Distribution of Missing Sums in Correlated Sumsets

Dylan King
Wake Forest University
kingda16@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

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

- 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, $P + P \supseteq 2\mathbb{N} \setminus \{0, 2\}$

Setting

- 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 $\mathrm{Var}(|A + A|)$?

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

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

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

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

How to compute expected value?

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

$$\mathbb{E}[|A+A|] = \frac{\sum_{A \subset \{0,...,n-1\}} |A+A|}{2^n}$$

$$= \sum_{i=0}^{2n-2} \mathbb{P}(i \in A+A)$$

$$= \sum_{i=0}^{2n-2} \left(1 - \mathbb{P}(i \notin A+A)\right)$$

For other p this fails

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

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

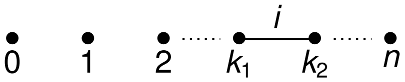
A Solution

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

$$\begin{aligned}\mathbb{E}[|A+A|] &= \sum_{r=0}^{n} \mathbb{P}(|A| = r) \sum_{i=0}^{2n-2} \mathbb{P}(i \in A + A \mid |A| = r) \\ &= \sum_{r=0}^{n} \binom{n}{r} p^r q^{n-r} \sum_{i=0}^{2n-2} \left(1 - \mathbb{P}(i \notin A + A \mid |A| = r)\right)\end{aligned}$$

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 $i \notin A + A$
- This graph is a collection of disjoint edges and isolated vertices



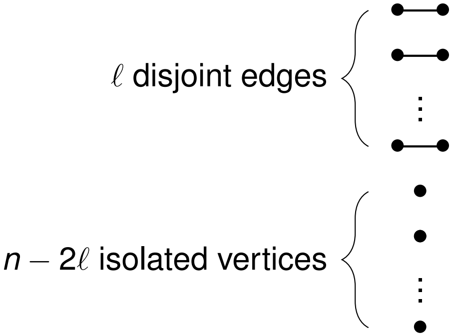
Vertex Cover Definition

Vertex Cover Definition

a vertex cover V' of an undirected graph $G = (V, E)$ is a subset of V such that $uv \in E \Rightarrow u \in V' \vee v \in V'$

A Graph Theoretic Solution

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



Computing $\mathbb{E}[|A+A|]$

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

$$\mathbb{P}[k \notin S+S \mid |S|=r] = \begin{cases} \dfrac{\sum_{i=\frac{k+1}{2}}^{k+1} 2^{k+1-i}\binom{\frac{k+1}{2}}{i-\frac{k+1}{2}}\binom{n-k-1}{r-i}}{\binom{n}{r}} & \text{for } k \text{ odd} \\ \dfrac{\sum_{i=\frac{k}{2}}^{k} 2^{k-i}\binom{\frac{k}{2}}{i-\frac{k}{2}}\binom{n-k-1}{r-1-i}}{\binom{n}{r}} & \text{for } k \text{ even} \end{cases}$$

Implementing Combinatorics

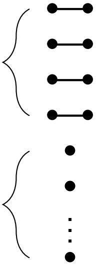
$$\mathbb{P}[k \notin S+S \mid |S|=r] = \frac{\text{ways to place r vertices and get cover}}{\text{ways to choose r vertices from n}}$$

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

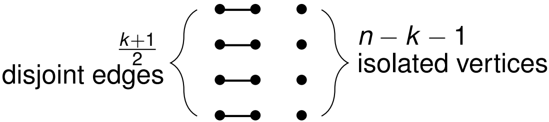
$$\begin{aligned} r &= r-i+i \\ &= r-i+\frac{k+1}{2}+i-\frac{k+1}{2} \end{aligned}$$

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

$n-k-1$ isolated vertices



Demo Slide



Implementing Combinatorics

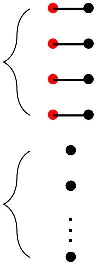
$$\mathbb{P}[k \notin S + S \mid |S| = r] = \frac{\text{ways to place r vertices and get cover}}{\text{ways to choose r vertices from n}}$$

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

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

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

$n-k-1$ isolated vertices



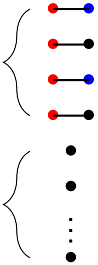
Implementing Combinatorics

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

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

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

$n - k - 1$ isolated vertices



$$\begin{aligned} r &= r - i + i \\ &= r - i + \frac{k+1}{2} + i - \frac{k+1}{2} \end{aligned}$$

Implementing Combinatorics

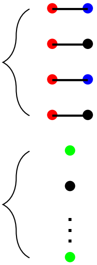
$$\mathbb{P}[k \notin S + S \mid |S| = r] = \frac{\text{ways to place r vertices and get cover}}{\text{ways to choose r vertices from n}}$$

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

$$\begin{aligned} r &= r - i + i \\ &= r - i + \frac{k+1}{2} + i - \frac{k+1}{2} \end{aligned}$$

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

$n - k - 1$ isolated vertices



Computing the Variance of $|A + A|$

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

- The major component in computing $E[|A+A|^2]$ is $\mathbb{P}(i, j \notin A + A)$
- Analyzed for the $p = \frac{1}{2}$ case in [LMO]
- We work on the problem for generic p

A Problem with Dependencies

- $\mathbb{P}[i \notin A + A]$ is well known.

Lemma (Martin and O'Bryant '06)

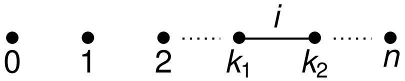
Let $q = 1 - p$. If $i \leq n - 1$,

$$\mathbb{P}[i \notin A + A] = \begin{cases} (2q - q^2)^{(i+1)/2} & \text{for } i \text{ odd} \\ q(2q - q^2)^{i/2} & \text{for } i \text{ even} \end{cases}$$

- However, $\mathbb{P}[i, j \notin A + A]$ is laden with dependencies
- Example: $\mathbb{P}[0 \notin A + A] = 1 - p$, $\mathbb{P}[1 \notin A + A] = 1 - p^2$, but $\mathbb{P}[0, 1 \notin A + A] = 1 - p^2$

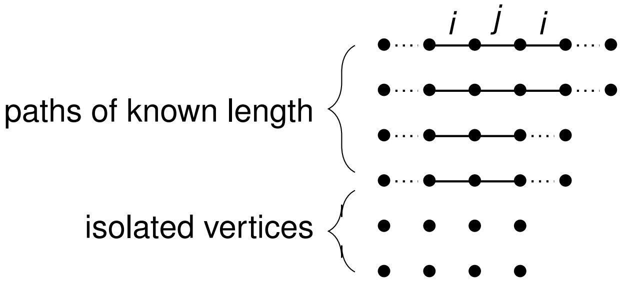
A Graph Theoretic Solution

- $G = (V, E)$, $V = \{0, \dots n-1\}$
- 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 $i, j \notin A + A$



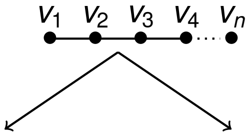
Structure of the Graph

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



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?

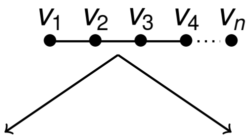


Case 1 : $1 \in S$
12 edge is covered

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?



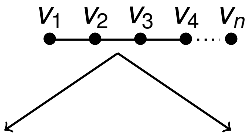
Case 1 : $1 \in S$
12 edge is covered

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

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

Vertex Cover Probabilities

Lemma

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

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

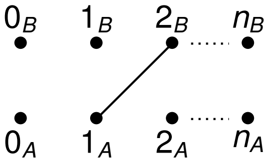
This can be used to compute the variance of $|A + A|$.

A Generalization to Correlated Sumsets

- Introduced by 2013 SMALL REU group [DKMMW]
- Replace $A + A$ with $A + B$, where
    - $\mathbb{P}(i \in A) = p$
    - $\mathbb{P}(i \in B \mid i \in A) = p_1$
    - $\mathbb{P}(i \in B \mid i \notin A) = p_2$
- $p_1 = 1, p_2 = 0$ reduces to $A + A$
- Once again, determining $\mathbb{P}(i, j \notin A + B)$ is difficult

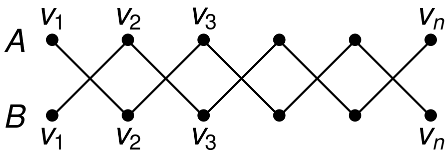
Generalizing the Graph Framework

- $G = (V, E)$, $V = \{0_A, \ldots (n-1)_A, 0_B, \ldots (n-1)_B\}$
- 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 $i, j \notin A + B$



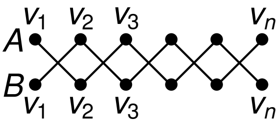
Generalizing the Graph Framework

- $G = (V, E)$, $V = \{0_A, \ldots (n-1)_A, 0_B, \ldots (n-1)_B\}$
- 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 $i, j \notin A + B$



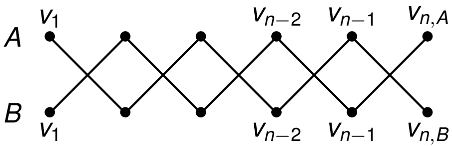
Correlated Set Recurrence Relation

- 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



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

then we find that

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

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 $\mathrm{Var}(|A+B|)$ for any correlated sumset $A+B$.

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.
- T. Do, A. Kulkarni, S.J. Miller, D. Moon, and J. Wellens, *Sums and Differences of Correlated Random Sets*, Journal of Number Theory **147** (2015), 44–68.