{k+1-i} \binom{k+1}{i} \binom{n-k-1}{r-i} \binom{n}{r} \]

\[ r = r - i + i \]

\[ = r - i + \frac{k+1}{2} + i - \frac{k+1}{2} \]

\[ \frac{k+1}{2} \text{ disjoint edges} \]

\[ n - k - 1 \text{ isolated vertices} \]
Implementing Combinatorics

$$\mathbb{P}[k \not\in S + S \mid |S| = r] = \frac{\text{ways to place } r \text{ vertices and get cover}}{\text{ways to choose } r \text{ vertices from } n}$$

$$= \frac{\sum_{i=\frac{k+1}{2}}^{k+1} 2^{k+1-i} \left(\frac{k+1}{2}\right) \binom{n-k-1}{r-i}}{\binom{n}{r}}$$

$$r = r - i + i$$

$$= r - i + \frac{k+1}{2} + i - \frac{k+1}{2}$$

$$\frac{k+1}{2} \text{ disjoint edges}$$

$$n - k - 1 \text{ isolated vertices}$$
Computing the Variance of $|A + A|$ 

$$\text{Var}(|A + A|) = E[|A + A|^2] - E[|A + A|^2].$$

- The major component in computing $E[|A + A|^2]$ is $\mathbb{P}(i, j \not\in A + A)$
- Analyzed for the $p = \frac{1}{2}$ case in [LMO]
- We work on the problem for generic $p$
A Problem with Dependencies

- \( P[i \not\in A + A] \) is well known.

**Lemma (Martin and O’Bryant ’06)**

Let \( q = 1 - p \). If \( i \leq n - 1 \),

\[
P[i \not\in A + A] = \begin{cases} 
(2q - q^2)^{(i+1)/2} & \text{for } i \text{ odd} \\
q(2q - q^2)^{i/2} & \text{for } i \text{ even}
\end{cases}
\]

- However, \( P[i, j \not\in A + A] \) is laden with dependencies.
- Example: \( P[0 \not\in A + A] = 1 - p \),
  \( P[1 \not\in A + A] = 1 - p^2 \), but \( P[0, 1 \not\in A + A] = 1 - p^2 \)
A Graph Theoretic Solution

- $G = (V, E)$, $V = \{0, \ldots n - 1\}$
- Edge $(k_1, k_2)$ if $k_1 + k_2 = i$ or $k_1 + k_2 = j$
- $A$ corresponds to a subset of these vertices
- A vertex cover of missing elements corresponds to $i, j \not\in A + A$
Structure of the Graph

- This graph is the union of disjoint paths [LMO]
- Understanding vertex covers reduces to understanding vertex covers of paths

paths of known length

isolated vertices

\[ i \quad j \quad i \]
Prior Work - Fibonacci Numbers

- When $p = \frac{1}{2}$, all sets equally likely, [LMO] only needed to count vertex covers
- How can we count vertex covers on a path of length $n$?

\[ V_1 \quad V_2 \quad V_3 \quad V_4 \quad V_n \]

Case 1: $1 \in S$  
12 edge is covered

Case 2: $1 \notin S$, then necessarily $2 \in S$, 23 edge is covered
Prior Work - Fibonacci Numbers

- When \( p = \frac{1}{2} \), all sets equally likely, [LMO] only needed to count vertex covers
- How can we count vertex covers on a path of length \( n \)?

\[
V_1 \quad V_2 \quad V_3 \quad V_4 \quad V_n
\]

Case 1: \( 1 \in S \)  
12 edge is covered

Case 2: \( 1 \notin S \), then necessarily \( 2 \in S \), 23 edge is covered

\[
F_n := \# \text{ of vertex covers}
\]

\[
F_n = F_{n-1} + F_{n-2}; \text{ Fibonacci!}
\]
Vertex Cover Probabilities

- How can we compute the probability of finding a vertex cover on a path of length $n$?

$$V_1 \quad V_2 \quad V_3 \quad V_4 \quad V_n$$

- Case 1: $1 \in S$, then necessarily $12$ edge is covered
- Case 2: $1 \notin S$, then necessarily $2 \in S$, $23$ edge is covered

- $a_n := \mathbb{P}(\text{a vertex cover})$
- $a_n = qa_{n-1} + pqa_{n-2}$; a recurrence relation we can solve
Lemma

Set $\phi(p) := \sqrt{1 + 2p - 3p^2}$. Then

$$a_n = \frac{(\phi(p) - 1 - p)(1 - p - \phi(p))^n + (\phi(p) + 1 + p)(1 - p + \phi(p))^n}{2^{n+1}\phi(p)}$$

This can be used to compute the variance of $|A + A|$. 
A Generalization to Correlated Sumsets

- Introduced by 2013 SMALL REU group [DKMMW]

- Replace $A + A$ with $A + B$, where
  - $\mathbb{P}(i \in A) = p$
  - $\mathbb{P}(i \in B \mid i \in A) = p_1$
  - $\mathbb{P}(i \in B \mid i \not\in A) = p_2$

- $p_1 = 1, p_2 = 0$ reduces to $A + A$

- Once again, determining $\mathbb{P}(i, j \not\in A + B)$ is difficult
Generalizing the Graph Framework

- $G = (V, E)$, $V = \{0_A, \ldots (n - 1)_A, 0_B, \ldots (n - 1)_B\}$
- Edge $(k_1, k_2)$ if $k_1 + k_2 = i$ or $k_1 + k_2 = j$
- $A$ corresponds to a subset of these vertices
- A vertex cover of missing elements corresponds to $i, j \not\in A + B$
Generalizing the Graph Framework

- \( G = (V, E), \quad V = \{0_A, \ldots, (n-1)_A, 0_B, \ldots, (n-1)_B\} \)
- Edge \((k_1, k_2)\) if \(k_1 + k_2 = i\) or \(k_1 + k_2 = j\)
- \(A\) corresponds to a subset of these vertices
- A vertex cover of missing elements corresponds to \(i, j \notin A + B\)
Correlated Set Recurrence Relation

- How can we compute the probability of finding a vertex cover on a pair of paths of length $n$?

- Many cases, based on whether we have $1 \in A$ and/or $1 \in B$

- Solution: system of recurrence relations
Correlated Set Recurrence Relation

\begin{center}
\begin{tikzpicture}[scale=0.8]
  \node (v1) at (0,0) [shape=circle,fill,inner sep=2pt] {$v_1$};
  \node (vn) at (4,0) [shape=circle,fill,inner sep=2pt] {$v_n$};
  \node (vn1) at (2,1) [shape=circle,fill,inner sep=2pt] {$v_{n-1}$};
  \node (vn2) at (2,-1) [shape=circle,fill,inner sep=2pt] {$v_{n-2}$};
  \node (vna) at (4,1) [shape=circle,fill,inner sep=2pt] {$v_{n,A}$};
  \node (vnb) at (4,-1) [shape=circle,fill,inner sep=2pt] {$v_{n,B}$};

  \draw (v1) -- (vn2);
  \draw (v1) -- (vn1);
  \draw (vn1) -- (vn);
  \draw (vn2) -- (vna);
  \draw (vn2) -- (vnb);

  \node at (0,-2) {$A$};
  \node at (4,-2) {$B$};

correlation of sets $A$ and $B$
\end{tikzpicture}
\end{center}

$a_n := \mathbb{P}(\text{a vertex cover})$

$b_n := \mathbb{P}(\text{a vertex cover AND } n_A \in A)$

$c_n := \mathbb{P}(\text{a vertex cover AND } n_B \in B)$

then we find that

\begin{align*}
a_n &= qq_2 a_{n-1} + qp_2 b_{n-1} + pq_1 c_{n-1} + pp_1 qq_2 a_{n-2} \\
b_n &= qq_2 a_{n-1} + qp_2 b_{n-1} \\
c_n &= qq_2 a_{n-1} + pq_1 c_{n-1}
\end{align*}
Future Work

- Handle the bounds on $\mathbb{E}[|A + A|]$ for $p \leq \frac{1}{2}$.
- Find and analyze a closed form for $a_n$ in the correlated sets case.
- Find $\mathbb{E}[|A + B|]$ and $\text{Var}(|A + B|)$ for any correlated sumset $A + B$. 
