Slow Decay and Missing Term Distributions in Generalized Sum and Difference Sets
Eric Winsor (rcwnsr@umich.edu), Advisor: Steven J. Miller
Number Theory and Probability Group - SMALL 2017 - Williams College

1. Background

Definition 1.1. Fix \( A \subset \mathbb{Z} \) and integers \( m, n \geq 0 \). The generalized sum and difference set with \( m \) positive summands and \( n \) negative summands is

\[
\mathcal{A} = \{ \sum_{j \in I} a_j - \sum_{k \in J} b_k : \sum_{j \in I} a_j \leq m, \sum_{k \in J} b_k \leq n \}.
\]

MSTD Sets: Classically, we are interested in the size of \( A + A \) and \( A - A \). Because addition is commutative and subtraction is not, we expect \( A + A \) to be smaller than \( A - A \) (given distinct \( a \in A \) and \( b \in A \), \( a + b = b + a \) is a single term of \( A + A \), but \( a - b \neq b - a \) are two distinct terms of \( A - A \)). This intuition turns out to be correct in some sense, and we expect that for most sets \( A \), we have \(|A + A| < |A - A|\). Iyer, Lazarev, Miller, and Zhang proved that given nonnegative integers \( m, n \geq 0 \), the probability that \( A + A \) and \( A - A \) are expected to be

\[
\mathbb{P}(|A + A| < |A - A|) = \frac{|A|}{n + |A|} - \frac{|A|}{n + |A|} + \frac{|A|}{n + |A|}.
\]

In general, we are in-

2. Generalized Sum and Difference Sets with Decay

We now let \( A \) be a randomly chosen subset of \([0, \ldots, N]\) where each element of this set has probability \( p \) of being in \( A \). We are specifically interested in the case where \( p(N) = N^{-\delta} \) for some \( \delta \in (0, 1) \). Given this distribution on \( \mathcal{A} \), we wish to investigate the relative sizes of all \( m - A \) and \( -n + A \) for fixed \( m > 0 \) and \( n \geq 0 \). We have yet to find a way of dealing with these dependencies. Thus, we turn to numerics to illuminate the missing term distribution for \( m - A \) and \( -n + A \). This suggests that we may be able to write this missing term distribution as a combination of 4 distributions constructed from the set \( \{0, \ldots, N\} \). The case of slow decay with \( \delta \geq 1/2 \) is strongly concentrated about its mean. While some details must be modified for the case of \( \delta < 1/2 \), the argument is fairly similar. In the general case, the representations of an element of \( m - A \) are not all independent, and we have yet to find a way of dealing with these dependencies. Thus, we turn to numerics to illuminate the missing term distribution for \( m - A \) and \( -n + A \). This intuition turns out to be correct in some sense, and we expect that for most sets \( A \), we have \(|A + A| < |A - A|\). Iyer, Lazarev, Miller, and Zhang proved that given nonnegative integers \( m, n \geq 0 \), the probability that \( A + A \) and \( A - A \) are expected to be

\[
\mathbb{P}(|A + A| < |A - A|) = \frac{|A|}{n + |A|} - \frac{|A|}{n + |A|} + \frac{|A|}{n + |A|}.
\]

In general, we are in-

3. Obstructions in Analysis of Slow Decay

We start by analyzing the distribution of missing terms from \( m - A \) when each element of \( A \) is chosen from \([0, \ldots, N]\) with fixed probability \( p \). We ran simulations of \( 2^{17} \) trials with \( N = 2^7 \) for all choices of \( m \geq 0 \) and \( n \leq 5 \) with \( 2 \leq m + n \leq 7 \). We include a graph of the missing term distribution for \( m = 4 \) as this distribution is fairly representative of the patterns we see in general.

4. Missing Term Distributions

The case of slow decay with \( \delta < 1/2 \) is not well understood. To analyze this case for \( \delta = 1/2 \), Hegarty and Miller analyzed the distribution of missing terms from \( m - A \) using the random variable \( X \) defined in (\ref{slow}.) This suggests that we may be able to write this missing term distribution as a combination of 4 distributions constructed from a single underlying distribution. Finding the underlying distribution may aid in computing the expectation of the number of missing terms.

5. Future Work

We hope to apply our work to analyzing the expectation of the number of missing terms from \( m - A \) and \( -n + A \). Currently, numerics seem to suggest that a result similar to Hegarty and Miller’s holds in this case. A plot of the expectation of missing terms from \( 1,4 \) with \( p(N) = N^{-\delta} \) as a function of \( \delta \) is shown in Figure 3. The expectation seems to grow as \( N^{-\frac{1}{1+\delta}} \) for some constant \( c \), which is similar to the case of \( \delta = 1/2 \). However, values of \( \delta \) are fairly small for this simulation, so we cannot be sure that the limiting behavior has set in.

6. Acknowledgements

We wish to thank Williams College, whose generous support made this research possible. This work is funded by NSF grant DMS1561945 and NSF grant DMS1659307.

References


Figure 3: The expectation of the number of missing sums from \( 1,4 \) (the random variable \( S = 4N + 1 - |A + 1| \) as a function of \( \delta \) given \( A \subset \{0, \ldots, N\} \), where each element is chosen with probability \( p(n) = N^{-\delta} \). Each \( N \) is simulated with \( 2^{17} \) trials.