Distribution of Missing Sums in Correlated Sumsets

Dylan King
Wake Forest University
kingdal6@wfu.edu

Joint work with Hung Chu, Noah Luntzlara, Thomas Martinez, Steven J. Miller, Lily Shao, Chenyang Sun, and Victor Xu

Combinatorial and Additive Number Theory, June 5 2020
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

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

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

- Remember \( A \subset \{0, \ldots, n - 1\} \) gives \( A + A \subset \{0, \ldots, 2n - 2\} \).
Numerical Experimentation

[LMO] plotted the frequency with which $k$ sums are missing from $A + A$. 
We are adding elements of $0, \ldots, n - 1$ to get elements of $0, \ldots, 2n - 2$; we need to estimate how many we will find.
Problem Intuition

How likely is it that $n \in A + A$?

\[
\begin{align*}
0,1,... & \quad n-1 \\
+ & \\
0,1,... & \quad n-1 \\
\downarrow & \\
0,1,... & \quad 2n-2
\end{align*}
\]
There are $O(n)$ distinct pairs which, if present, sum to $n$. 

$$0, 1, \ldots, n-1$$

$3$ +

$$0, 1, \ldots, n-1$$

$\downarrow$

$$0, 1, \ldots, n$$

$2n-2$
Problem Intuition

For smaller numbers, there are fewer pairs. For example, 7 has only $0 + 7, 1 + 6, 2 + 5, 3 + 4$, not matter the value of $n > 7$. 

\[ \begin{array}{c}
\text{0,1,...} \\
3 \\
\text{n-1} \\
\hline
\text{0,1,...} \\
4 \\
\text{n-1} \\
\hline
\text{0,1,...} \\
7 \\
\text{2n-2}
\end{array} \]
The takeaway: numbers in the middle of $0, \ldots, 2n - 2$ are likely to appear, while numbers at the upper and lower extremes are more likely to be missing.
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...
Numerics

We can plot still try numerics and investigate the frequency with which $k$ sums are missing from $A + A$, for different $p \in [0, 1]$. 

![Graph showing frequency of $k$ sums missing from $A + A$ for different $p$ values]
Theorem (CHKLMSX)

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

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

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

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

For any $p$, all subsets of $\{0, ..., n - 1\}$ with equal cardinality have the same probability of occurring.

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

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

Our target: $\mathbb{P}(k \not\in A + A \mid |A| = r)$
A Graph Theoretic Solution

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

![Graph diagram]
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 $n - r$-vertex vertex covers

$\ell$ disjoint edges

$n - 2\ell$ isolated vertices
Computing $E[|A + A|]$

**Lemma (CHKLMSX)**

\[
P[k \not\in A + A \mid |A| = r] = \begin{cases} 
\sum_{i=\frac{k+1}{2}}^{k+1} 2^{k+1-i} \binom{k+1}{i-\frac{k+1}{2}} \binom{n-k-1}{r-i} & \text{for } k \text{ odd} \\
\sum_{i=\frac{k}{2}}^{k} 2^{k-i} \binom{k}{i-\frac{k}{2}} \binom{n-k-1}{r-1-i} & \text{for } k \text{ even}
\end{cases}
\]
Implementing Combinatorics

\[ P[k \not\in A + A \mid |A| = n - r] = \frac{\text{ways to place } r \text{ vertices and get cover}}{\text{ways to choose } r \text{ vertices from } n} \]
\[ = \sum_{i=\frac{k+1}{2}}^{k+1} 2^{k+1-i} \binom{\frac{k+1}{2}}{i} \binom{n-k-1}{r-i} \binom{n}{r} \]

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

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

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

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

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

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

\[ = \frac{k+1}{2} \text{ disjoint edges} \]
\[ n - k - 1 \text{ isolated vertices} \]
Implementing Combinatorics

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

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

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

\[ \left\{ \begin{array}{l}
\frac{k+1}{2} \text{ disjoint edges} \\
 n - k - 1 \text{ isolated vertices}
\end{array} \right. \]
Implementing Combinatorics

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

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

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

\( \frac{k+1}{2} \) disjoint edges

\( n - k - 1 \) isolated vertices
Variance: A Problem with Dependencies

- $\mathbb{P}[i \not\in A + A]$ is straightforward.

**Lemma (Martin and O’Bryant ’06)**

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

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

- However, $\mathbb{P}[i, j \not\in A + A]$, required to compute Variance, is laden with dependencies.
- Example: $\mathbb{P}[0 \not\in A + A] = 1 - p$, $\mathbb{P}[1 \not\in A + A] = 1 - p^2$, but $\mathbb{P}[0, 1 \not\in A + A] = 1 - p^2$.
- Analyzed for the $p = \frac{1}{2}$ case in [LMO].
- We work on the problem for generic $p$. 
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 $i, j \not\in A + A$

```
0 1 2       i
            \-
    k1   k2   n
```
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
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 $n$?

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

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 $n$?

![Graph of a path with vertices $v_1, v_2, v_3, v_4, v_n$]

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

- $F_n := \#$ of vertex covers
- $F_n = F_{n-1} + F_{n-2}$; Fibonacci!
Vertex Cover Probabilities

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

Case 1: $1 \in S$, 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 $i, j \notin A + B$
Future Work

- Find a lower bound 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 $A + B$. 


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 \(i, j \notin A + B\)
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

\[ a_n := \mathbb{P}( \text{a vertex cover} ) \]
\[ b_n := \mathbb{P}( \text{a vertex cover AND } n_A \in A) \]
\[ c_n := \mathbb{P}( \text{a vertex cover AND } n_B \in B) \]

then we find that

\[ a_n = qq_2 a_{n-1} + qp_2 b_{n-1} + pq_1 c_{n-1} + pp_1 qq_2 a_{n-2} \]
\[ b_n = qq_2 a_{n-1} + qp_2 b_{n-1} \]
\[ c_n = qq_2 a_{n-1} + pq_1 c_{n-1} \]