GENERALIZING THE DISTRIBUTION OF MISSING SUMS IN SUMSETS

HÙNG VIỆT CHU, DYLAN KING, NOAH LUNTZLARA, THOMAS C. MARTINEZ, STEVEN J. MILLER, LILY SHAO, CHENYANG SUN, AND VICTOR XU

ABSTRACT. Given a finite set of integers A, its sumset is $A + A := \{a_i + a_j \mid a_i, a_j \in A\}$. We examine $|A + A|$ as a random variable, where $A \subset I_n = [0, n-1]$, the set of integers from 0 to $n-1$, so that each element of I_n is in A with a fixed probability $p \in (0, 1)$. Recently, Martin and O'Bryant studied the case in which $p = 1/2$ and found a closed form for $\mathbb{E}[|A + A|]$. Lazarev, Miller, and O'Bryant extended the result to find a numerical estimate for Var($|A + A|$) and bounds on $m_{n\,;\,p}(k) := \mathbb{P}(2n - 1 - |A + A| = k)$. Their primary tool was a graph-theoretic framework which we now generalize to provide a closed form for $\mathbb{E}[|A+A|]$ and Var($|A + A|$) for all $p \in (0, 1)$ and establish good bounds for $\mathbb{E}[|A + A|]$ and $m_{n\,;\,p}(k)$.

We continue to investigate $m_{n\,;\,p}(k)$ by studying $m_p(k) = \lim_{n\to\infty} m_{n\,;\,p}(k)$, proven to exist by Zhao. Lazarev, Miller, and O'Bryant proved that, for $p = 1/2$, $m_{1/2}(6) > m_{1/2}(7) < m_{1/2}(8)$. This distribution is not unimodal, and is said to have a "divot" at 7. We report results investigating this divot as p varies, and through both theoretical and numerical analysis prove that for $p \geq 0.68$ there is a divot at 1; that is, $m_{1/2}(0) > m_{1/2}(1) < m_{1/2}(2)$.

Finally, we extend the graph-theoretic framework originally introduced by Lazarev, Miller, and O'Bryant to correlated sumsets $A + B$ where B is correlated to A by the probabilities $\mathbb{P}(i \in B \mid i \in A) = p_1$ and $\mathbb{P}(i \in B \mid i \notin A) = p_2$. We provide some preliminary results using the extension of this framework.

CONTENTS

1. INTRODUCTION

Many problems in additive number theory, such as Fermat's Last Theorem or the Goldbach conjecture or the infinitude of twin primes, can be cast as problems involving sum or difference sets. For example, if P_n is the set of n^{th} powers of positive integers, Fermat's Last Theorem is equivalent to $(P_n + P_n) \cap P_n = \varnothing$ for

Date: May 16, 2020.
Key words and phrases. Sumsets, More Sums Than Differences sets, vertex cover, correlated sets, divot.
This work was partially supported by NSF grants DMS1659037 and DMS1561945.

$n \geq 3$. Given a finite set of non-negative integers A, we define the sumset $A+A := \{a_i + a_j \mid a_i, a_j \in A\}$ and the difference set $A - A := \{a_i - a_j \mid a_i, a_j \in A\}$. The set A is said to be

- *sum-dominant* if $|A+A| > |A-A|$ (also called *MSTD*, or *More Sums Than Differences*),
- *balanced* if $|A+A| = |A-A|$, and
- *difference-dominant* if $|A+A| < |A-A|$.

By $[a, b]$, we mean the set of integers $\{a, a+1, \dots, b\}$. As addition is commutative and subtraction is not, it was expected that in the limit almost all sets would be difference-dominant, though there were many constructions of infinite families of MSTD sets.[1] There is an extensive literature on such sets, their constructions, and generalizations to settings other than subsets of the integers; see for example [AMMS, BELM, CLMS, CMMXZ, DKMMW, He, HLM, ILMZ, Ma, MOS, MS, MPR, MV, Na1, Na2, PW, Ru1, Ru2, Ru3, Sp, Zh1].

We are interested in studying $|A+A|$ as we randomly choose A using a Bernoulli process. Explicitly, we fix a $p \in (0,1)$ and construct $A \subset [0, n-1]$ by independently including each $i \in [0, n-1]$ to be in A with probability p. Martin and O'Bryant [MO] studied the distributions of $|A+A|$ and $|A-A|$ when $p = 1/2$, including computing the expected values. Contrary to intuitions, they proved a positive percentage of these sets are MSTD in the limit as $n \to \infty$. Note that $p = 1/2$ is equivalent to the model where each subset of $[0, n-1]$ is equally likely to be chosen. Their work extends to any fixed $p > 0$, though if p is allowed to decay to zero with n then the intuition is correct and almost all sets are difference dominated [HM].

Lazarev, Miller and O'Bryant [LMO] continued this program in the special but important case of $p = 1/2$. They computed the variance of $|A+A|$, showed that the distribution is asymptotically exponential, and proved the existence of a "divot", which we now explain. From [MO], the expected number of missing sums is 10 as $n \to \infty$; thus almost all sets are missing few sums, making it more convenient to plot the distribution of the number of missing sums. For $A \subset [0, n-1]$, we set $m_{n\,;\,p}(k) := \mathbb{P}(2n - 1 - |A+A| = k)$, and examine the distribution of $m_p(k) := \lim_{n\to\infty} m_{n\,;\,p}(k)$, proven to exist by Zhao [Zh2]. The distribution does not just rise and fall, but forms a 'divot', with $m_{1/2}(6) > m_{1/2}(7) < m_{1/2}(8)$; see Figure 1 for data and [LMO] for a proof.

We extend the methodologies developed in [LMO] to study the distribution of $|A+A|$ for generic p not necessarily equal to $1/2$; there are many technical issues that arise which greatly complicate the combinatorial analysis when $p \neq 1/2$. To do so, we generalize many previous results in Section 2, and use them to derive a formula for the expected value of $|A+A|$, which we then analyze.

Theorem 1.1. *Let $A \subseteq [0, n-1]$ with $\mathbb{P}(i \in A) = p$ for $p \in (0,1)$. Then $\mathbb{E}[|A+A|]$ equals*

$$\sum_{r=0}^{n} \binom{n}{r} p^r q^{n-r} \left(2 \sum_{i=0}^{n-2} (1 - \mathbb{P}(i \notin A+A \mid |A| = r)) + (1 - \mathbb{P}(n-1 \notin A+A \mid |A| = r)) \right), \tag{1.1}$$

where $q = 1 - p$ and

$$\mathbb{P}(i \notin A+A \mid |A| = r) = \begin{cases} \dfrac{\displaystyle\sum_{k=\frac{i+1}{2}}^{i+1} 2^{i+1-k} \binom{\frac{k+1}{2}}{k - \frac{i+1}{2}} \binom{n-i-1}{n-r-k}}{\dbinom{n}{n-r}} & \text{for } i \text{ odd} \\[4ex] \dfrac{\displaystyle\sum_{k=\frac{i}{2}}^{i} 2^{i-k} \binom{\frac{i}{2}}{k - \frac{i}{2}} \binom{n-i-1}{n-r-1-k}}{\dbinom{n}{n-r}} & \text{for } i \text{ even.} \end{cases} \tag{1.2}$$

[1]The proportion of sets in $[0, n-1]$ in these families tend to zero as $n \to \infty$. In the early constructions these densities tended to zero exponentially fast, but recent methods have found significantly larger ones where the decay is polynomial.

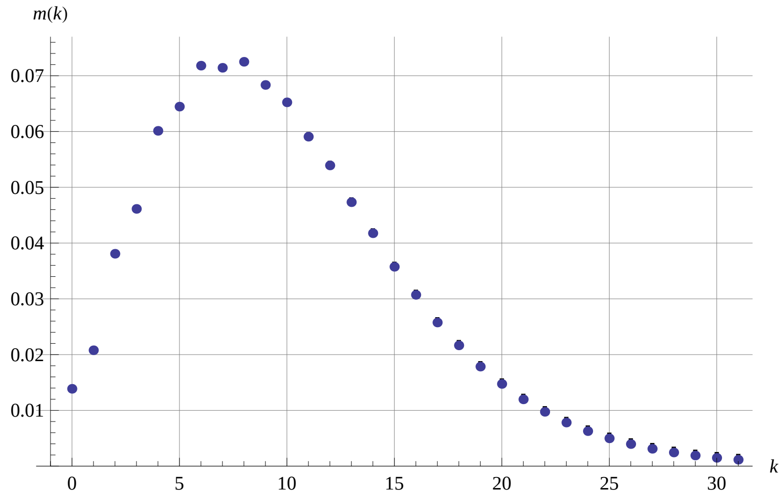

FIGURE 1. From [LMO]: Experimental values of $m_p(k)$, where $p = 1/2$, with vertical bars depicting the values allowed by our rigorous bounds. In most cases, the allowed interval is smaller than the dot indicating the experimental value. The data comes from generating 2^{28} sets uniformly forced to contain 0 from $[0, 256)$.

As we need to compute on the order of n^3 sums to compute $\mathbb{E}[|A + A|]$, an useful bound is needed for numerical investigations.

Theorem 1.2. *Let $A \subseteq [0, n-1]$ with $\mathbb{P}(i \in A) = p$ for $p \in (0, 1)$ and set $q = 1 - p$. Then*

$$\mathbb{E}[|A + A|] \ \leq \ 2n - 1 - 2q \, \frac{1 - q^{\frac{n-1}{2}}}{1 - \sqrt{q}}. \tag{1.3}$$

If $p > 1/2$, then we also get

$$\mathbb{E}[|A + A|] \ \geq \ 2n - 1 - 2q \, \frac{1}{1 - \sqrt{2q}} - (2q)^{\frac{n-1}{2}} \, . \tag{1.4}$$

The proofs of Theorems 1.1 and 1.2 are given in Section 4. The proofs require an extension of the graph-theoretic framework of [LMO], which is done in Section 3.

We also compute the variance of $|A + A|$.

Theorem 1.3. *Let $A \subseteq [0, n-1]$ with $\mathbb{P}(i \in A) = p$ with $p \in (0, 1)$. Then*

$$\mathrm{Var}(|A+A|) = \sum_{r=0}^{n} \binom{n}{r} p^r \, q^{n-r} \left(2 \sum_{0 \leq i < j \leq 2n-2} 1 - P_r(i,j) + \sum_{0 \leq i \leq 2n-2} 1 - P_r(i) \right) - \mathbb{E}[|A+A|]^2, \tag{1.5}$$

where $q = 1 - p$, $\mathbb{E}[|A + A|]$ is as calculated in Theorem 1.1, $P_r(i) = \mathbb{P}(i \notin A + A \mid |A| = r)$ and $P_r(i,j) = \mathbb{P}(i \text{ and } j \notin A + A \mid |A| = r)$.

As opposed to the calculation for the expected value, the variance does not have an easily calculable closed form, as $\mathbb{P}(i \text{ and } j \notin A + A \mid |A| = r)$ has on the order of $p(n)$ terms to calculate, where $p(n)$ is the partition number of n and grows faster than any polynomial.[2] We discuss these issues in Section 5, where we prove Theorem 1.3. The difficulty arises as we go from $\mathbb{P}(i \notin A + A \mid |A| = r)$ to $\mathbb{P}(i \text{ and } j \notin A + A \mid |A| = r)$ because we introduce many more dependencies between nodes in the graph-theoretic framework. We were, however, able to show that the number of missing sums is asymptotically exponential.

[2]One has $\log p(n) \sim \pi\sqrt{3/2}n^{1/2}$.

Theorem 1.4. *Let $A \subseteq [0, n-1]$ with $\mathbb{P}(i \in A) = p$ for $p \in (0,1)$ and recall that $m_{n\,;\,p}(k) := \mathbb{P}(2n - 1 - |A+A| = k)$. If $n > 2\frac{\log(1-p)}{\log(1-p^2)}k$, then*

$$q^{k/2} \ \ll \ m_{n\,;\,p}(k) \ \ll \ \left(\frac{1-p+\phi(p)}{2}\right)^k,$$

where $\phi(p) \ = \ \sqrt{1+2p-3p^2}$.

The proof of Theorem 1.4 is structurally equivalent to the proof of Theorem 1.2 in [LMO]. The full proof is in Appendix A, but the idea is to study specific scenarios that are less likely or more likely to happen, for the lower bound and the upper bound, respectively, using many of the results proven in Section 2.

Finally, we investigate the shape of the distribution of $m_p(k)$. Recall that Zhao proved that $m_p(k)$ existed by fringe analysis [Zh2]. The technique of fringe analysis for estimating numbers of missing sums is the means by which most results about sum-distributions have been obtained and is the method we will follow. The technique grows out of the observation that sumsets usually have fully populated centers: there is a very low probability that there will be any element missing that is not near one of the ends. When a suitable distance from the edge is chosen, this observation can be made precise by bounding the probability of missing any elements in the middle. It follows that most of the time, all missing sums must be near the ends of the interval; the only contribution to these elements is from the upper and lower fringes of the randomly chosen set. Conveniently, as long as they are short relative to the length of the whole set, the fringes are independent and can be analyzed separately from the rest of the elements. As long as they are reasonably sized (between 20 and 30 elements, usually) a computer can numerically check by brute force all the possible fringe arrangements, and give exact data for the number of missing sums near the edges.

Working with difference sets is *orders of magnitude* more challenging than working with sumsets. This is because the fringe method fails, since there are interactions between the upper and lower fringes when we consider difference sets and thus the computational time is the square of that for sumsets. No suitable alternative technique has been found, and so rigorous numerical results about difference sets are scarce. Most work, including ours, focuses on distributions of sums, though see [H-AMP] for some results on differences.

The results for $p = 1/2$ were possible because there were nice interpretations for the terms that simplified the analysis; we do not have that in general, which is why our results are concentrated on the larger values of p; see Figure 2. In Section 6, we look for divots other than that at $m_{1/2}(7)$, and our main theorem is the following.

Theorem 1.5. *For $p \geq 0.68$, there is a divot at 1; that is, $m_p(0) > m_p(1) < m_p(2)$.*

Our final result looks at the generalization of our work to correlated sumsets (see [DKMMW] for earlier work and results). We examine the random variable $|A+B|$, where, for a given triplet (p, p_1, p_2) and any $i \in \{0, \dots, n-1\}$ we have

- $\mathbb{P}(i \in A) = p$,
- $\mathbb{P}(i \in B \mid i \in A) = p_1$, and
- $\mathbb{P}(i \in B \mid i \notin A) = p_2$.

We extend our graph-theoretic framework to analyze this system and find $\mathbb{P}(k \notin A+B)$ and $\mathbb{P}(i \text{ and } j \notin A+B)$ in Section 7. We end by considering some work that can be done in continuation of that presented here, in Section 8.

Note on the Computations.

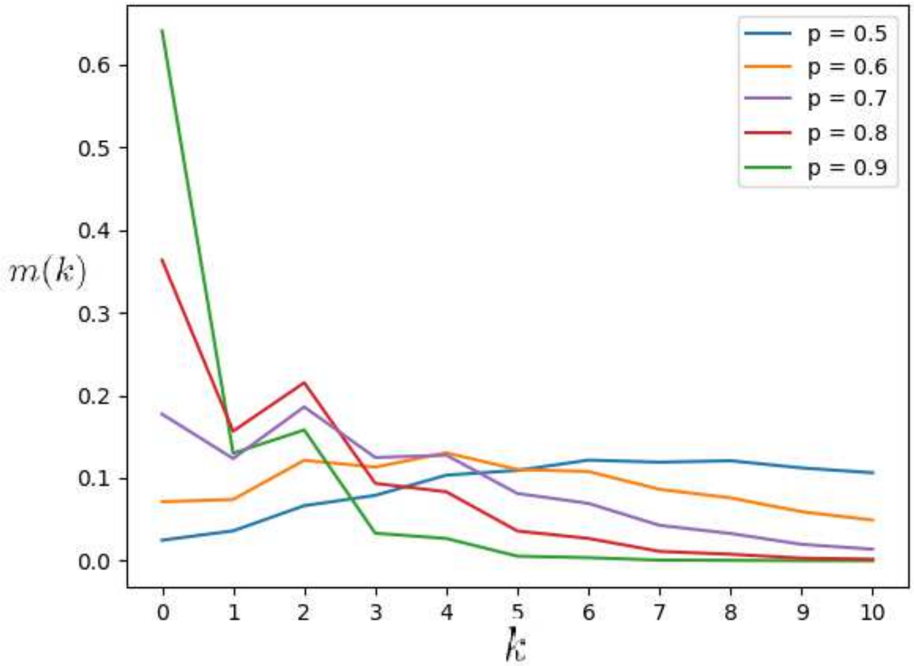

FIGURE 2. Plot of numerical approximations to $m_p(k)$, varying p by simulating 10^6 subsets of $\{0, 1, 2, \ldots, 400\}$. The simulation shows that: for $p = 0.9$ and 0.8, there is a divot at 1, for $p = 0.7$, there are divots at 1 and 3, for $p = 0.6$, there is a divot at 3 and for $p = 0.5$, there is a divot at 7.

We end this section by mentioning our method of collecting data. Our program uses the technique of fringe analysis, and is able to do the calculations for any value of p simultaneously by doing one pre-computation. To do this, the program constructs all the fringe sets of a given size at once; the probability we choose each set of a given size can be computed easily. In particular, let ℓ be the fringe size (i.e. the width of the fringe) and let k be the number of elements in our set. The probability we choose this set is $p^k(1-p)^{\ell-k}$. We then compute how many missing sums it has and store this information in an array, whose elements correspond to the probabilities $p^\ell, p^{\ell-1}(1-p), p^{\ell-2}(1-p)^2, \ldots, (1-p)^\ell$. For each set, we only save its number of missing sums and its number of elements. Using this method, we can find the probability of missing different number of sums by simply changing the value of p.

We picked a fringe size of $\ell = 30$ to compute all the data necessary for Theorem 1.5 about the divot at 1 (see Section §6 and Appendix C for details). The process took three days on a single machine without code parallelization. We were initially hoping to get results about the divot at 3 with the same fringe analysis by using a larger fringe size $\ell = 40$ and parallelizing the code. Unfortunately, this was too big a search space. We were able to use the shared Linux computing cluster at Williams College for six months, which was enough time to do three-fourths of the computation. As the run-time for difference set computations is on the order of the square of the length of sumsets (as the two fringes interact), these run-times illustrate the challenges in numerically exploring difference sets.

2. Generalizations of [MO]

We need to extend many of the lemmas and propositions from [MO], and prove they are true for general p and not just $p = 1/2$. The arguments typically do not change, we only introduce notation as necessary[3], thus, we just state the results we use and how we generalized the argument. The full proofs are in Appendix A.

Lemma 2.1 (Lemma 5 of [MO]). *Let n, ℓ, u be integers with $n \geq \ell + u$. Fix $L \subset \{0, \dots, \ell - 1\}$ and $U \subset \{n-u, \dots, n-1\}$. Suppose R is a random subset of $\{\ell, \dots, n-u-1\}$, where each element of $\{\ell, \dots, n-u-1\}$ is in R with independent probability $p \in (0,1)$, and define $A := L \cup R \cup U$ and $q := 1 - p$. Then for any integer i satisfying $2\ell - 1 \leq i \leq n - u - 1$, we have*

$$\mathbb{P}(i \notin A + A) \;=\; \begin{cases} q^{|L|}(1-p^2)^{\frac{i+1}{2}-\ell} & \text{if } i \text{ odd} \\ q^{|L|+1}(1-p^2)^{\frac{i}{2}-\ell} & \text{if } i \text{ even.} \end{cases} \tag{2.1}$$

Lemma 2.2 (Lemma 6 of [MO]). *Let n, ℓ, u be integers with $n \geq \ell + u$. Fix $L \subset \{0, \dots, \ell - 1\}$ and $U \subset \{n-u, \dots, n-1\}$. Suppose R is a random subset of $\{\ell, \dots, n-u-1\}$, where each element of $\{\ell, \dots, n-u-1\}$ is in R with independent probability $p \in (0,1)$, and define $A := L \cup R \cup U$ and $q := 1 - p$. Then for any integer i satisfying $n + \ell - 1 \leq i \leq 2n - 2u - 1$, we have*

$$\mathbb{P}(i \notin A + A) \;=\; \begin{cases} q^{|U|}(1-p^2)^{n-\frac{i+1}{2}-u} & \text{if } i \text{ odd} \\ q^{|U|+1}(1-p^2)^{n-1-\frac{i}{2}-u} & \text{if } i \text{ even.} \end{cases} \tag{2.2}$$

Lemma 2.3. *Choose $A \subseteq [0, n-1]$ by including each element with probability p. Set $q = 1 - p$. Then, for $0 \leq i \leq n - 1$, the probability*

$$\mathbb{P}(i \notin A + A) \;=\; \begin{cases} (2q - q^2)^{(i+1)/2} & \text{if } i \text{ odd} \\ q\,(2q-q^2)^{i/2} & \text{if } i \text{ even,} \end{cases} \tag{2.3}$$

while for any integer $n - 1 \leq i \leq 2n - 2$ the probability

$$\mathbb{P}(i \notin A + A) \;=\; \begin{cases} (2q - q^2)^{n-(i+1)/2}, & \text{if } i \text{ odd} \\ q\,(2q-q^2)^{n-1-i/2}, & \text{if } i \text{ even.} \end{cases} \tag{2.4}$$

These give us a generalization of Proposition 8 from [MO].

Proposition 2.4 (Proposition 8 of [MO]). *Let n, ℓ, u be integers with $n \geq \ell + u$. Fix $L \subset \{0, \dots, \ell - 1\}$ and $U \subset \{n-u, \dots, n-1\}$. Suppose R is a random subset of $\{\ell, \dots, n-u-1\}$, where each element of $\{\ell, \dots, n-u-1\}$ is in R with independent probability $p \in (0,1)$, and define $A := L \cup R \cup U$ and $q := 1 - p$. Then the probability that*

$$\{2\ell - 1, \dots, n-u-1\} \cup \{n + \ell - 1, \dots, 2n - 2u - 1\} \;\subseteq\; A + A \tag{2.5}$$

is greater than $1 - \frac{1+q}{p^2}\left(q^{|L|} + q^{|U|}\right)$.

3. Graph-Theoretic Framework

We develop a graph-theoretic framework which has proved powerful in computing various probabilities used in calculations. As we have shown in Section 2, we have an explicit formula for $\mathbb{P}(i \notin A + A)$ (Lemmas 2.1 and 2.2). However for generic i and j, $\mathbb{P}(i \notin A + A)$ and $\mathbb{P}(j \notin A + A)$ are dependent, and therefore $\mathbb{P}(i \text{ and } j \notin A + A)$ requires more work. To understand the dependencies between these two events, we create a *condition graph*, as defined in [LMO], with some slight modifications. In [LMO], $V = [0, \max\{i, j\}]$, while we use $V = [0, n-1]$. This distinction is because in [LMO] there was no need to consider the unconnected vertices, but here they will prove meaningful for computations.

[3] We just replace $1/2$ and $3/4$ with q and $1 - p^2$, respectively, as these are the representations of the exact values used in [MO].

Definition 3.1. *For a set $F \subseteq [0, 2n-2]$ we define the* condition graph $G_n = (V, E)$ *induced on F where $V = [0, n-1]$, and for two vertices k_1 and k_2, $(k_1, k_2) \in E$ if $k_1 + k_2 \in F$.*

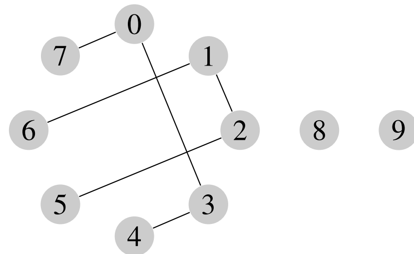

FIGURE 3. Condition Graph for $\mathbb{P}(3 \text{ and } 7 \notin A + A), n = 9$.

See Figure 3 for the condition graph G_9 induced on $F = \{3, 7\}$.

By construction, as [LMO] explains, viewing our vertices as the integers $[0, n-1]$ we have a bijection between edges and pairs of elements whose sum belongs to F. If we suppose that $F \cap (A + A) = \varnothing$, then, for each pair of elements whose sum belongs to F, at least one of the pair must be excluded from A. The corresponding criteria in the condition graph is that each edge must be incident to a vertex which corresponds to an integer missing from A. That is, $F \cap (A + A) = \varnothing$ exactly when $[0, n-1] \setminus A$ forms a vertex cover on G_n (recall a vertex cover of a graph is a set of vertices such that each edge is incident to at least one vertex in the set). We therefore find the following result (Lemma 2.1 from [LMO]).

Lemma 3.2. *For a set $F \subseteq [0, 2n-2]$, $\mathbb{P}(F \cap (A + A) = \varnothing)$ is the probability that we choose a vertex cover for the condition graph G_n induced on F.*

Note that when we choose vertices in the condition graph for our vertex cover, we are choosing elements to **exclude** from A.

We now find a closed form for $\mathbb{P}(i \text{ and } j \notin A + A)$. By Lemma 3.2, we only need to study the condition graph G_n induced on $\{i, j\}$. From Proposition 3.1 of [LMO], each component in our condition graph G_n is a *segment graph*, a graph that consists of a sequence of vertices such that each vertex is connected only to the vertices to its immediate left and right, or an isolated vertex. Since we only add isolated vertices to [LMO]'s definition of a condition graph, the proof of their Proposition 3.1 applies to our condition graph. As we are interested in counting vertex covers and there are no edges between different components, the behavior of each component is independent. That is, the probability of finding a vertex cover for the entire graph is the product of the probability of finding a vertex cover on each component. In this way, we reduce the problem at hand to computing the probability of finding a vertex cover on a segment graph, which we do in the following proposition.

Lemma 3.3. *Let S be a subset of the vertices of a segment graph on the n vertices V, with each vertex included in S with probability p. If we set $a_n = \mathbb{P}(V \setminus S \text{ is a vertex cover})$, then*

$$a_n = \frac{(\phi(p) - 1 - p)\,(1 - p - \phi(p))^n + (\phi(p) + 1 + p)\,(1 - p + \phi(p))^n}{2^{n+1}\,\phi(p)}, \tag{3.1}$$

where $\phi(p) = \sqrt{1 + 2p - 3p^2}$.

Proof. The proof follows from a simple recurrence relation. We see that $a_1 = 1$, as this path does not have loops, so we cannot have an edge if only one vertex exists. Also, $a_2 = 1 - p^2$, as the only case in which we do not get a vertex cover is when both vertices $v_1, v_2 \in S$. This happens with probability p^2, as each event is independent. We now find a recurrence relation to compute a_n.

In Figure 4 we see that if $v_n \notin S$, then we can recur on the remaining $n - 1$ vertices, as the edge connecting v_n to v_{n-1} has an incident vertex in the complement of S. This corresponds to $(1 - p)\,a_{n-1}$.

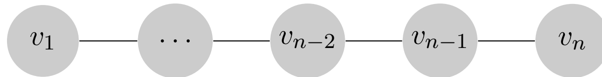

FIGURE 4. A path with n vertices; we may count vertex covers recursively by handling the behavior of the n^{th} vertex.

However, if $v_n \in S$, we must necessarily have $v_{n-1} \notin S$ for $V \setminus S$ to be a vertex cover, and then we can recur on the remaining $n-2$ vertices. This corresponds to $p\,(1-p)\,a_{n-2}$. So, we find that

$$a_n \;=\; (1-p)\,a_{n-1} + p\,(1-p)\,a_{n-2},$$

with $a_1 = 1$ and $a_2 = 1 - p^2$. Solving this gives us (3.1), as desired. $\square$

Now, to find $\mathbb{P}(i \text{ and } j \notin A + A)$, we only need the number and size of the segment graphs of the condition graph G_n induced by $\{i, j\}$. Fortunately, [LMO] derived formulas for the number and size of segment graphs (their Proposition 3.5). Using these, we find

Proposition 3.4. *Consider i, j such that $i < j$.*
For i, j both odd:

$$\mathbb{P}(i \textit{ and } j \notin A + A) \;=\; a_q^s\, a_{q+2}^{s'} \tag{3.2}$$

where

$$\begin{aligned} q &= 2\left\lceil \frac{i+1}{j-i} \right\rceil \\ s &= \frac{1}{2}\left((j-i)\left\lceil \frac{i+1}{j-i} \right\rceil - (i+1) \right) \\ s' &= \frac{1}{2}\left(j+1-(j-i)\left\lceil \frac{i+1}{j-i} \right\rceil \right). \end{aligned} \tag{3.3}$$

For i even, j odd:

$$\mathbb{P}(i \textit{ and } j \notin A + A) \;=\; a_o\, a_q^s\, a_{q+2}^{s'} \tag{3.4}$$

where

$$\begin{aligned} o &= 2\left\lceil \frac{i/2+1}{j-i} \right\rceil - 1 \\ q &= 2\left\lceil \frac{i+1}{j-i} \right\rceil \\ s &= \frac{1}{2}\left((j-i-1)\left\lceil \frac{i+1}{j-i} \right\rceil - (i+1) + o \right) \\ s' &= \frac{1}{2}\left(j-(j-i-1)\left\lceil \frac{i+1}{j-i} \right\rceil - o \right). \end{aligned} \tag{3.5}$$

For i odd, j even:

$$\mathbb{P}(i \textit{ and } j \notin A + A) \;=\; a_{o'}\, a_q^s\, a_{q+2}^{s'} \tag{3.6}$$

where

$$\begin{aligned}
o' &= 2\left\lceil \frac{j/2+1}{j-i} \right\rceil - 2 \\
q &= 2\left\lceil \frac{i+1}{j-i} \right\rceil \\
s &= \frac{1}{2}\left((j-i-1)\left\lceil \frac{i+1}{j-i} \right\rceil - (i+1) + o' \right) \\
s' &= \frac{1}{2}\left(j - (j-i-1)\left\lceil \frac{i+1}{j-i} \right\rceil - o' \right).
\end{aligned} \tag{3.7}$$

For i, j both even:

$$\mathbb{P}(i \textit{ and } j \notin A+A) \;=\; a_o\, a_{o'}\, a_q^s\, a_{q+2}^{s'} \tag{3.8}$$

where

$$\begin{aligned}
o &= 2\left\lceil \frac{i/2+1}{j-i} \right\rceil - 1 \\
o' &= 2\left\lceil \frac{j/2+1}{j-i} \right\rceil - 2 \\
q &= 2\left\lceil \frac{i+1}{j-i} \right\rceil \\
s &= \frac{1}{2}\left((j-i-2)\left\lceil \frac{i+1}{j-i} \right\rceil - (i+1) + o + o' \right) \\
s' &= \frac{1}{2}\left(j - 1 - (j-i-2)\left\lceil \frac{i+1}{j-i} \right\rceil - o - o' \right).
\end{aligned} \tag{3.9}$$

The proof of Proposition 3.4 is structurally identical to that of Proposition 3.5 of [LMO], as we have already shown independence of segment graphs, so we must show how to obtain the number of segment graphs and their size, which was done in [LMO]. We now bound $\mathbb{P}(i \text{ and } j \notin A+A)$. We note, from Equation (3.1), that if n is even, then

$$a_n \;\le\; \frac{(\phi(p)+1+p)\,(1-p+\phi(p))^n}{2^{n+1}\,\phi(p)}. \tag{3.10}$$

Since q and $q+2$ are always even, for odd i, j, we have

$$\begin{aligned}
\mathbb{P}(i \text{ and } j \notin A+A) &= a_q^r\, a_{q+2}^{r'} \\
&\le \left(\frac{(\phi(p)+1+p)\,(1-p+\phi(p))^q}{2^{q+1}\,\phi(p)} \right)^r \left(\frac{(\phi(p)+1+p)\,(1-p+\phi(p))^{q+2}}{2^{q+3}\,\phi(p)} \right)^{r'} \\
&= \frac{(\phi(p)+1+p)^{r+r'}\,(1-p+\phi(p))^{qr+(q+2)r'}}{2^{(q+1)\,r+(q+3)\,r'}\,\phi(p)^{r+r'}} \\
&= \left(\frac{\phi(p)+1+p}{2\phi(p)} \right)^{r+r'} \left(\frac{1-p+\phi(p)}{2} \right)^{qr+(q+2)r'} \\
&= \left(\frac{\phi(p)+1+p}{2\phi(p)} \right)^{\frac{j-i}{2}} \left(\frac{1-p+\phi(p)}{2} \right)^{j+1},
\end{aligned} \tag{3.11}$$

where the last equality comes from (3.18) in [LMO]. We can use Proposition 3.4 to show (3.11) holds for all i, j.

4. Expected Value

To compute $\mathbb{E}[|A+A|]$, we see that

$$\begin{aligned}
\mathbb{E}[|A+A|] &= \sum_{A\subseteq\{0,\dots,n-1\}} |A+A|\cdot\mathbb{P}(A) \\
&= \sum_{r=0}^{n}\binom{n}{r}p^r q^{n-r}\sum_{i=0}^{2n-2}\sum_{\substack{A\subseteq\{0,\dots,n-1\},|A|=r\\ i\,\in\, A}} 1 \\
&= \sum_{r=0}^{n}\binom{n}{r}p^r q^{n-r}\sum_{i=0}^{2n-2}\mathbb{P}(i\in A+A\mid |A|=r). \qquad (4.1)
\end{aligned}$$

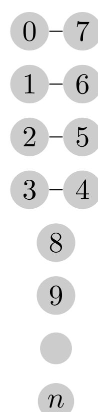

Figure 5. Condition Graph for $\mathbb{P}(7\notin A+A)$.

Now we compute $\mathbb{P}(i\in A+A\mid |A|=r) = 1-\mathbb{P}(i\notin A+A\mid |A|=r)$. To compute this probability, we refer once again to the condition graph G_n induced by i. If we assume[4] that $i\le n-1$, this graph has n vertices and $\frac{i+1}{2}$ and $\frac{i}{2}$ disjoint edges with one loop. See Figure 5 for a visualization.

By Lemma 3.2, the event $i\notin A+A$ corresponds to when the elements not in S form a vertex cover of G_n. Since we are conditioning that $|S|=r$, we must count the number of ways that the $n-r$ missing elements may be chosen so that they form a vertex cover. Then we obtain the following:

[4]The $i>n-1$ case is identical after reflection about $n-1$.

Lemma 4.1. *Let $i \in [0, 2n-2]$ be given. Then*

$$\mathbb{P}(i \notin A+A \mid |A| = r) \;=\; \frac{\text{\# ways to place } n-r \text{ vertices on disjoint edges to get cover}}{\text{\# ways to choose } n-r \text{ vertices from } n}$$
$$= \begin{cases} \dfrac{\displaystyle\sum_{k=\frac{i+1}{2}}^{i+1} 2^{i+1-k}\binom{\frac{i+1}{2}}{k-\frac{i+1}{2}}\binom{n-i-1}{n-r-k}}{\dbinom{n}{n-r}} & \text{for } i \text{ odd} \\[2ex] \dfrac{\displaystyle\sum_{k=\frac{i}{2}}^{i} 2^{i-k}\binom{\frac{i}{2}}{k-\frac{i}{2}}\binom{n-i-1}{n-r-1-k}}{\dbinom{n}{n-r}} & \text{for } i \text{ even.} \end{cases}$$

Proof. The derivation for the odd cases is as follows; the even cases can be handled similarly. We divide G_n into two components; G_0 containing the $\frac{i+1}{2}$ disjoint edges and G_1 containing $n-i-1$ isolated vertices. To count the number of vertex covers, we denote by k the number of our $n-r$ missing vertices which are placed in G_0. First, having fixed k, we must choose $n-r-k$ vertices from the $n-i-1$ vertices in G_1, with no edge restrictions. Second, inside of G_0 we must determine which edges are twice-covered and which edges are once-covered; a factor of $\binom{\frac{i+1}{2}}{k-\frac{i+1}{2}}$. Finally, those edges which are once-covered may be covered both on the right and on the left; a factor of 2^{i+1-k}. □

Then, if in (4.1) we use the symmetry around $n-1$ to double terms and account for $n-1 < i < 2n-2$, we find that

$$\mathbb{E}[|A+A|] \;=\; \sum_{r=0}^{n}\binom{n}{r}p^r q^{n-r}\left(2\sum_{i=0}^{n-2}(1-\mathbb{P}(i \notin A+A \mid |A|=r)) + (1-\mathbb{P}(n-1 \notin A+A \mid |A|=r))\right). \tag{4.2}$$

This proves Theorem 1.1, because equation (4.2) is exactly the claim. □

While this closed form is exact and easily approximated numerically, there are $O(n^3)$ sums to execute. We wish to place effective upper and lower bounds on this sum. First notice that

$$\begin{cases} \displaystyle\sum_{k=\frac{i+1}{2}}^{i+1} 2^{i+1-k}\binom{\frac{i+1}{2}}{k-\frac{i+1}{2}}\binom{n-i-1}{n-r-k} & \text{for } i \text{ odd} \\ \displaystyle\sum_{k=\frac{i}{2}}^{i} 2^{i-k}\binom{\frac{i}{2}}{k-\frac{i}{2}}\binom{n-i-1}{n-r-1-k} & \text{for } i \text{ even} \end{cases} \geq \begin{cases} \dbinom{n-\frac{i+1}{2}}{n-r-\frac{i+1}{2}} & \text{for } i \text{ odd} \\ \dbinom{n-\frac{i}{2}-1}{n-r-1-\frac{i}{2}} & \text{for } i \text{ even.} \end{cases} \tag{4.3}$$

As discussed before, the left-hand side counts the number of vertex covers using r vertices on our graph G. The right-hand side undercounts the number of such vertex covers, by first (in the odd case) assigning $\frac{i+1}{2}$ vertices to cover the edges, and then choosing the remaining $n-r-\frac{k+1}{2}$ vertices freely from the remaining $n-\frac{i+1}{2}$ vertices. Substituting this into (4.2), we find that

$$\begin{aligned}\mathbb{E}[|A+A|] \quad &\leq \quad \sum_{r=0}^{n} p^r q^{n-r}\Bigg(2\sum_{i=0}^{n-2}\left(\binom{n}{r}-\begin{cases}\binom{n-\frac{i+1}{2}}{r} & \text{for } i \text{ odd}\\ \binom{n-\frac{i}{2}-1}{r} & \text{for } i \text{ even}\end{cases}\right)\\ &\qquad + \binom{n}{r}-\begin{cases}\binom{n-\frac{n}{2}}{r} & \text{for } n-1 \text{ odd}\\ \binom{n-\frac{n-1}{2}-1}{r} & \text{for } n-1 \text{ even}\end{cases}\Bigg). \end{aligned} \tag{4.4}$$

We organize these by collecting those terms of the form $\binom{n}{r}$ to find a binomial which necessarily sums to 1. Specifically,

$$\begin{aligned}\mathbb{E}[|A+A|] \;\leq\; \sum_{r=0}^{n} p^r q^{n-r}\binom{n}{r}\left(2\left(\sum_{i=0}^{n-2}1\right)+1\right) \;&-\; \sum_{r=0}^{n} p^r q^{n-r} 2\sum_{i=0}^{n-2}\begin{cases}\binom{n-\frac{i+1}{2}}{r} & \text{for } i \text{ odd}\\ \binom{n-\frac{i}{2}-1}{r} & \text{for } i \text{ even}\end{cases}\\ &-\; \sum_{r=0}^{n} p^r q^{n-r}\begin{cases}\binom{n-\frac{n}{2}}{r} & \text{for } n-1 \text{ odd}\\ \binom{n-\frac{n-1}{2}-1}{r} & \text{for } n-1 \text{ even.}\end{cases}\end{aligned} \tag{4.5}$$

The first sum over r we see is binomial in r, and for a fixed value of i gives us

$$\sum_{r=0}^{n} p^r q^{n-r}\begin{cases}\binom{n-\frac{i+1}{2}}{r} & \text{for } i \text{ odd}\\ \binom{n-\frac{i}{2}-1}{r} & \text{for } i \text{ even}\end{cases} \tag{4.6}$$

by factoring $q^{\frac{i+1}{2}}$ or $q^{\frac{i}{2}}$ out of this sum we get, once again, a sum of probabilities of events under a binomial distribution which must sum to 1. We omit the last term corresponding to $n-1$, as we seek an upper bound. Then

$$\begin{aligned}\mathbb{E}[|A+A|] \;&\leq\; 2\sum_{i=0}^{n-2}1+1-\sum_{i=0}^{n-2}\begin{cases}q^{\frac{i+1}{2}} & \text{for } i \text{ odd}\\ q^{\frac{i+2}{2}} & \text{for } i \text{ even}\end{cases}\\ &=\; 2n-1-2q\sum_{i=0}^{n-2}(\sqrt{q})^i\\ &=\; 2n-1-2q\frac{1-q^{\frac{n-1}{2}}}{1-\sqrt{q}}\end{aligned} \tag{4.7}$$

as needed to prove the first statement of Theorem 1.2.

To derive a lower bound, we first see that

$$\begin{cases}\sum_{k=\frac{i+1}{2}}^{i+1} 2^{i+1-k}\binom{\frac{i+1}{2}}{k-\frac{i+1}{2}}\binom{n-i-1}{n-r-k} & \text{for } i \text{ odd}\\ \sum_{k=\frac{i}{2}}^{i} 2^{i-k}\binom{\frac{i}{2}}{k-\frac{i}{2}}\binom{n-i-1}{n-r-1-k} & \text{for } i \text{ even}\end{cases} \leq \begin{cases}2^{\frac{i+1}{2}}\binom{n-\frac{i+1}{2}}{n-r-\frac{i+1}{2}} & \text{for } i \text{ odd}\\ 2^{\frac{i}{2}}\binom{n-\frac{i}{2}-1}{n-r-1-\frac{i}{2}} & \text{for } i \text{ even.}\end{cases} \tag{4.8}$$

Similar to before, on the right-hand side we are counting each way to choose our isolated vertices, and overcounting the number of ways to position the vertices adjacent to edges.

Then by substitution,

$$\begin{aligned}\mathbb{E}[|A+A|] \;\geq\; \sum_{r=0}^{n} p^r q^{n-r} 2\sum_{i=0}^{n-2}\Bigg(\binom{n}{r}&-\begin{cases}2^{\frac{i+1}{2}}\binom{n-\frac{i+1}{2}}{r} & \text{for } i \text{ odd}\\ 2^{\frac{i}{2}}\binom{n-\frac{i}{2}-1}{r} & \text{for } i \text{ even}\end{cases}\\ &+\binom{n}{r}-\begin{cases}2^{\frac{n}{2}}\binom{n-\frac{n}{2}}{r} & \text{for } n-1 \text{ odd}\\ 2^{\frac{n-1}{2}}\binom{n-\frac{n-1}{2}-1}{r} & \text{for } n-1 \text{ even}\end{cases}\Bigg).\end{aligned} \tag{4.9}$$

If we distribute this sum into individual components,

$$\mathbb{E}[|A+A|] \;\geq\; \sum_{r=0}^{n} p^r q^{n-r}\left(2\left(\sum_{i=0}^{n-2}\binom{n}{r}\right) + 2\sum_{i=0}^{n-2}\left(-\begin{cases} 2^{\frac{i+1}{2}}\binom{n-\frac{i+1}{2}}{r} & \text{for } i \text{ odd}\\ 2^{\frac{i}{2}}\binom{n-\frac{i}{2}-1}{r} & \text{for } i \text{ even}\end{cases}\right)\right.$$
$$\left. + \left(\binom{n}{r} - \begin{cases} 2^{\frac{n}{2}}\binom{n-\frac{n}{2}}{r} & \text{for } n-1 \text{ odd}\\ 2^{\frac{n-1}{2}}\binom{n-\frac{n-1}{2}-1}{r} & \text{for } n-1 \text{ even}\end{cases}\right)\right), \quad (4.10)$$

before exchanging the order of summation,

$$\mathbb{E}[|A+A|] \;\geq\; \sum_{i=0}^{n-2}\left(2\sum_{r=0}^{n}\left(p^r q^{n-r}\binom{n}{r}\right) + 2\sum_{r=0}^{n}\left(-p^r q^{n-r}\begin{cases} 2^{\frac{i+1}{2}}\binom{n-\frac{i+1}{2}}{r} & \text{for } i \text{ odd}\\ 2^{\frac{i}{2}}\binom{n-\frac{i}{2}-1}{r} & \text{for } i \text{ even}\end{cases}\right)\right)$$
$$+ \sum_{r=0}^{n} p^r q^{n-r}\left(\binom{n}{r} - \begin{cases} 2^{\frac{n}{2}}\binom{n-\frac{n}{2}}{r} & \text{for } n-1 \text{ odd}\\ 2^{\frac{n-1}{2}}\binom{n-\frac{n-1}{2}-1}{r} & \text{for } n-1 \text{ even}\end{cases}\right). \quad (4.11)$$

Now the first sum over r is 1, a binomial sum. That is,

$$\mathbb{E}[|A+A|] \;\geq\; \sum_{i=0}^{n-2}\left(2 - 2\sum_{r=0}^{n}\left(p^r q^{n-r}\begin{cases} 2^{\frac{i+1}{2}}\binom{n-\frac{i+1}{2}}{r} & \text{for } i \text{ odd}\\ 2^{\frac{i}{2}}\binom{n-\frac{i}{2}-1}{r} & \text{for } i \text{ even}\end{cases}\right)\right)$$
$$+ \sum_{r=0}^{n} p^r q^{n-r}\left(\binom{n}{r} - \begin{cases} 2^{\frac{n}{2}}\binom{n-\frac{n}{2}}{r} & \text{for } n-1 \text{ odd}\\ 2^{\frac{n-1}{2}}\binom{n-\frac{n-1}{2}-1}{r} & \text{for } n-1 \text{ even}\end{cases}\right). \quad (4.12)$$

Let us consider, for a moment, the remaining terms inside. They depend on both r and i, but, for a fixed value of i, they resemble a binomial sum. That is, for fixed i,

$$\sum_{r=0}^{n} p^r q^{n-r}\begin{cases} 2^{\frac{i+1}{2}}\binom{n-\frac{i+1}{2}}{r} & \text{for } i \text{ odd}\\ 2^{\frac{i}{2}}\binom{n-\frac{i}{2}-1}{r} & \text{for } i \text{ even}\end{cases} = \begin{cases} (2q)^{\frac{i+1}{2}} & \text{for } i \text{ odd}\\ (2q)^{\frac{i+2}{2}} & \text{for } i \text{ even.}\end{cases}$$

Thus we have

$$\mathbb{E}[|A+A|] \;\geq$$
$$\sum_{i=0}^{n-2}\left(2 - 2\begin{cases} (2q)^{\frac{i+1}{2}} & \text{for } i \text{ odd}\\ (2q)^{\frac{i+2}{2}} & \text{for } i \text{ even.}\end{cases}\right) + \sum_{r=0}^{n} p^r q^{n-r}\left(\binom{n}{r} - \begin{cases} 2^{\frac{n}{2}}\binom{n-\frac{n}{2}}{r} & \text{for } n-1 \text{ odd}\\ 2^{\frac{n-1}{2}}\binom{n-\frac{n-1}{2}-1}{r} & \text{for } n-1 \text{ even}\end{cases}\right).$$

Now consider those terms independent of i. The first, $\binom{n}{r}$, is once again a simple binomial sum. The last term corresponding to $n-1$ may be handled the same way by factoring out $(2q)^{\frac{n}{2}}$ or $(2q)^{\frac{n-1}{2}}$, depending on parity; for a lower bound we choose to subtract the larger $(2q)^{\frac{n-1}{2}}$, and find that, using our assumption that $p > 1/2$ implies $q < 1/2$, so we may apply the geometric series formulae to obtain

$$\begin{aligned}\mathbb{E}[|A+A|] \;&\geq\; 2\sum_{i=0}^{n-2}\left(1 - \begin{cases} (2q)^{\frac{i+1}{2}} & \text{for } i \text{ odd}\\ (2q)^{\frac{i+2}{2}} & \text{for } i \text{ even}\end{cases}\right) + 1 - (2q)^{\frac{n-1}{2}}\\ &=\; 2n - 1 - 2q\sum_{i=0}^{n-2}(\sqrt{2q})^i - (2q)^{\frac{n-1}{2}}\\ &=\; 2n - 1 - \frac{2q}{1-\sqrt{2q}} - (2q)^{\frac{n-1}{2}}, \end{aligned} \quad (4.13)$$

completing the proof of Theorem 1.2. $\square$

5. VARIANCE

We now find the variance. Recall

$$\mathrm{Var}(|A+A|) \;=\; \mathbb{E}[|A+A|^2] - \mathbb{E}[|A+A|]^2. \tag{5.1}$$

In the previous section, we computed $\mathbb{E}[|A+A|]$, so we need only to determine $\mathbb{E}[|A+A|^2]$. We apply the same technique used to compute the expected value and condition each probability on the size of A.

$$\begin{aligned}
\mathbb{E}[|A+A|^2] \;&=\; \sum_{A\subseteq\{0,\dots,n-1\}} |A+A|^2\cdot\mathbb{P}(A)\\
&=\; \sum_{r=0}^{n}\binom{n}{r}p^r q^{n-r}\sum_{A\subseteq[0,n-1]\,,\,|A|=r}|A+A|^2\\
&=\; \sum_{r=0}^{n}\binom{n}{r}p^r q^{n-r}\sum_{0\le i,j\le 2n-2}\;\sum_{\substack{A\subseteq[0,n-1]\,,\,|A|=r\\ i,j\in A+A}} 1\\
&=\; \sum_{r=0}^{n}\binom{n}{r}p^r q^{n-r}\sum_{0\le i,j\le 2n-2}\mathbb{P}(i \text{ and } j\in A+A\mid |A|=r)\\
&=\; \sum_{r=0}^{n}\binom{n}{r}p^r q^{n-r}\left(2\sum_{0\le i<j\le 2n-2}\mathbb{P}(i\text{ and } j\in A+A\mid |A|=r)+\sum_{0\le i\le 2n-2}\mathbb{P}(i\in A+A\mid |A|=r)\right).
\end{aligned}\tag{5.2}$$

Similarly to the Expected Value, we compute

$$\mathbb{P}(i \text{ and } j\in A+A\mid |A|=r) \;=\; 1-\mathbb{P}(i\text{ and } j\notin A+A\mid |A|=r).$$

Once again, this reduces to a question about graph coverings. In Proposition 3.4, we state formulas for the number, and size, of paths in the dependency graph G associated to i and j. We choose $n-r$ elements to be missing from A, and seek to compute the number of vertex covers.

Unlike the dependency graph used for expected value, for a single i here we have many options for distributing our $n-r$ chosen vertices. We attack this program in generality, and derive a solution which can then take, as input, the number and size of paths we know are in G_n. Suppose we wish to compute the number of vertex covers on a graph consisting of m paths, each of length ℓ_i for $1\le i\le m$, with the remaining $n-\sum_{i=1}^m \ell_i$ vertices isolated. Then, given the number of vertices distributed to each path, we may compute the number of afforded vertex covers. Summing over all such possible distribution schemes, we find the total number of vertex covers. We state a lemma that will be important in computing $\mathbb{P}(i$ and $j\notin A+A\mid |A|=r)$.

Lemma 5.1. *Given a graph G consisting of n vertices with m disjoint paths, with lengths ℓ_i for $1\le i\le m$, and $t=n-\sum_{i=1}^m\ell_i$ isolated vertices, then the number of vertex covers of G using exactly $n-r$ vertices is equal to*

$$\sum_{\substack{r_0,r_1,\dots,r_m\in\mathbb{N}_0\\ r_0+r_1+\cdots+r_m=n-r}}\binom{t}{r_0}\prod_{i=1}^{m}\binom{r_i+1}{\ell_i-r_i}.$$

Proof. We must distribute the $n-r$ vertices amongst the pieces of our graph. This is exactly the internal sum. The $\binom{t}{r_0}$ term controls how many ways we may place those vertices in the edge-less block. Each of the $f(r_i,\ell_i)=\binom{r_i+1}{\ell_i-r_i}$ terms controls how many ways we may place the r_i vertices in that path. Note that if $r_i>\ell_i$ or, conversely, $r_i<\ell_i/2$, then f is zero since we cannot place that many vertices there, or get a vertex cover, respectively. The same occurs if we attempt to place more than t vertices in the edge-less block. $\square$

We want to use Lemma 5.1 to compute $\mathbb{P}(i \text{ and } j \notin A+A \mid |A| = r)$, in conjunction with Proposition 3.4, plugging in the lengths and number of these paths. We find the following proposition.

Proposition 5.2. *Let $i, j \in [0, 2n-2]$ with $i < j$. Then, denoting $P_r(i,j) = \mathbb{P}(i \text{ and } j \notin A+A \mid |A| = r)$ for i, j both odd:*

$$P_r(i,j) \;=\; \sum_{\substack{r_0, r_1, \ldots, r_m \in \mathbb{N}_0 \\ r_0 + r_1 + \ldots + r_m = n-r}} \binom{t}{r_0} \prod_{i=1}^{s} \binom{r_i+1}{q-r_i} \prod_{i=s+1}^{s+s'} \binom{r_i+1}{q+2-r_i},$$

where s, s' and q are as defined in Proposition 3.4, $m = s+s'$ and $t = n - (qs + (q+2)s')$.

For i even and j odd:

$$P_r(i,j) \;=\; \sum_{\substack{r_0, r_1, \ldots, r_m \in \mathbb{N}_0 \\ r_0 + r_1 + \ldots + r_m = n-r}} \binom{t}{r_0} \binom{r_m+1}{o-r_m} \prod_{i=1}^{s} \binom{r_i+1}{q-r_i} \prod_{i=s+1}^{s+s'} \binom{r_i+1}{q+2-r_i},$$

where s, s', o and q are as defined in Proposition 3.4, $m = s+s'+1$ and $t = n - (qs + (q+2)s' + o)$.

For i odd and j even:

$$P_r(i,j) \;=\; \sum_{\substack{r_0, r_1, \ldots, r_m \in \mathbb{N}_0 \\ r_0 + r_1 + \ldots + r_m = n-r}} \binom{t}{r_0} \binom{r_m+1}{o'-r_m} \prod_{i=1}^{s} \binom{r_i+1}{q-r_i} \prod_{i=s+1}^{s+s'} \binom{r_i+1}{q+2-r_i},$$

where s, s', o' and q are as defined in Proposition 3.4, $m = s+s'+1$ and $t = n - (qs + (q+2)s' + o')$.

For i, j both even:

$$P_r(i,j) \;=\; \sum_{\substack{r_0, r_1, \ldots, r_m \in \mathbb{N}_0 \\ r_0 + r_1 + \ldots + r_m = n-r}} \binom{t}{r_0} \binom{r_{m-1}+1}{o-r_{m-1}} \binom{r_m+1}{o'-r_m} \prod_{i=1}^{s} \binom{r_i+1}{q-r_i} \prod_{i=s+1}^{s+s'} \binom{r_i+1}{q+2-r_i},$$

where s, s', o, o' and q are as defined in Proposition 3.4, $m = s+s'+2$ and $t = n-(qs+(q+2)s'+o+o')$.

We find that

$$\mathbb{E}[|A+A|^2] \;=\; \sum_{r=0}^{n} \binom{n}{r} p^r q^{n-r} \Big(2 \sum_{0 \le i < j \le 2n-2} 1 - P_r(i,j) + \sum_{0 \le i \le 2n-2} 1 - P_r(i) \Big) \tag{5.3}$$

where $P(i,j) = \mathbb{P}(i \text{ and } j \notin A+A \mid |A| = r)$ and $P(i) = \mathbb{P}(i \notin A+A \mid |A| = r)$. We calculated $P_r(i,j)$ in Proposition 5.2, and $P_r(i)$ in Lemma 4.1. Since we have already calculated $\mathbb{E}[|A+A|]$ with Theorem 1.1, we have

$$\mathrm{Var}(|A+A|) \;=\; \sum_{r=0}^{n} \binom{n}{r} p^r q^{n-r} \Big(2 \sum_{0 \le i < j \le 2n-2} 1 - P_r(i,j) + \sum_{0 \le i \le 2n-2} 1 - P_r(i) \Big) - \mathbb{E}[|A+A|]^2, \tag{5.4}$$

which proves Theorem 1.3. $\square$

6. Divot Computations

In this section we prove Theorem 1.5. Fringe analysis has historically been the most successful technique for estimating probabilities of missing certain numbers of sums, and this is the method we follow. The technique grows out of the observation that sumsets usually have fully populated centers: there is a very low probability that an element from the bulk center of $[0, 2n-2]$ is missing.[5] When a suitable distance from the edge is chosen, this observation can be made precise. It follows that the number of missing sums is essentially controlled by the upper and lower fringes of the randomly chosen set. As long as they are short relative to the length of the whole set, the fringe behaviors at the top and bottom are independent and can be analyzed separately from the rest of the elements and each other. Furthermore, as long as they are reasonably sized (on the order of 20 to 30 elements) a computer can exhaustively check all the possible fringe arrangements, and give exact data for the number of missing sums.

Our approach is to represent a general set A as the union of a left, middle, and right part, where the left and right parts have fixed length ℓ and the middle size $n-2\ell$. Then, we establish sharp upper and lower bounds for $m_p(k)$, and use this to prove the existence of divots. First we develop some specialized notation for dealing with these fringe sets.

Fix a positive integer $\ell \leq n/2$; this will be the "fringe width". Write $A = L \cup M \cup R$, where $L \subseteq [0, \ell-1]$, $M \subseteq [\ell, n-\ell-1]$ and $R \subseteq [n-\ell, n-1]$. We look at $m_p(k) = \lim_{n\to\infty} m_{n\,;\,p}(k)$ for $k \in \mathbb{N}_0$, the limiting distribution of missing sums.

Let L_k be the event that $L+L$ misses exactly k elements in $[0, \ell-1]$. Let L_k^a be the event that $L+L$ misses exactly k elements in $[0, \ell-1]$ and contains $[\ell, 2\ell-a]$. Similar notations are applied to R; see below.

$$\begin{aligned}
L_k:&\quad |[0,\ell-1] \setminus (L+L)| \;=\; k,\\
L_k^a:&\quad |[0,\ell-1] \setminus (L+L)| \;=\; k \text{ and } [\ell, 2\ell-a] \subseteq L+L,\\
R_k:&\quad |[2n-\ell-1, 2n-2] \setminus (R+R)| \;=\; k,\\
R_k^a:&\quad |[2n-\ell-1, 2n-2] \setminus (R+R)| \;=\; k \text{ and } [2n-2\ell+a-2, 2n-\ell-2] \subseteq R+R.
\end{aligned} \tag{6.1}$$

Next, let $\min L_k$ be the minimal size of L for which the event L_k occurs, and similarly for the other events just defined; see below.

$$\begin{aligned}
\min L_k &= \min\{|L| : L_k \text{ occurs}\},\\
\min R_k &= \min\{|R| : R_k \text{ occurs}\},\\
\min L_k^a &= \min\{|L| : L_k^a \text{ occurs}\},\\
\min R_k^a &= \min\{|R| : R_k^a \text{ occurs}\}.
\end{aligned} \tag{6.2}$$

Let

$$\begin{aligned}
\mathcal{M}_{L,k} &= \{L \subset [0,\ell-1] \mid L_k \text{ occurs}\}\\
\mathcal{M}_{L,a,k} &= \{L \subset [0,\ell-1] \mid L_k^a \text{ occurs}\}\\
\mathcal{M}_{R,k} &= \{R \subset [0,\ell-1] \mid R_k \text{ occurs}\}\\
\mathcal{M}_{R,a,k} &= \{R \subset [0,\ell-1] \mid R_k^a \text{ occurs}\}
\end{aligned} \tag{6.3}$$

and

$$\begin{aligned}
\tau(L_k^a) &= \min_{L\in\mathcal{M}_{L,k}} |L \cap [0, \ell-a+1]|\\
\tau(R_k^a) &= \min_{R\in\mathcal{M}_{R,k}} |R \cap [n-\ell+a-2, n-1]|.
\end{aligned} \tag{6.4}$$

[5]If each element of $[0, n-1]$ is chosen with probability p, the number of elements in A is of size pn with fluctuations of order $\sqrt{n}$. There are thus of order p^2n^2 pairs of sums but only $2n-1$ possible sums, and most possible sums are realized.

By symmetry, for each $k \in \mathbb{N}_0$ we have $\min L_k = \min R_k$, $\min L_k^a = \min R_k^a$ and $\tau\big(L_k^a\big) = \tau\big(R_k^a\big)$. However, for clarity we will still distinguish these numbers despite that they are equal.

6.1. **An Upper Bound on $m_p(k)$.** In this section we place an upper bound on $m_p(k)$. First we show formally that our fringe events are independent.

Lemma 6.1 (Independence of the Fringes)**.** *Pick a fringe width ℓ. For any $k_1, k_2 \in [0, \ell]$, the events L_{k_1} and R_{k_2} are independent.*

Proof. The elements of $[0, n-1]$ are all chosen to be in A or not independently. The only elements of $[0, n-1]$ which can contribute to $(A+A) \cap [0, \ell-1]$ are those in the interval $[0, \ell-1]$, since any larger element will sum to at least $\ell + 0 \notin [0, \ell-1]$. Similarly, the only elements of $[0, n-1]$ which can contribute to $(A+A) \cap [2n-\ell-1, 2n-2]$ are those in the interval $[n-\ell, n-1]$, since any smaller element will sum to at most $(n-\ell-1)+(n-1) \notin [2n-\ell-1, 2n-2]$. Since $\ell \leq n/2$,

$$[0, \ell-1] \cap [n-\ell, n-1] \;=\; \varnothing. \tag{6.5}$$

Therefore the elements of $[0, \ell-1] \cap (L+L)$ and $[2n-\ell-1, 2n-2] \cap (R+R)$ are independent, so in particular the events L_{k_1} and R_{k_2} are independent. □

Next, we place an upper bound on the probability that the bulk of $A+A$ is missing at least one element.

Lemma 6.2. *Let C denote the event that $(([0, n-1] + [0, n-1]) \setminus A) \cap [\ell, 2n-\ell-2] \neq \varnothing$. Then*

$$\mathbb{P}(C) \;\leq\; 2 \begin{cases} \dfrac{(1+q)(2q-q^2)^{j+1}}{(1-q)^2}, & \ell = 2j+1; \\[2ex] \dfrac{(3q-q^2)(2q-q^2)^{j}}{(1-q)^2}, & \ell = 2j. \end{cases} \tag{6.6}$$

Proof. Because the event C implies that $[\ell, 2n-\ell-2] \not\subseteq A+A$,

$$\begin{aligned} \mathbb{P}(C) \;&\leq\; \mathbb{P}([\ell, 2n-\ell-2] \not\subseteq A+A) \\ &\leq \sum_{i=\ell}^{2n-\ell-2} \mathbb{P}(i \notin A+A) \\ &= \sum_{i=\ell}^{n-1} \mathbb{P}(i \notin A+A) + \sum_{i=n}^{2n-\ell-2} \mathbb{P}(i \notin A+A) \\ &= \sum_{\substack{i=\ell \\ i \text{ odd}}}^{n-1} (2q-q^2)^{(i+1)/2} + \sum_{\substack{i=\ell \\ i \text{ even}}}^{n-1} q(2q-q^2)^{i/2} \\ &\quad + \sum_{\substack{i=n \\ i \text{ odd}}}^{2n-\ell-2} (2q-q^2)^{n-(i+1)/2} + \sum_{\substack{i=n \\ i \text{ even}}}^{2n-\ell-2} q(2q-q^2)^{n-1-i/2}. \end{aligned} \tag{6.7}$$

The last equality uses Lemma 2.3. Each of the four sums on the RHS of inequality (6.7) can be bounded from above by an infinite geometric sum as follows:

$$\begin{aligned}
\sum_{\substack{i=\ell\\ i\text{ odd}}}^{n-1}(2q-q^2)^{(i+1)/2} &\le \sum_{\substack{i=\ell\\ i\text{ odd}}}^{\infty}(2q-q^2)^{(i+1)/2} &&= \begin{cases}\frac{(2q-q^2)^{j+1}}{(1-q)^2}, & \ell=2j+1\\ \frac{(2q-q^2)^{j+1}}{(1-q)^2}, & \ell=2j\end{cases}\\
\sum_{\substack{i=\ell\\ i\text{ even}}}^{n-1} q(2q-q^2)^{i/2} &\le \sum_{\substack{i=\ell\\ i\text{ even}}}^{\infty} q(2q-q^2)^{i/2} &&= \begin{cases}\frac{q(2q-q^2)^{j+1}}{(1-q)^2}, & \ell=2j+1\\ \frac{q(2q-q^2)^{j}}{(1-q)^2}, & \ell=2j\end{cases}\\
\sum_{\substack{i=n\\ i\text{ odd}}}^{2n-\ell-2}(2q-q^2)^{n-(i+1)/2} &\le \sum_{\substack{i=-\infty\\ i\text{ odd}}}^{2n-\ell-2}(2q-q^2)^{n-(i+1)/2} &&= \begin{cases}\frac{(2q-q^2)^{j+1}}{(1-q)^2}, & \ell=2j+1\\ \frac{(2q-q^2)^{j+1}}{(1-q)^2}, & \ell=2j\end{cases}\\
\sum_{\substack{i=n\\ i\text{ even}}}^{2n-\ell-2} q(2q-q^2)^{n-1-i/2} &\le \sum_{\substack{i=-\infty\\ i\text{ even}}}^{2n-\ell-2} q(2q-q^2)^{n-1-i/2} &&= \begin{cases}\frac{q(2q-q^2)^{j+1}}{(1-q)^2}, & \ell=2j+1\\ \frac{q(2q-q^2)^{j}}{(1-q)^2}, & \ell=2j.\end{cases}
\end{aligned} \tag{6.8}$$

Adding these together, we obtain the desired bound (inequality (6.6)). □

Remark 6.3. *In the first step of the above proof, we could replace*

$$\sum_{i=\ell}^{2n-\ell-2}\mathbb{P}(i\notin A+A) \tag{6.9}$$

with

$$\sum_{i=\ell;i\equiv l(2)}^{2n-\ell-2}\mathbb{P}(i,i+1\notin A+A), \tag{6.10}$$

and then use the results of Proposition 3.4 to place a tighter upper bound. However, these terms are already quite small and will play little role in our upper bound, so this would not significantly improve our result.

Using these lemmas, we prove an upper bound on the probability of missing exactly k elements.

Theorem 6.4. *Pick a fringe width ℓ. For any $k\in[0,\ell]$,*

$$m_{n\,;\,p}(k) \ \le\ \sum_{i=0}^{k}\mathbb{P}(L_i)\mathbb{P}(L_{k-i})+2\begin{cases}\frac{(1+q)(2q-q^2)^{j+1}}{(1-q)^2}, & \ell=2j+1\\[2ex] \frac{(3q-q^2)(2q-q^2)^{j}}{(1-q)^2}, & \ell=2j.\end{cases} \tag{6.11}$$

Proof. We divide the interval $[0,2n{-}2]$ into three subintervals: $[0,\ell{-}1]$, $[\ell,2n{-}\ell{-}2]$ and $[2n{-}\ell{-}1,2n{-}2]$. Suppose that there are k missing sums. We separate into two cases.

Case I. There are no missing sums in the interval $[\ell,2n-\ell-2]$. In this case, let i be the number of missing sums in $[0,\ell-1]$. (Note that i can be any integer between 0 and k inclusive, because we chose $k\le\ell$.) Then the remaining $k{-}i$ sums are in $[2n{-}\ell{-}2,2n{-}2]$, and thus the events L_i and R_{k-i} both occur.

Case II. There is at least one missing sum in $[\ell,2n-\ell-2]$. This corresponds to the event C defined in Lemma 6.6.

The above casework gives us the expression

$$m_{n\,;\,p}(k) \ =\ \sum_{i=0}^{k}\mathbb{P}(L_i\text{ and }R_{k-i})+\mathbb{P}(C). \tag{6.12}$$

By Lemma 6.1, L_i and R_{k-i} are independent, so

$$\mathbb{P}(L_i \text{ and } R_{k-i}) \;=\; \mathbb{P}(L_i)\mathbb{P}(R_{k-i}) \;=\; \mathbb{P}(L_i)\mathbb{P}(L_{k-i}). \tag{6.13}$$

Using this in (6.12), along with the bound on $\mathbb{P}(C)$ from Lemma 6.6, gives our desired bound (inequality (6.11)), completing the proof. □

Corollary 6.5. *Let $k \in \mathbb{N}_0$ and $p \in (0,1)$ be chosen. Given $\ell \geq k$, then*

$$m_p(k) \;\leq\; \sum_{i=0}^{k} \mathbb{P}(L_i)\mathbb{P}(L_{k-i}) + 2\begin{cases} \frac{(1+q)(2q-q^2)^{j+1}}{(1-q)^2}, & \ell = 2j+1 \\ \\ \frac{(3q-q^2)(2q-q^2)^{j}}{(1-q)^2}, & \ell = 2j. \end{cases} \tag{6.14}$$

Proof. This result follows immediately from Theorem 6.4 by taking the limit as n goes to infinity of both sides. In particular, $\lim_{n\to\infty} m_{n\,;\,p}(k) = m_p(k)$ while the right side is independent of n. □

6.2. **A Lower Bound on $m_p(k)$.** We now attack the more challenging problem of finding a lower bound for the number of missing sums. This will allow us to prove the existence of a divot at 1 by showing that the probability of missing nothing and the probability of missing two sums have lower bounds that are greater than the upper bound for missing one sum. Once again, we begin by observing that our fringe events are indeed independent.

Lemma 6.6 (Independence of the Fringes)**.** *Fix a fringe width ℓ and a positive integer $a \leq \ell$. If $n \geq 4\ell - 2a + 1$, then for any $k_1, k_2 \in [0, \ell]$, the events $L^a_{k_1}$ and $R^a_{k_2}$ are independent.*

The proof of Lemma 6.6 is similar to that of Lemma 6.1. The following lemma is a generalization of Proposition 8 in [MO]. The lemma gives a lower bound that is independent the specific elements of the fringe. Instead, the bound only involves the cardinalities of L and R.

Lemma 6.7. *Choose a fringe width ℓ and let $L \subseteq [0, \ell-1]$ and $R \subseteq [n-\ell, n-1]$ be fixed. Let $S = L \cup M \cup R$ for $M \subseteq [\ell, n-\ell-1]$. Then for any $\varepsilon > 0$,*

$$\mathbb{P}([2\ell-1, 2n-2\ell-1] \subseteq A+A) \;\geq\; 1 - \frac{1+q}{(1-q)^2}(q^{|L|} + q^{|R|}) - \varepsilon \tag{6.15}$$

for all sufficiently large n.

Proof. We have

$$\begin{aligned}
&\mathbb{P}([2\ell-1, 2n-2\ell-1] \subseteq A+A) \\
&= \mathbb{P}([2\ell-1, n-\ell-1] \cup [n+\ell-1, 2n-2\ell-1] \subseteq A+A \\
&\quad \text{and } [n-\ell, n+\ell-2] \subseteq A+A) \\
&= 1 - \mathbb{P}([2\ell-1, n-\ell-1] \cup [n+\ell-1, 2n-2\ell-1] \not\subseteq A+A \\
&\quad \text{or } [n-\ell, n+\ell-2] \not\subseteq A+A) \\
&\geq 1 - \mathbb{P}([2\ell-1, n-\ell-1] \cup [n+\ell-1, 2n-2\ell-1] \not\subseteq A+A) \\
&\quad - \mathbb{P}([n-\ell, n+\ell-2] \not\subseteq A+A).
\end{aligned} \tag{6.16}$$

We find a lower bound for $\mathbb{P}([n-\ell, n+\ell-2] \subseteq A+A)$. Since $M+M \subseteq A+A$,

$$\mathbb{P}([n-\ell, n+\ell-2] \subseteq A+A) \;\geq\; \mathbb{P}([n-\ell, n+\ell-2] \subseteq M+M). \tag{6.17}$$

Applying the change of variable $N = n - 2\ell$, we estimate

$$\begin{aligned}
\mathbb{P}([n-2\ell, n-2] \subseteq M+M) &= \mathbb{P}([N, N+2\ell-2] \subseteq M+M) \\
&= 1 - \mathbb{P}(\exists k \in [N, N+2\ell-2], k \notin M+M) \\
&\geq 1 - \sum_{k=N}^{N+2\ell-2} \mathbb{P}(k \notin M+M) \\
&= 1 - \sum_{\substack{k=N \\ k \text{ even}}}^{N+2\ell-2} \mathbb{P}(k \notin M+M) - \sum_{\substack{k=N \\ k \text{ odd}}}^{N+2\ell-2} \mathbb{P}(k \notin M+M) \\
&= 1 - \sum_{\substack{k=N \\ k \text{ even}}}^{N+2\ell-2} q(2q-q^2)^{N-1-k/2} - \sum_{\substack{k=N \\ k \text{ odd}}}^{N+2\ell-2} (2q-q^2)^{N-(k+1)/2}. \qquad (6.18)
\end{aligned}$$

The last equality uses Lemma 2.3. In the last line, the exponents of $2q - q^2 \in (0,1)$ are all at least $N/2 - \ell = n/2 - 2\ell$, so the RHS approaches 1 as $n \to \infty$. Hence for any $\varepsilon > 0$, when n is sufficiently large we have $\mathbb{P}([n-\ell, n+\ell-2] \subseteq A+A) \geq 1 - \varepsilon$ and so

$$\mathbb{P}([n-\ell, n+\ell-2] \not\subseteq A+A) = 1 - \mathbb{P}([n-\ell, n+\ell-2] \subseteq A+A) \leq \varepsilon$$

Combining this with Lemma 2.4, we obtain

$$\mathbb{P}([2\ell-1, 2n-2\ell-1] \subseteq A+A) \geq 1 - \frac{1+q}{(1-q)^2}(q^{|L|} + q^{|R|}) - \varepsilon. \qquad (6.19)$$

This completes our proof. □

The even L_i^a prescribes, in some ways, the behavior of $i+\ell-a$ elements in $[0, 2\ell]$; i sums must be missing from the first ℓ, while $[\ell, 2\ell - a]$ are all present. The next lemma places a lower bound on the probability that the remaining $a - 3$ elements are also present in $A + A$.

Lemma 6.8. *For $n \geq 4\ell - 2a + 1$, we have*

$$\begin{aligned}
\mathbb{P}([2\ell-a+1, 2\ell-2] \subseteq A+A \mid L_i^a) &\geq 1 - (a-2)q^{\tau(L_i^a)}, \\
\mathbb{P}([2n-2\ell, 2n-2\ell+a-3] \subseteq A+A \mid R_i^a) &\geq 1 - (a-2)q^{\tau(R_i^a)}.
\end{aligned} \qquad (6.20)$$

Proof. We prove only the first inequality because the second follows identically. We have

$$\begin{aligned}
\mathbb{P}([2\ell-a+1, 2\ell-2] \subseteq A+A \mid L_i^a) &= 1 - \mathbb{P}([2\ell-a+1, 2\ell-2] \not\subseteq A+A \mid L_i^a) \\
&\geq 1 - \sum_{k=2\ell-a+1}^{2\ell-2} \mathbb{P}(k \notin A+A \mid L_i^a). \qquad (6.21)
\end{aligned}$$

Recall the definitions of $\mathcal{M}_{L,a,i}$ (the set of sets $L \subset [0, \ell-1]$ such that event L_i^a occurs) and

$$\tau\big(L_k^a\big) = \min_{L \in \mathcal{M}_{L,i}} |L \cap [0, \ell-a+1]|. \qquad (6.22)$$

Suppose from now on that L_i^a occurs. For each $k \in [2\ell-a+1, 2\ell-2]$, the probability that $k \notin A+A$ is equal to the probability that for each $x \in L$, the corresponding $x - k \notin L$. Since there are at least $\tau(L_k^a)$ elements of L, and the probability of excluding a certain integer from S is q, we can bound

$$\mathbb{P}(k \notin A+A \mid L_i^a) \leq q^{\tau(R_i^a)}. \qquad (6.23)$$

Hence

$$1 - \sum_{k=2\ell-a+1}^{2\ell-2} \mathbb{P}(k \notin A+A \mid L_i^a) \geq 1 - (a-2)q^{\tau(R_i^a)}. \qquad (6.24)$$

This completes our proof. $\square$

Given $k \in \mathbb{N}_0$, the following theorem gives us a lower bound for $m_{n\,;\,p}(k)$.

Theorem 6.9. *Fix $q \in (0,1)$ and pick a fringe length $\ell \geq 0$. Also choose $a \leq \ell$. For any $\varepsilon > 0$, the following holds for all sufficiently large n:*

$$m_{n\,;\,p}(k) \;\geq\; \sum_{i=0}^{k} \mathbb{P}(L_i^a)\mathbb{P}(R_{k-i}^a)\theta_{k,i}(q,\varepsilon), \tag{6.25}$$

where

$$\theta_{k,i}(q,\varepsilon) \;=\; 1-(a-2)(q^{\tau(L_i^a)}+q^{\tau(R_{k-i}^a)})-\varepsilon-\frac{1+q}{(1-q)^2}(q^{\min L_i^a}+q^{\min R_{k-i}^a}). \tag{6.26}$$

Proof. The probability that $A+A$ is missing exactly k sums is greater than the probability that all these sums are missing from the two fringes. Thus for each $k \in [0,\ell]$, we have

$$\begin{aligned} m_{n\,;\,p}(k) \;&\geq\; \sum_{i=0}^{k} \mathbb{P}(L_i^a \text{ and } R_{k-i}^a \text{ and } [2\ell-a+1, 2n-2\ell+a-3] \subseteq A+A) \\ &=\; \mathbb{P}(L_i^a \text{ and } R_i^a)\mathbb{P}([2\ell-a+1,2n-2\ell+a-3] \subseteq A+A \mid L_i^a \text{ and } R_{k-i}^a) \\ &=\; \mathbb{P}(L_i^a)\mathbb{P}(R_i^a)\mathbb{P}([2\ell-a+1,2n-2\ell+a-3] \subseteq A+A \mid L_i^a \text{ and } R_{k-i}^a). \end{aligned} \tag{6.27}$$

This last equality follows from Lemma 6.6. We can bound $\mathbb{P}([2\ell-a+1,2n-2\ell+a-3] \subseteq A+A \mid L_i^a \text{ and } R_{k-i}^a)$ below by splitting into three subintervals.

$$\begin{aligned} &\mathbb{P}([2\ell-a+1,2\ell-2]\cup[2\ell-1,2n-2\ell-1]\cup[2n-2\ell,2n-2\ell+a-3] \subseteq A+A \mid L_i^a \text{ and } R_{k-i}^a) \\ &\quad=\; 1-\mathbb{P}([2\ell-a+1,2\ell-2] \not\subseteq A+A \text{ or } [2\ell-1,2n-2\ell-1] \not\subseteq A+A \\ &\qquad\qquad \text{or } [2n-2\ell,2n-2\ell+a-3] \not\subseteq A+A \mid L_i^a \text{ and } R_{k-i}^a) \\ &\quad\geq\; 1-\mathbb{P}([2\ell-a+1,2\ell-2] \not\subseteq A+A \mid L_i^a \text{ and } R_{k-i}^a) \\ &\qquad -\mathbb{P}([2\ell-1,2n-2\ell-1] \not\subseteq A+A \mid L_i^a \text{ and } R_{k-i}^a) \\ &\qquad -\mathbb{P}([2n-2\ell,2n-2\ell+a-3] \not\subseteq A+A \mid L_i^a \text{ and } R_{k-i}^a) \\ &\quad\geq\; 1-(a-2)q^{\tau(L_i^a)}-\frac{1+q}{(1-q)^2}(q^{\min L_i^a}+q^{\min R_{k-i}^a})-\varepsilon-(a-2)q^{\tau(R_{k-i}^a)} \end{aligned}$$

The last inequality uses Lemma 6.8, as well as Lemma 6.7 with the observation that for any L such that L_i^a occurs, $q^{|L|} \leq q^{\min L_i^a}$ (respectively $q^{|R|} \leq q^{\min R_i^a}$). Hence

$$\mathbb{P}(L_i^a \text{ and } R_{k-i}^a \text{ and } [2\ell-a+1,2n-2\ell+a-3] \subseteq A+A) \;\geq\; \mathbb{P}(L_i^a)\mathbb{P}(R_{k-i}^a)\theta_{k,i}(q,\varepsilon). \tag{6.28}$$

This completes our proof. $\square$

Corollary 6.10. *Let $k \in \mathbb{N}_0$ and $p \in (0,1)$ be chosen. Given $\ell \geq k$, then*

$$\begin{aligned} m_p(k) \;\geq\; \sum_{i=0}^{k} \mathbb{P}(L_i^a)\mathbb{P}(L_{k-i}^a)\Big[&1-(a-2)\left(q^{\tau(L_i^a)}+q^{\tau(L_{k-i}^a)}\right) \\ &-\frac{1+q}{(1-q)^2}\left(q^{\min L_i^a}+q^{\min L_{k-i}^a}\right)\Big]. \end{aligned} \tag{6.29}$$

Proof. This follows immediately from Theorem 6.9 by taking the limit as n goes to infinity of both sides. In particular, $\lim_{n\to\infty} m_{n\,;\,p}(k) = m_p(k)$ while the right side is independent of n. $\square$

These upper and lower bounds enable a proof of Theorem 1.5.

Remark 6.11. *Given $0 \le k \le \ell$, $\mathbb{P}(L_k)$ is a polynomial of p. Recall that $\mathcal{M}_{L,k}$ is the set of all sets L such that event L_k occurs. For $0 \le i \le \ell$, let*

$$c(i) \;=\; |\{L \in \mathcal{M}_{L,k} \;\text{ such that }\; |L| = i\}| \tag{6.30}$$

Then

$$\mathbb{P}(L_k) \;=\; \sum_{i=0}^{\ell} c(i)p^i(1-p)^{\ell-i}. \tag{6.31}$$

As long as ℓ, the fringe size, is not too large[6]*, we can numerically compute $c(i)$. Similarly, define*

$$c_{k,a}(i) \;=\; |\{L \in \mathcal{M}_{L,a,k} \;\text{ such that }\; |L| = i\}|. \tag{6.32}$$

Then

$$\mathbb{P}(L_k^a) = \sum_{i=0}^{\ell} c_{k,a}(i)p^i(1-p)^{\ell-i}. \tag{6.33}$$

So we can numerically compute the upper and lower bounds for $m_p(k)$ found in § 6.1 and § 6.2.

Finally, we employ these numerical techniques through exhaustive search to prove Theorem 1.5.

Proof of Theorem 1.5. For our argument for the divot at 1, we use $\ell = 30$ and $a = 12$. For clarity, we summarize our bounds:

$$\begin{aligned} m_p(0) &\;\ge\; \mathrm{LB}(0,p), \\ m_p(1) &\;\le\; \mathrm{UB}(1,p), \\ m_p(2) &\;\ge\; \mathrm{LB}(2,p), \end{aligned} \tag{6.34}$$

where

$$\begin{aligned} \mathrm{LB}(0,p) &:= \mathbb{P}(L_i^{12})\mathbb{P}(L_0^{12})\left[1 - 10(q^{\tau(L_0^{12})} + q^{\tau(L_0^{12})}) - \frac{1+q}{(1-q)^2}(q^{\min L_0^{12}} + q^{\min L_0^{12}})\right], \\ \mathrm{UB}(1,p) &:= \sum_{i=0}^{1} \mathbb{P}(L_i)\mathbb{P}(L_{1-i}) + 2\frac{(3q-q^2)(2q-q^2)^{15}}{(1-q)^2}, \\ \mathrm{LB}(2,p) &:= \mathbb{P}(L_i^{12})\mathbb{P}(L_{2-i}^{12}) \sum_{i=0}^{2} \Bigg[1 - 10(q^{\tau(L_i^{12})} + q^{\tau(L_{2-i}^{12})}) \\ &\qquad\qquad - \frac{1+q}{(1-q)^2}(q^{\min L_i^{12}} + q^{\min L_{2-i}^{12}})\Bigg]. \end{aligned} \tag{6.35}$$

Using Remark 6.11, we can plot each function $\mathrm{LB}(0,p), \mathrm{UB}(1,p)$ and $\mathrm{LB}(2,p)$ ($q = 1-p$). We provide the values for $\min L_i^{12}, \tau(L_i^{12}), c_k(i)$ and $c_{k,a}(i)$, which are crucial values for explicitly plotting functions $\mathrm{LB}(0,p), \mathrm{UP}(1,p), \mathrm{LB}(2,p)$ in Appendix C. Figure 6 is the plot. From it, we see that for $p \ge 0.68$, $\mathrm{LB}(0,p) > \mathrm{UB}(1,p) < \mathrm{LB}(2,p)$; thus, there is a divot at 1. Numerical evidence shows that our upper bounds are very good; we also discuss this in Appendix B. □

[6]Later, we pick $\ell = 30$ because computer can run through $2^{30} \approx 10^9$ subsets in a reasonable amount of time. The computer (no parallel running) took approximately 3 days to gather all the information we need. For $p \ge 0.67$, there is no need to go further since our results are already very close to the true value.

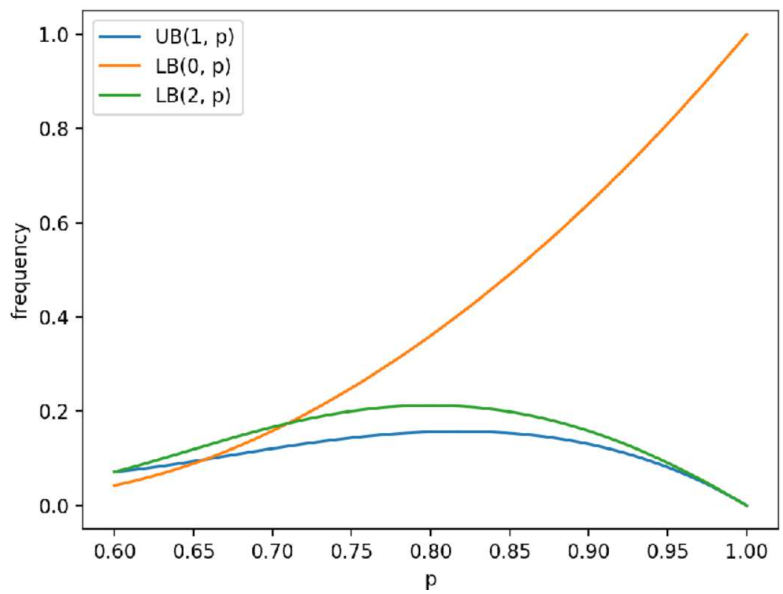

FIGURE 6. Plot of the bounds $\mathrm{LB}(0,p), \mathrm{UB}(1,p)$ and $\mathrm{LB}(2,p)$. Since $\mathrm{LB}(0,p) > \mathrm{UB}(1,p) < \mathrm{LB}(2,p)$ for $p \geq 0.68$, there exists a divot at 1 for $p \geq 0.68$.

7. CORRELATED SUMSETS

Up unto this point, we have studied the random variable $|A+A|$, where each element is included in A with probability p. Now, we examine the random variable $|A+B|$, where, for a given triplet (p, p_1, p_2) and any $i \in \{0, \dots, n-1\}$

- $\mathbb{P}(i \in A) = p$
- $\mathbb{P}(i \in B \mid i \in A) = p_1$
- $\mathbb{P}(i \in B \mid i \notin A) = p_2$.

For example, if $p_1 = 1, p_2 = 0$ we recover the problem of $|A+A|$, while if $p_1 = 0, p_2 = 1$ we get $|A+A^c|$.

Our first objective is, as before, to use graph theory to compute $\mathbb{P}(i, j \notin A+B)$. The probability of missing a single element, $\mathbb{P}(i \notin A+B)$, was computed in [DKMMW]. The clear choice of graph-theoretic generalization is to form a bipartite graph CG.

Definition 7.1. *For a set $F \subseteq [0, 2n-2]$ we define the bipartite* correlated condition graph $CG_n = (V, E)$ *induced on F where $V = A \cup B = \{0_A, 1_A, \dots, (n-1)_A, 0_B, 1_B, \dots, (n-1)_B\}$, and for two vertices $k_1 \in A$ and $k_2 \in B$, $(k_1, k_2) \in E$ if $k_1 + k_2 \in F$.*

Then, just as before with Lemma 3.2, the event $i, j \notin A+B$ is the same as having a vertex cover on this graph of those elements missing from A and B.

Fortunately, the structure of this graph is entirely analogous to that found in § 3. If $k \in \{0, \dots, n-1\}$, and we denote by k_A and k_B the copies of k potentially present in A, B respectively, then we know that if $k_A^1 + k_B^2 = i$, then also $k_B^1 + k_A^2 = i$, and so each edge in G_n has a "partner". Thus, [LMO]'s Proposition 3.1 still applies and we once again find ourself with a collection of disjoint paths, present in pairs where one element is in A and the other is in B.

Definition 7.2. *Given nonnegative integers i, j, an* accordion path *of length n on CG_m is a pair of paths in CG_m given by vertices specified by a sequence of integers k_s for $1 \leq s \leq n$, so that for each $s > 1$,*

$k_s + k_{s-1} \in \{i, j\}$. *Then the accordion path is given by* $\bigcup_{k_A, k_B \in k_s} \{k_A, k_B\}$ *and the edges (inherited from the condition graph) between them.*

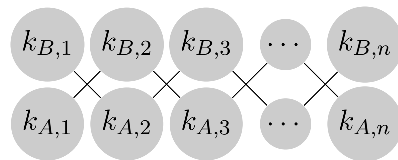

FIGURE 7. A generic accordion path.

An example of such an accordion path is given in Figure 7. In Section 3, we were able to compute the probability of finding a vertex cover on a path using a one-dimensional recurrence relation. Here, we still have that vertices in distinct pairs of paths are independent, but within a pair of paths composing a single accordion we have serious dependencies, since each element is included in B with probability conditioned on whether or not that element was included in A. However, by extending our recurrence relation to include 2 new variables, setting

$$\begin{aligned} x_n &= \mathbb{P}(\text{vertex cover}) \\ y_n &= \mathbb{P}(\text{vertex cover and } k_n \in S) \\ z_n &= \mathbb{P}(\text{vertex cover and } k_n^* \in S). \end{aligned} \tag{7.1}$$

We then get the recurrence relations

$$\begin{aligned} x_n &= qq_2 x_{n-1} + qp_2 y_{n-1} + pq_1 z_{n-1} + pp_1 qq_2 x_{n-2} \\ y_n &= qq_2 x_{n-1} + qp_2 y_{n-1} \\ z_n &= qq_2 x_{n-1} + pq_1 z_{n-1}. \end{aligned} \tag{7.2}$$

where $x_1 = y_1 = z_1 = 1$ and

$$\begin{aligned} x_2 &= q(q_2 + p_2 q) + p(q_1 q q_2 + q_1^2 p + p_1 q q_2) \\ y_2 &= q(1 - pp_2) \\ z_2 &= pq_1(qq_2 + pq_1) + qq_2. \end{aligned} \tag{7.3}$$

This larger recurrence relation generalizes the one previously derived in §3. To find asymptotics, we can examine the eigenvalues of the governing 4×4 matrix;

$$\begin{pmatrix} x_{n+1} \\ y_{n+1} \\ z_{n+1} \\ x_n \end{pmatrix} = \begin{pmatrix} qq_2 & qp_2 & pq_1 & pp_1 qq_2 \\ qq_2 & qp_2 & 0 & 0 \\ qq_2 & 0 & pq_1 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} x_n \\ y_n \\ z_n \\ x_{n-1} \end{pmatrix}. \tag{7.4}$$

In fact, we can find the closed form for the eigenvalues of the governing matrix from Equation (7.4); however, we do not include it in this paper as it is not very informative to the behavior of the probability of obtaining a vertex cover. Instead, we propose that one might fix one or more of p, p_1, and p_2, and then find the eigenvalues, to obtain a more meaningful result.

From this, we are able to find a preliminary result for the event $k \notin A + B$, noting the similarities to cases discussed in Section 4 and [MO].

Proposition 7.3. *For* $k \in [0, 2n - 2]$, *we have*

$$\mathbb{P}(k \notin A + B) \;=\; \begin{cases} x_2^{k/2}(1 - pp_1) & \text{if } k \text{ is even,} \\ x_2^{k+1/2} & \text{if } k \text{ is odd,} \end{cases}$$

where x_2 is as defined in Equation 7.3.

Proof. Consider CG_2 induced on $k \in [0, 2n-2]$. We notice this graph is very similar to the graph displayed in Figure 5, with disjoint edges and isolated vertices. We also note that Lemma 3.2 still applies, so we find a vertex cover on this graph.

By definition, the probability of obtaining a vertex cover on a disjoint edge is x_2. We must count how many of these disjoint edges there are; then we can multiply these together and find the probability of obtaining a vertex cover on the graph.

If k is odd, then there are $(k+1)/2$ disjoint edges. So, we get $x_2^{(k+1)/2}$.

If k is even, then there are $k/2$ disjoint edges, however, there is also an edge between $(k/2)_A$ and $(k/2)_B$ with no "partner". The probability of obtaining a vertex cover for this edge is $1 - pp_1$. So, we get $x_2^{k/2}(1 - pp_1)$. $\square$

And using the framework developed in Section 3, we are able to find the following Proposition.

Proposition 7.4. *For $i, j \in [0, 2n-2]$, we have*

$$\mathbb{P}(i, j \notin A + B) \;=\; \begin{cases} x_q^s\, x_{q+2}^{s'} & i, j \text{ both odd,} \\ x_o\, x_q^s\, x_{q+2}^{s'} & i \text{ even, } j \text{ odd,} \\ x_{o'}\, x_q^s\, x_{q+2}^{s'} & i \text{ odd, } j \text{ even,} \\ x_o\, x_{o'}\, x_q^s\, x_{q+2}^{s'} & i, j \text{ both even,} \end{cases} \tag{7.5}$$

where q, s, s', o, o' are as defined in Proposition 3.4.

Proof. Consider CG_n induced on $\{i, j\}$. To find $\mathbb{P}(i, j \notin A + B)$, we must find the probability of obtaining a vertex cover on this graph. Thankfully, the structure of this graph has been well-studied, from Proposition 3.4. The difference is we now have accordion paths as opposed to paths, however x_n gives us the probability of obtaining a vertex cover on an accordion path of length n. So, we can use Proposition 3.4 to find the number and lengths of these accordion paths, to obtain our desired result. $\square$

8. Future Work

We list some natural questions for future research. We first list questions relating to Sections 4, 5 and 7.

- Does there exist a "good" lower bound for $\mathbb{E}[|A + A|]$ for $p \le 1/2$?
- Can a "good" bound be found for $\mathrm{Var}(|A + A|)$?
- Does there exist a closed formula for $\mathbb{E}[|A + B|]$ and $\mathrm{Var}(|A + B|)$?

Now we list questions relating to Section 6. For convenience, we present Figure 2 again.

- As p decreases, the divot appears to shift to the right, from 1 at $p = .8$, to 3 for $p = .6$, to 7 for $p = .5$. How does the position of the divot depend on p? Do divots move monotonically with p?
- At $p = .7$ there appear to be two divots at 1 and 3; for what values of p are there more than one divot?
- Is there a value p_0 where for $p > p_0$ the distribution of the number of missing sums has a divot, and for $p < p_0$ the divot disappears. Where is this phase transition point p_0?
- In our theoretical and numerical investigations, we have never seen a divot at an even number. Are there no divots at even values?

These results all apply to the sumset $A+A$. In general, can any of these results be applied to the difference set $A - A$? What is needed to apply these results to the difference set?

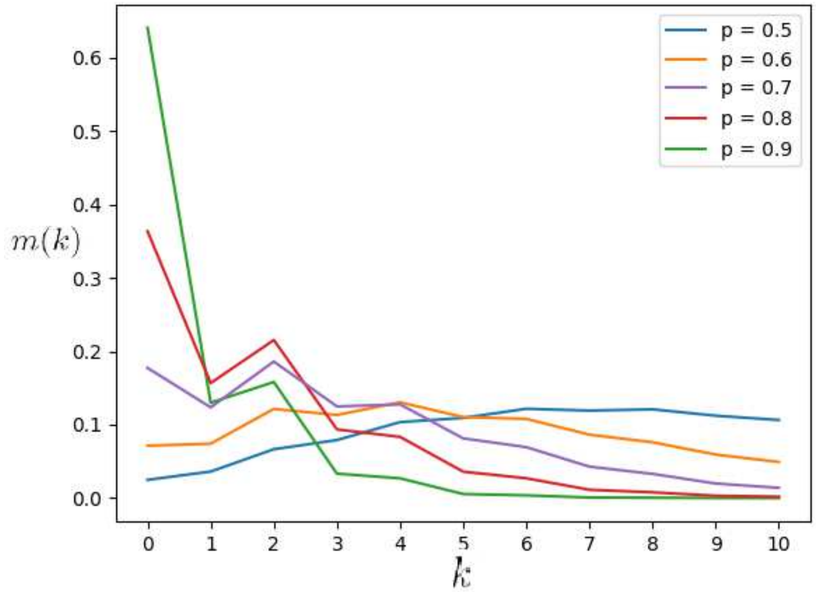

FIGURE 8. Plot of the distribution of missing sums, varying p by simulating 10^6 subsets of $\{0, 1, 2, \ldots, 400\}$. The simulation shows that: for $p = 0.9$ and 0.8, there is a divot at 1, for $p = 0.7$, there are divots at 1 and 3, for $p = 0.6$, there is a divot at 3 and for $p = 0.5$, there is a divot at 7.

APPENDIX A. PROOFS OF GENERALIZATIONS

Here we provide the full proofs of the many generalizations of lemmas originally proven by [MO]. Note that we have only introduced new notation that generalizes the previous arguments made. We also provide the full proof of Theorem 1.4, that is structurally equivalent to Theorem 1.2 of [LMO].

Proof of Lemma 2.1. Define random variables X_j by setting $X_j = 1$ if $j \in A$ and $X_j = 0$ otherwise. By the definition of A, the variables X_j are independent random variables for $\ell \le j \le n - u - 1$, each taking the values 0 and 1 with probability q and p respectively, while the variables X_j for $0 \le j \le \ell - 1$ and $n - u \le j \le n - 1$ have values that are fixed by the choices of L and U.

We have $k \notin A + A$ if and only if $X_j X_{k-j} = 0$ for all $0 \le j \le k/2$; the key point is that these variables $X_j X_{k-j}$ are independent of one another. Therefore

$$\mathbb{P}(k \notin A + A) = \prod_{0 \le j \le k/2} \mathbb{P}(X_j X_{k-j} = 0). \tag{A.1}$$

If k is odd, this becomes

$$\begin{aligned}
\mathbb{P}(k \notin A + A) &= \prod_{j=0}^{\ell-1} \mathbb{P}(X_j X_{k-j} = 0) \prod_{j=\ell}^{(k-1)/2} \mathbb{P}(X_j X_{k-j} = 0) \\
&= \prod_{j \in L} \mathbb{P}(X_{k-j} = 0) \prod_{j=\ell}^{(k-1)/2} \mathbb{P}(X_j = 0 \text{ or } X_{k-j} = 0) \\
&= q^{|L|}(1 - p^2)^{(k+1)/2 - \ell}.
\end{aligned} \tag{A.2}$$

On the other hand, if k is even then

$$\begin{aligned}
\mathbb{P}\left(k \notin A+A\right) &= \prod_{j=0}^{\ell-1}\mathbb{P}\left(X_j X_{k-j}=0\right)\left(\prod_{j=\ell}^{k/2-1}\mathbb{P}\left(X_j X_{k-j}=0\right)\right)\mathbb{P}\left(X_{k/2}X_{k/2}=0\right)\\
&= \prod_{j\in L}\mathbb{P}\left(X_{k-j}=0\right)\left(\prod_{j=\ell}^{k/2-1}\mathbb{P}\left(X_j=0 \text{ or } X_{k-j}=0\right)\right)\mathbb{P}\left(X_{k/2}=0\right)\\
&= q^{|L|}(1-p^2)^{k/2-\ell}\cdot q. \qquad \text{(A.3)}
\end{aligned}$$

□

Proof of Lemma 2.2. This follows from Lemma 2.1 applied to the parameters $\ell'=u$ and $L'=n-1-U$, $u'=\ell$ and $U'=n-1-L$, and $A'=n-1-A$ and $k'=2n-2-k$. □

Proof of Proposition 2.4. We employ the crude inequality

$$\begin{aligned}
&\mathbb{P}\left(\{2\ell-1,\ldots,n-u-1\}\cup\{n+\ell-1,\ldots,2n-2u-1\}\not\subseteq A+A\right)\\
&\leq \sum_{k=2\ell-1}^{n-u-1}\mathbb{P}\left(k\notin A+A\right)+\sum_{k=n+\ell-1}^{2n-2u-1}\mathbb{P}\left(k\notin A+A\right). \qquad \text{(A.4)}
\end{aligned}$$

The first sum can be bounded, using Lemma 2.1, by

$$\begin{aligned}
\sum_{k=2\ell-1}^{n-u-1}\mathbb{P}\left(k\notin A+A\right) &< \sum_{\substack{k\geq 2\ell-1\\ k \text{ odd}}} q^{|L|}(1-p^2)^{(k+1)/2-\ell}+\sum_{\substack{k\geq 2\ell-1\\ k \text{ even}}} q^{|L|+1}(1-p^2)^{k/2-\ell}\\
&= q^{|L|}\sum_{m=0}^{\infty}(1-p^2)^m+q^{|L|+1}\sum_{m=0}^{\infty}(1-p^2)^m\\
&= q^{|L|}\frac{1}{p^2}+q^{|L|+1}\frac{1}{p^2} \;=\; \frac{1+q}{p^2}\,q^{|L|}. \qquad \text{(A.5)}
\end{aligned}$$

The second sum can be bounded in a similar way using Lemma 2.2, yielding

$$\sum_{k=n+\ell-1}^{2n-2u-1}\mathbb{P}\left(k\notin A+A\right) < \frac{1+q}{p^2}\,q^{|U|}. \qquad \text{(A.6)}$$

Therefore $\mathbb{P}\left(\{2\ell-1,\ldots,n-u-1\}\cup\{n+\ell-1,\ldots,2n-2u-1\}\not\subseteq A+A\right)$ is bounded above by $\frac{1+q}{p^2}\left(q^{|L|}+q^{|U|}\right)$, which is equivalent to the statement of the proposition. □

Proof of Theorem 1.4. For the lower bound, we construct many A such that $A+A$ is missing k elements. First suppose that k is even. Let the first $k/2$ non-negative integers not be in A. Then let the rest of the elements of A be any subset A' that fills in (so $A'+A'$ has no missing elements between its largest and smallest elements); that is $M_{n-k/2}(A')=0$. By Proposition 2.4, we can show that $\mathbb{P}(M_{[0,n-1]}(A')=0)$ is a constant independent of n. If $L\subseteq[0,\ell-1]$ and $U\subseteq[n-u,n-1]$ are fixed, then Proposition 2.4 says that

$$\mathbb{P}([2\ell-1,2n-2u-1]\subseteq A'+A' \mid A'\cap[0,\ell-1]=L, A'\cap[n-u,n-1]=U) > 1-\frac{1+q}{p^2}(q^{|L|}+q^{|U|}), \qquad \text{(A.7)}$$

independent of n. Therefore,

$$\begin{aligned}
&\mathbb{P}([2\ell-1,2n-2u-1]\subseteq A'+A' \text{ and } A'\cap[0,\ell-1]=L, A'\cap[n-u,n-1]=U)\\
&> \left(1-\frac{1+q}{p^2}(q^{|L|}+q^{|U|})\right)q^{\ell}q^{u}. \qquad \text{(A.8)}
\end{aligned}$$

By letting $L = [0, \ell - 1], U = [n - u, n - 1]$ so the ends fill in, we get that

$$\mathbb{P}(A' + A' = [0, 2n - 2]) \ > \ \left(1 - \frac{1+q}{p^2}(q^\ell + q^u)\right) q^\ell q^u. \tag{A.9}$$

Pick ℓ, u large enough so that the first term in the product is positive, we get that

$$\mathbb{P}(A' + A' = [0, 2n - 2]) \ > \ \left(1 - \frac{1+q}{p^2}(q^s + q^s)\right) q^s q^s \ = \ \left(1 - \frac{1+q}{p^2} 2q^s\right) q^{2s}, \tag{A.10}$$

which is a constant independent of n, as desired.

As $A = k/2 + A'$, we have $A + A = k + A' + A' = [k, 2n - 2]$ and so $M_{[0,n-1]}(A) = k$. Thus

$$\begin{aligned} \mathbb{P}(M_{[0,n-1]}(A) = k) \ &\geq \ \mathbb{P}(A = k/2 + A' \text{ and } M_{n-k/2}(A') = 0) \\ &= \ q^{k/2}\mathbb{P}(M_{n-k/2}(A') = 0) \\ &\gg \ q^{k/2}. \end{aligned} \tag{A.11}$$

This proves the lower bound in Theorem 1.4 when k is even.

If k is odd, then we can let $L = [0, \ell - 1] \setminus \{2, 3\}$ and $U = [n - u, n - 1]$ so that only the element 3 is missing from $A' + A'$. Then we get a bound for $\mathbb{P}(M_{[0,n-1]}(A') = 1)$. Letting $A = (k - 1)/2 + A'$, we get the desired lower bound in Theorem 1.4 for when k is odd.

Now, we find the upper bound. For this, we introduce some notation. We set

$$M_{[0,n-1]} \ := \ |[0, 2n - 2]\backslash(A + A)| = 2n - 1 - |A + A|. \tag{A.12}$$

For the upper bound, we have the following inequality for the probability of missing k elements in $[0, n/2]$:

$$\begin{aligned} \mathbb{P}(|[0, n/2]\backslash(A + A)| = k) \ &\leq \ \mathbb{P}(j \notin A + A, j \in [k, n/2]) \\ &\leq \ 2\sum_{j\geq k}(1 - p^2)^{j/2} \\ &\ll \ (1 - p^2)^{k/2}, \end{aligned} \tag{A.13}$$

and similarly for $\mathbb{P}(|[3n/2, 2n]\backslash(A + A)| = k)$. Furthermore, there is an equation ((7.27) from [LMO]) that connects the probability of missing k elements to the probability of missing elements in $[0, n/2]$ and $[3n/2, 2n]$:

$$\mathbb{P}(M_{[0,n-1]}(A) = k) = \sum_{i+j=k} \mathbb{P}(|[0, n/2]\backslash(A+A)| = i)\mathbb{P}(|[3n/2, 2n]\backslash(A+A)| = j) + O\left((1 - p^2)^{n/4}\right). \tag{A.14}$$

Combining A.13 and A.14, we get

$$\begin{aligned} &\mathbb{P}(M_{[0,n-1]}(A) = k) \\ &= \ \sum_{i+j=k} \mathbb{P}(|[0, n/2]\backslash(A + A)| = i)\mathbb{P}(|[3n/2, 2n]\backslash(A + A)| = j) + O\left((1 - p^2)^{n/4}\right) \\ &\ll \ \sum_{i+j=k} (1 - p^2)^{i/2}(1 - p^2)^{j/2} + (1 - p^2)^{n/4} \\ &\ll \ k(1 - p^2)^{k/2} + (1 - p^2)^{n/4}. \end{aligned} \tag{A.15}$$

Therefore, if $k/2 < n/4$, we get

$$\mathbb{P}(M_{[0,n-1]}(A) = k) \ \ll \ k(1 - p^2)^{k/2}. \tag{A.16}$$

However, [LMO] shows we can improve this bound as follows, with the use of (3.11):

$$\begin{aligned}
\mathbb{P}(|[0,n/2]\backslash(A+A)|)=k \ &\leq \ \mathbb{P}(A+A \text{ misses 2 elements greater than } k-3) \\
&= \ \mathbb{P}(i,j\notin A+A, i,j\in[k-3,n/2]) \\
&= \sum_{k-3<i<j}\mathbb{P}(i,j\notin A+A) \\
&\ll \sum_{k-3<i<j}\left(\frac{\phi(p)+1+p}{2\phi(p)}\right)^{\frac{j-i}{2}}\left(\frac{1-p+\phi(p)}{2}\right)^{j+1} \\
&\ll \left(\frac{\phi(p)+1+p}{2\phi(p)}\right)^{\frac{k-k}{2}}\left(\frac{1-p+\phi(p)}{2}\right)^{k+1} \\
&= \left(\frac{1-p+\phi(p)}{2}\right)^{k+1} \ < \ \left(\frac{1-p+\phi(p)}{2}\right)^{k}.
\end{aligned} \tag{A.17}$$

Note that as in (A.15), we always have an extra $(1-p^2)^{n/4}$ term. To make this term negligible, we need to have $(1-p^2)^{n/4}<((1-p+\phi(p))/2)^k$, which means $n>k\cdot 4\log((1-p+\phi(p))/2)/\log(1-p^2)$. This condition is sufficient in this case where we have the bound $((1-p+\phi(p))/2)^k$. However, in general, we know that we have a lower bound of $(1-p)^{k/2}$ for the distribution. Therefore, to make the $(1-p^2)^{n/4}$ term always negligible, we can have $(1-p^2)^{n/4}<(1-p)^{k/2}$, which means $n>k\cdot 2\log(1-p)/\log(1-p^2)$, as in the statement of Theorem 1.4. Note that then the implied constants are independent of n. Combining (A.11) and (A.17), we get Theorem 1.4. □

Appendix B. Our Bounds for $\mathbb{P}(|B|=k)$ Are Good

To observe numerically how good our bounds are, we must compare our bounds to the true values of $\mathbb{P}(|B|=k)$. However, $\mathbb{P}(|B|=k)$ cannot be computed directly; thus, we run simulations to estimate $\mathbb{P}(|B|=k)$. We pick $p\in(0,1)$ and run 10^6 simulations to form subsets of $\{0,1,\dots,400\}$ and find the frequency of each number of missing sums within these 10^6 simulations. We then compare the plot of the simulated distribution with our bound functions mentioned in Corollary 6.5 and Corollary 6.10.

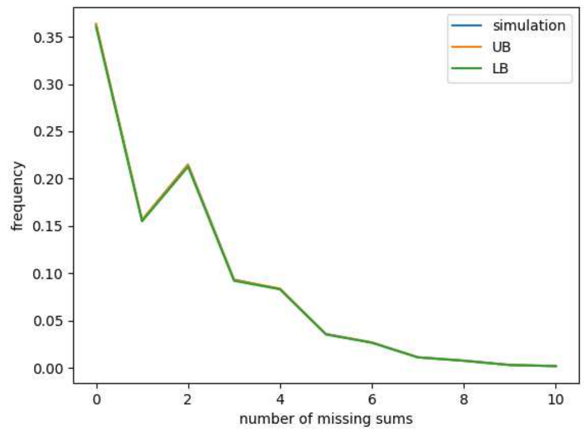

FIGURE 9. For $p = 0.8$, lower bound, upper bound and simulation of $\mathbb{P}(|B| = k)$. At $p = 0.8$, the lower bound and upper bound for $\mathbb{P}(|B| = k)$ are so good that we cannot differentiate the lines. The two bounds and the simulation seem to closely coincide at all points.

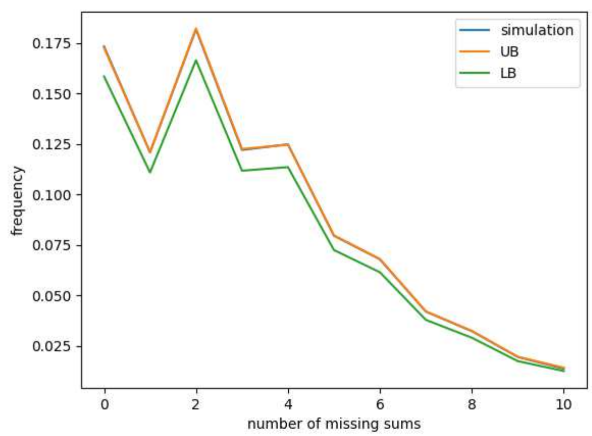

FIGURE 10. At $p = 0.7$, the lower bound and upper bound for $\mathbb{P}(|B| = k)$ are still close to each other. The upper bound seems to coincide with the simulation everywhere. However, the bounds are relatively worse compared to the case $p = 0.8$.

APPENDIX C. DATA FOR DIVOT COMPUTATIONS

All the data we provide below corresponds to $\ell = 30$ and $a = 12$. Our program computes all the quantities required by Inequalities (6.5) and (6.10) to find lower and upper bounds for $m_p(k)$ when p varies. Again, our method of storing and collecting data are mentioned at the end of Section 1.

C.0.1. *Data for* $\min L_i^{12}$ *and* $\tau(L_i^{12})$.

i	0	1	2	3	4	5
$\min L_i^{12}$	12	11	11	11	11	11
$\tau(L_i^{12})$	7	7	6	6	6	6

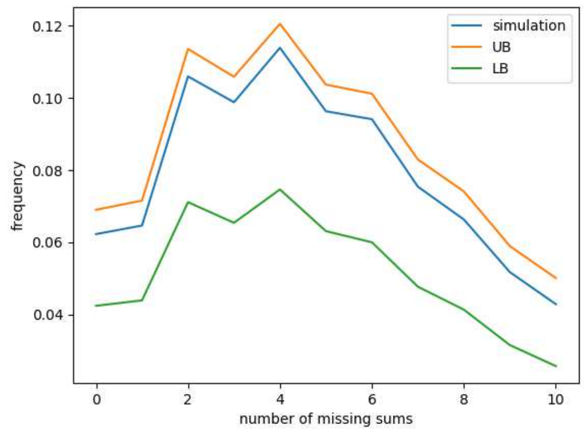

FIGURE 11. At $p = 0.6$, the upper bound is fairly good while the lower bound is much worse compared to previous cases.

C.0.2. *Data for* $c_k(i)$ *for* $0 \leq k \leq 2$.

i	$c_0(i)$	$c_1(i)$	$c_2(i)$
0-7	0	0	0
8	0	0	12
9	58	1552	13955
10	10629	82696	276434
11	190349	704139	1495762
12	1164105	2613360	4544680
13	3879603	6121208	9753610
14	8720201	10586952	16142608
15	14730206	14526747	21525160
16	19817016	16371555	23716940
17	21916190	15421977	21913801
18	20269375	12251022	17114758

i	$c_0(i)$	$c_1(i)$	$c_2(i)$
19	15817037	8237988	11333625
20	10452359	4689056	6359012
21	5847957	2251741	3010077
22	2759881	906081	1192562
23	1090747	302191	390638
24	356894	82172	103915
25	95055	17777	21870
26	20099	2947	3501
27	3248	352	400
28	377	27	29
29	28	1	1
30	1	0	0

C.0.3. *Data for* $c_{k,a}(i)$.

i	$c_{0,12}(i)$	$c_{1,12}(i)$	$c_{2,12}(i)$
0-10	0	0	0
11	0	5	27
12	400	3352	18890
13	39072	198265	589832
14	685029	1857746	3772428
15	3664341	6033358	9993760
16	9311984	10449491	15740073
17	14592372	12237242	17647078
18	16358625	10909486	15345556
19	14202656	7803902	10775362
20	9943771	4584649	6229436
21	5729373	2234092	2989242
22	2740544	904203	1190494
23	1088774	302096	390543
24	356799	82172	103915
25	95055	17777	21870
26	20099	2947	3501
27	3248	352	400
28	377	27	29
29	28	1	1
30	1	0	0

REFERENCES

[AMMS] M. Asada, S. Manski, S. J. Miller, and H. Suh, *Fringe pairs in generalized MSTD sets*, International Journal of Number Theory **13** (2017), no. 10, 2653–2675.

[BELM] A. Bower, R. Evans, V. Luo and S. J. Miller, *Coordinate sum and difference sets of d-dimensional modular hyperbolas*, INTEGERS #A31, 2013, 16 pages.

[CLMS] H. Chu, N. Luntzlara, S. J. Miller and L. Shao, *Generalizations of a Curious Family of MSTD Sets Hidden By Interior Blocks*, to appear in Integers.

[CMMXZ] H. Chu, N. McNew, S. J. Miller, V. Xu and S. Zhang, *When Sets Can and Cannot Have MSTD Subsets*, Journal of Integer Sequences **21** (2018), Article 18.8.2.

[DKMMW] 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.

[H-AMP] S. Harvey-Arnold, S. J. Miller and F. Peng, *Distribution of missing differences in diffsets*, preprint.

[He] P. V. Hegarty, *Some explicit constructions of sets with more sums than differences* (2007), Acta Arithmetica **130** (2007), no. 1, 61–77.

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

[HLM] A. Hemmady, A. Lott and S. J. Miller, *When almost all sets are difference dominated in $\mathbb{Z}/n\mathbb{Z}$*, Integers **17** (2017), Paper No. A54, 15 pp.

[ILMZ] 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.

[LMO] O. Lazarev, S. J. Miller, K. O'Bryant, *Distribution of Missing Sums in Sumsets* (2013), Experimental Mathematics **22**, no. 2, 132–156.

[Ma] J. Marica, *On a conjecture of Conway*, Canad. Math. Bull. **12** (1969), 233–234.

[MO] 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.

[MOS] S. J. Miller, B. Orosz and D. Scheinerman, *Explicit constructions of infinite families of MSTD sets*, Journal of Number Theory **130** (2010) 1221–1233.

[MS] S. J. Miller and D. Scheinerman, *Explicit constructions of infinite families of mstd sets,* Additive Number Theory, Springer, 2010, pp. 229-248.

[MPR] S. J. Miller, S. Pegado and L. Robinson, *Explicit Constructions of Large Families of Generalized More Sums Than Differences Sets*, Integers **12** (2012), #A30.

[MV] S. J. Miler and K. Vissuet, *Most Subsets are Balanced in Finite Groups*, Combinatorial and Additive Number Theory, CANT 2011 and 2012 (Melvyn B. Nathanson, editor), Springer Proceedings in Mathematics & Statistics (2014), 147–157.

[Na1] M. B. Nathanson, *Problems in additive number theory, 1*, Additive combinatorics, 263–270, CRM Proc. Lecture Notes **43**, Amer. Math. Soc., Providence, RI, 2007.

[Na2] M. B. Nathanson, *Sets with more sums than differences*, Integers : Electronic Journal of Combinatorial Number Theory **7** (2007), Paper A5 (24pp).

[PW] D. Penman and M. Wells, On sets with more restricted sums than differences, *Integers* **13** (2013), #A57.

[Ru1] I. Z. Ruzsa, *On the cardinality of $A + A$ and $A - A$*, Combinatorics year (Keszthely, 1976), vol. 18, Coll. Math. Soc. J. Bolyai, North-Holland-Bolyai Tàrsulat, 1978, 933–938.

[Ru2] I. Z. Ruzsa, *Sets of sums and differences*. In: Séminaire de Théorie des Nombres de Paris 1982-1983, pp. 267–273. Birkhäuser, Boston (1984).

[Ru3] I. Z. Ruzsa, *On the number of sums and differences*, Acta Math. Sci. Hungar. **59** (1992), 439–447.

[Sp] W. G. Spohn, On Conway's conjecture for integer sets, *Canad. Math. Bull* **14** (1971), 461-462.

[Zh1] Y. Zhao, *Constructing MSTD sets using bidirectional ballot sequences*, Journal of Number Theory **130** (2010), no. 5, 1212–1220.

[Zh2] Y. Zhao, *Sets characterized by missing sums and differences*, Journal of Number Theory **131** (2011), no. 11, 2107–2134.

Email address: chuh19@mail.wlu.edu

DEPARTMENT OF MATHEMATICS, WASHINGTON AND LEE UNIVERSITY, LEXINGTON, VA 24450

Email address: kingda16@wfu.edu

DEPARTMENT OF MATHEMATICS, WAKE FOREST UNIVERSITY, WINSTON-SALEM, NC 27109

Email address: nluntzla@umich.edu

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF MICHIGAN, ANN ARBOR, MI 48109

Email address: tmartinez@hmc.edu

DEPARTMENT OF MATHEMATICS, HARVEY MUDD COLLEGE, CLAREMONT, CA 91711

Email address: sjm1@williams.edu, Steven.Miller.MC.96@aya.yale.edu

DEPARTMENT OF MATHEMATICS AND STATISTICS, WILLIAMS COLLEGE, WILLIAMSTOWN, MA 01267

Email address: ls12@williams.edu

DEPARTMENT OF MATHEMATICS AND STATISTICS, WILLIAMS COLLEGE, WILLIAMSTOWN, MA 01267

Email address: cs19@williams.edu

DEPARTMENT OF MATHEMATICS AND STATISTICS, WILLIAMS COLLEGE, WILLIAMSTOWN, MA 0126

Email address: vzx@andrew.cmu.edu

DEPARTMENT OF MATHEMATICS, CARNEGIE MELLON UNIVERSITY, PITTSBURGH, PA 15289