Distribution of Missing Sums in Correlated Sumsets

Thomas Martinez (Harvey Mudd College, tmartinez@hmc.edu), Dylan King (Wake Forest University, kingda16@wfu.edu)

Joint research with Chenyang Sun. Advised by Steven J. Miller.

SMALL 2019 REU, Williams College

Introduction

Considering a set \( A \subseteq \{0,1,\ldots,n-1\} \), we define a **sumset** as

\[
A + A = \{i + j \mid i,j \in A\}.
\]

We look at random sets \( A \) where \( \Pr(i \in A) = p \).

**Notation:** \( q = 1 - p \),

\[
M_{n_{r-1}} := \left\lfloor \frac{2^n - 2}{(A + A)} \right\rfloor = 2^n - 1 - |A + A|,
\]

**Uniform Distribution:** Most previous work focused on \( p = 1/2 \). This implies every subset \( A \) has an equal probability of occurring.

Previous Results

Let \( p = \frac{1}{2} \). Martinez and O’Bryant found

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

Let \( n > 5k \). Lazarev, Miller and O’Bryant found

\[
2^{-k/2} \leq m_n(k) \leq (\phi/2)^k,
\]

where the implied constants are independent of \( k \) and \( n \), and \( \phi \) is the golden ratio.

Expected Value Generalization

We generalize previous results for any \( p \). The explicit formula for \( \mathbb{E}[|A + A|] \) is stated below.

**Expected Value Formula**

For \( p \in [0,1) \), \( \mathbb{E}[|A + A|] \) is

\[
\sum_{i=0}^{n-1} p^{i+1}q^{n-i-1} \left(\frac{1}{i+1} \left( 1 - \frac{f(k)}{i+1} \right) - \frac{f(n-1)}{i+1} \right),
\]

where

\[
f(k) = \begin{cases} 
\frac{\sqrt{n-k+1}}{\sqrt{k+1}}, & \text{for } k \text{ odd} \\
\frac{\sqrt{n-k+1}}{\sqrt{k}}, & \text{for } k \text{ even}
\end{cases}
\]

We also find bounds for the Expected Value.

Expected Value Bounds

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

\[
\mathbb{E}[|A + A|] < 2n - 1 - \sqrt{n}.
\]

For \( p \in (0,5,1) \),

\[
2n - 1 - \frac{\sqrt{n}}{1 - \sqrt{n}} \leq \mathbb{E}[|A + A|].
\]

Figure 1: Above is a plot for \( n = 35 \).

Approach

Previously, every subset had an equal chance of occurring, which allowed for a simpler calculation. Now, we must take into account the probability that \( |A| = r \), where \( r \in [0,n] \). Using similar techniques, we found

\[
\mathbb{E}[|A + A|] = \sum_{i=0}^{n-1} p^{i+1}q^{n-i-1}(1 - \Pr(k \notin A + A \mid |A| = r)).
\]

We compute \( \Pr(k \notin A + A \mid |A| = r) \) using graph theory. We define \( G = (V,E) \) with vertices representing elements in \([0,n-1]\) and an edge is present if \( v_1 + v_2 = k \). We are interested in finding a **vertex cover** with \( n - r \) vertices.

\[
f(k) = \begin{cases} 
\frac{\sqrt{n-k+1}}{\sqrt{k+1}}, & \text{for } k \text{ odd} \\
\frac{\sqrt{n-k+1}}{\sqrt{k}}, & \text{for } k \text{ even}
\end{cases}
\]

From this, we can derive \( f(k) \) from the Expected Value formula.

\[
m_n(k) \text{ Generalization}
\]

We generalize previous results of \( m_n(k) \) for any \( p \), and show a plot for \( k = 10 \).

\[
\mathbb{E}[n_{n_r}] < \left( \frac{1 - p + \phi(p)}{2} \right)^k.
\]

The lower bound is achieved by finding the probability that the first \( k/2 \) elements are not in \( A \), and showing that the probability that the rest of the elements in \( A \) is a subset \( A' \) such that \( A' + A' \) has no missing elements (which is much more likely that the first condition). We get

\[
m_n(k) \geq (A = k/2 + A' \text{ and } M_{n-k/2}(A') = 0) = q^{k/2} \Pr(M_{n-k/2}(A') = 0) \gg q^{k/2}.
\]

The upper bound is achieved from noting that missing an element at least \( k/2 \) elements away from the ends of \([0,2n-2]\) is very unlikely. For an upper bound, notice that missing \( k \) elements implies that missing an element at least \( k/2 \) elements away from the ends of \([0,2n-2]\). This event is unlikely, because there are so many pairs of numbers that add up to an element in the middle of \( A + A \), so we look at the probability of this event as our upper bound.

After some manipulation, we find

\[
m_n(k) \leq \Pr(A + A \text{ misses } 2 \text{ elements greater than } k - 3) = \sum_{k-3 < j < k} \Pr(i, j \notin A + A) \leq \left( \frac{1 - p + \phi(p)}{2} \right)^k.
\]

Where the last term comes from a generalization of graph theory introduced by Lazarev, Miller and O’Bryant. We seek to find \( \Pr(i, j \notin A + A) \) using a similar definition of a graph as earlier, this is the same as finding a vertex cover of missing elements on a path, as below.

\[
\begin{align*}
&v_1 \quad v_2 \quad v_3 \quad \cdots \quad v_k \\
&v_1 \quad v_2 \quad v_3 \quad \cdots \quad v_k \\
&v_1 \quad v_2 \quad v_3 \quad \cdots \quad v_k
\end{align*}
\]

Letting \( n \) denote the number of paths and \( \alpha \) denoting the probability of a vertex cover, we derive \( \alpha = q^{a_{n-1}} + pq^{a_{n-2}} \) and get a closed form

\[
\alpha_n = \frac{(\phi(p) - 1) - \phi(p)(1 - p - \phi(p))^n}{2^n + n + 1 \phi(p)}.
\]

References

- G. Martin, K. O’Bryant, Many sets have more sums than differences, in Additive Combinatorics. CRM Proceedings and Lecture Notes, vol. 43 (American Mathematical Society, Providence, 2007), pp. 287-305
- https://arxiv.org/abs/1109.4700

Acknowledgements

We would like to thank our mentor Dr. Steven J. Miller, the rest of the Number Theory & Probability Group and the SMALL REU Program. This project was funded by NSF Grants DMS1561945 and DMS1659037, Williams College and Wake Forest University.