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*

$$mA - nA := \left\{ \sum_{i=1}^{m} a_i - \sum_{j=1}^{n} b_j \mid a_i, b_j \in A \right\}$$

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 $a, b \in A$, $a+b = b+a$ is a single term of $A+A$, but $a-b$ and $b-a$ 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 $|A+A| < |A-A|$. If $|A+A| > |A-A|$, we call A a more sum than difference set, or MSTD set. For example,

$$\{0, 2, 3, 4, 7, 11, 12, 14\}$$

is an MSTD set, and a number of infinite families of MSTD sets are known. Martin and O'Bryant showed that if A is selected uniformly at random from the set of all subsets of $\{0, \ldots, N\}$, then the probability that A is an MSTD set stays above some positive lower bound for large N. Essentially, given a randomly chosen set A, A is an MSTD set a positive percentage of the time. Martin and O'Bryant also showed that $A+A$ and $A-A$ are expected to contain almost all possible sums and differences. Lazarev, Miller, and O'Bryant further analyzed the distribution of missing terms from $A+A$, and a graph of this distribution is shown below.

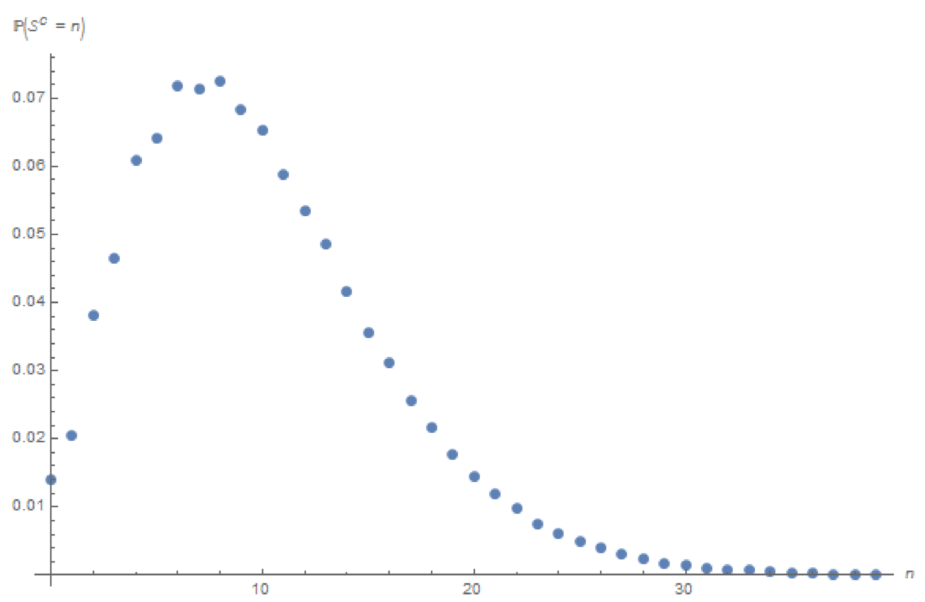

Figure 1: *The distribution of the number of missing sums for $2A = A+A$ (the random variable $\mathcal{S}^c = 2N+1-|2A|$) given A chosen uniformly at random from $P(\{0, \ldots, N\})$ with $N = 2^9$ and 2^{17} trials.*

Generalized Sum and Difference Sets: In general, we are interested in the relative sizes of all $mA - nA$ where $h = m+n$ is fixed. Iyer, Lazarev, Miller, and Zhang proved that given non-negative integers m, n, m', n' with $m+n = m'+n' > 1$ and $m \neq m'$, $|mA - nA| > |m'A - n'A|$ a positive percentage of the time. When investigating these relative sizes, we can restrict to the case where $m \geq n$ (otherwise, we can negate the sum to get a set with the same cardinality). Using a similar intuitive argument as in the simple case, we expect sets with n close to m to be larger than sets with n close to 0 given fixed $m+n$.

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 $p(N)$ 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 A we wish to investigate the relative sizes of all $mA - nA$ for fixed $h = m+n$ as N tends to infinity. We require $m \geq n$ as otherwise we could simply negate our set to get a set of the same cardinality with this property.

$h = 2$: Hegarty and Miller investigated this case, allowing $p(N)$ to be any function such that $p(N)$ decays to 0 as N approaches infinity and $N^{-1} = o(p(N))$. They showed that as N tends to infinity $A-A$ is almost surely larger than $A+A$. Furthermore, they investigated the ratios of the sizes of $A-A$ and $A+A$ and found two regimes of behavior separated by a phase transition based on the decay rate of $p(N)$.

Theorem 2.1 (Hegarty-Miller). *Let $\mathcal{S}$ be a random variable representing $|A+A|$ and $\mathcal{D}$ be a random variable representing $|A-A|$ with A and $p(N)$ as described above. There are three possible behaviors for these random variables:*

- $p(N) = o\left(N^{-1/2}\right)$: *We have $\mathcal{D} \sim 2\mathcal{S} \sim (N \cdot p(N))^2$.*
- $p(N) = cN^{-1/2}$ *for some $c \in (0, \infty)$: We have $\mathcal{S} \sim g\left(\frac{c^2}{2}\right) N$ and $\mathcal{S} \sim g\left(c^2\right) N$ where $g(x) := 2\left(\frac{e^{-x} - (1-x)}{x}\right)$.*
- $N^{-1/2} = o(p(N))$: *Letting $\mathcal{S}^c = (2N+1) - \mathcal{S}$ and $\mathcal{D}^c = (2N+1) - \mathcal{D}$, we have $\mathcal{S}^c \sim 2\mathcal{D} \sim \frac{4}{p(N)^2}$.*

Generalizing to $h > 2$: In the previous case, we saw a phase transition at $\delta = \frac{1}{2}$. In general, we expect a phase transition at $\delta = \frac{h-1}{h}$. For $p(N) = cN^{-\delta}$ for some positive constant c with $\delta \geq \frac{h-1}{h}$, Hogan and Miller proved a result similar to the above theorem:

Theorem 2.2 (Hogan-Miller). *Let*

$$g(x; s, d) := \sum_{k=1}^{\infty} (-1)^{k-1} \frac{b_{h,k}}{(s!d!)^k} x^{(s+d)k},$$

where $b_{h,k}$ is a constant dependent on h and k. Fix integers $h \geq 2$ and $m_1, n_1, m_2, n_2 \geq 0$ s.t. $m_1 + n_1 = m_2 + n_2 = h$, $m_i \geq n_i$, and $n_1 > n_2$. Consider A and $p(N)$ as above. There are two possible behaviors for the sizes of the generalized sum and difference sets:

- $\delta > \frac{h-1}{h}$: *As $N \to \infty$, with probability one we have $|m_1A - n_1A| / |m_2A - n_2A| = (m_2!n_2!) / (m_1!n_1!) + o(1)$.*
- $\delta = \frac{h-1}{h}$: *Almost surely $|A_{m_i n_i}| \sim Ng(c; m_i, n_i)$ and thus with probability one $|m_1A - n_1A| / |m_2A - n_2A| = g(c; m_1, n_1) / g(c; m_2, n_2) + o(1)$.*

This theorem leaves open the case of slow decay, $\delta < \frac{h-1}{h}$, and we hope to shed some light on this case.

3. Obstructions in Analysis of Slow Decay

The case of slow decay with $\delta < \frac{h-1}{h}$ is not well understood. To analyze this case for $h = 2$, Hegarty and Miller analyze the distribution of missing terms from $mA - nA \subseteq \{-nA, -nA+1, \ldots, mA-1, mA\}$. For $A+A$ and $A-A$, these distributions are fairly well understood. Hegarty and Miller consider

$$\mathbb{E}[\mathcal{S}^c] = \sum_{n=0}^{2N} \mathbb{P}(\mathcal{E}_n)$$

where $\mathcal{E}_n$ is the event that $n \notin A+A$. Using the fact that any two ways of writing an element of $A+A$ as a sum of elements of A are independent along with the geometric series formula, they are able to prove that this expectation is $\frac{4}{p^2}$. They then prove that $\mathcal{S}^c$ is strongly concentrated about its mean. While some details must be modified for the case of $A-A$, the argument is fairly similar. In the general case, the representations of an element of $mA - nA$ 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 $mA - nA$.

4. Missing Term Distributions

We start by analyzing the distribution of missing terms from $mA - nA$ when each element of A is chosen from $\{0, \ldots, N\}$ with fixed probabilty p. We ran simulations of 2^{17} trials with $N = 2^9$ for all choices of $m \geq 0$ and $n \leq 0$ with $2 \leq m+n \leq 5$. We include a graph of the missing term distribution for $4A$ as this distribution is fairly representative of the patterns we see in general.

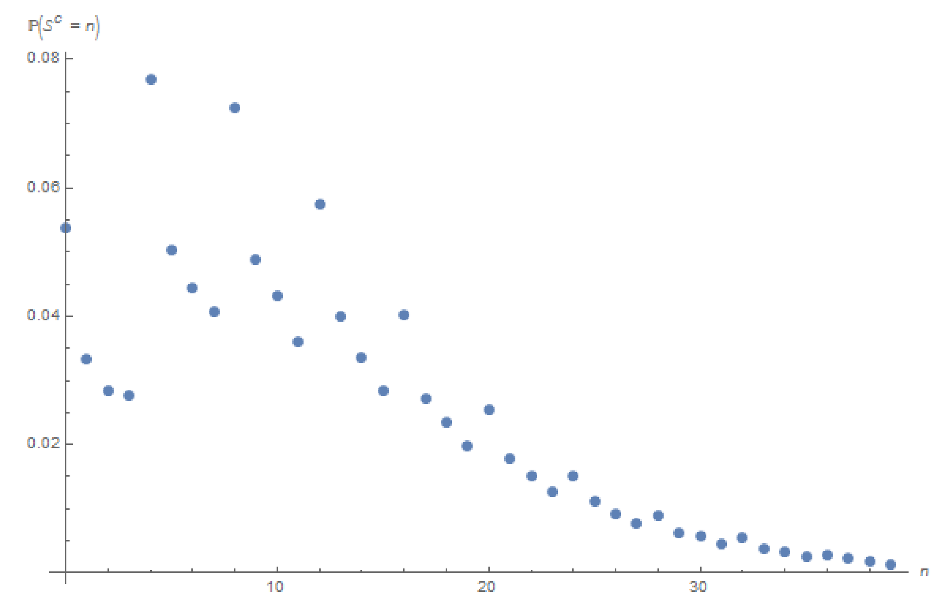

Figure 2: *The distribution of the number of missing sums for $4A$ (the random variable $\mathcal{S}^c = 4N+1-|4A|$) given A chosen uniformly at random from $P(\{0, \ldots, N\})$ with $N = 2^9$ and 2^{17} trials.*

The distribution in Figure 2 seems to have a repeating pattern of vertical ridges of 4 points each. If we look only at numbers of missing sums congruent to some fixed $k \mod 4$, we can pick out 4 seemingly smooth distributions (or discrete approximations of them). 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 $mA - nA$. 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 $4A$ with $p(N) = N^{-1/4}$ is shown in Figure 3. The expectation seems to grow as $cN^{1/2}$ for some constant c, which is similar to the case of $2A$. However, values of N are fairly small for this simulation, so we cannot be sure that the limiting behavior has set in.

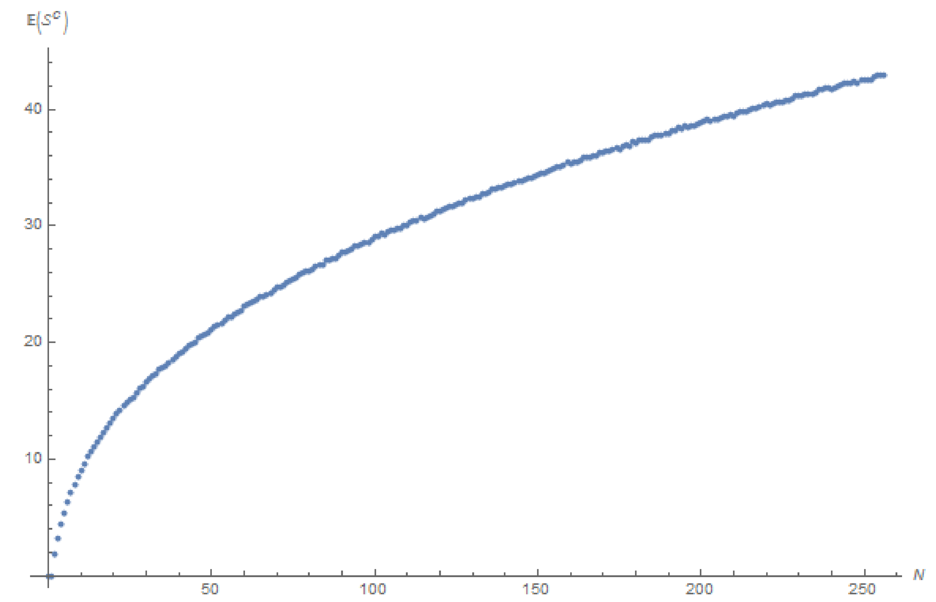

Figure 3: *The expectation of the number of missing sums from $4A$ (the random variable $\mathcal{S}^c = 4N+1-|4A|$) as a function of N given $A \subseteq \{0, \ldots, N\}$, where each element is chosen with probability $p(N) = N^{-1/4}$. Each N is simulated with 2^{16} trials.*

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 DMS1659037.

References

[1] V. Hogan and S. J. Miller, *When Generalized Sumsets Are Difference Dominated*, preprint. https://arxiv.org/abs/1301.5703

[2] P.V. Hegarty and S.J. Miller, *When almost all sets are difference dominated*, Random Structures and Algorithms **35** (2009), no. 1, 118-136.

[3] G. Iyer, O. Lazarev, S. J. Miller and L. Zhang, *Generalized More Sums Than Differences Sets*, Journal of Number Theory **132** (2012), no. 5, 1054-1073.

[4] O. Lazarev, S. J. Miller and K. O'Bryant, *Distribution of Missing Sums in Sumsets*, to appear in Experimental Mathematics. http://arxiv.org/abs//1109.4700.

[5] G. Martin and K. O'Bryant, *Many sets have more sums than differences*, Additive combinatorics, 287-305, CRM Proc. Lecture Notes **43**, Amer. Math. Soc., Providence, RI, 2007.