Phase Transitions in the Distribution of Missing Sums and a Powerful Family of MSTD Sets

Steven J Miller (Williams College)
Email: sjm1@williams.edu
https://web.williams.edu/Mathematics/sjmiller/public_html/math/talks/talks.html

With Hung Viet Chu, Noah Luntzlara, Lily Shao, Victor Xu

INTEGERS Conference, Augusta, October 4, 2018

Outline

- Introduction to MSTD sets
- Divot at 1
- A powerful family of MSTD sets
- Future research

Introduction

Statement

- A finite set of integers, $|A|$ its size.
- The sumset: $A+A=\{a_i+a_j|a_i,a_j\in A\}$.
- The difference set: $A-A=\{a_i-a_j|a_i,a_j\in A\}$.

Definition

A finite set of integers. A is called **sum-dominated** or MSTD (more-sum-than-difference) if $|A+A|>|A-A|$, **balanced** if $|A+A|=|A-A|$ and **difference-dominated** if $|A+A|<|A-A|$.

False conjecture

- Natural to think that $|A + A| \leq |A - A|$.
- Each pair (x, y), $x \neq y$ gives two differences: $x - y \neq y - x$, but only one sum $x + y$.
- However, sets A with $|A + A| > |A - A|$ do exist!
  Conway (1969): $\{0, 2, 3, 4, 7, 11, 12, 14\}$.

Martin and O'Bryant '06

Theorem

Consider $I_n = \{0, 1, ..., n-1\}$. The proportion of MSTD subsets of I_n is bounded below by a positive constant $c \approx 2 \cdot 10^{-7}$.

Results

Distribution of the Number of Missing Sums (Uniform Model)

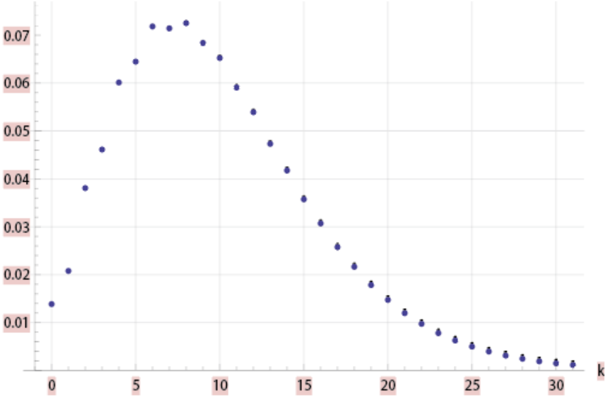


Figure: Frequency of the number of missing sums (q is probability of not choosing an element). The distribution is not unimodal. From Lazarev-Miller-O'Bryant.

Distribution of the Number of Missing Sums (Different Models)

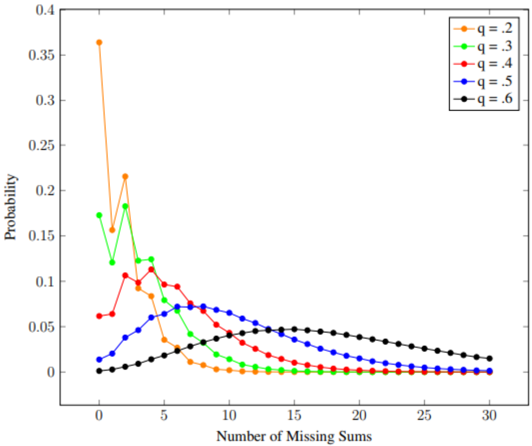


Figure: Missing sums (q is probability of not choosing an element) from simulating 1,000,000 subsets of $\{0, 1, 2, ..., 255\}$.

Sets of Missing Sums

- Let $I_n = \{0, 1, 2, ..., n-1\}$.
- Form $S \subseteq I_n$ randomly with probability p of picking an element in I_n ($q = 1 - p$: the probability of not choosing an element).
- $B_n = (I_n + I_n) \backslash (S + S)$ is the set of missing sums, $|B_n|$: the number of missing sums.

Distribution of Missing Sums

- Fix $p \in (0,1)$, study $\mathbb{P}(|B| = k) = \lim_{n\to\infty} \mathbb{P}(|B_n| = k)$. (Zhao proved that the limit exists.)
- $\mathbb{P}(|B| = k)$: the limiting distribution of missing sums.

Distribution of Missing Sums

- Fix $p \in (0,1)$, study $\mathbb{P}(|B| = k) = \lim_{n\to\infty} \mathbb{P}(|B_n| = k)$. (Zhao proved that the limit exists.)
- $\mathbb{P}(|B| = k)$: the limiting distribution of missing sums.

Divot

For some $k \geq 1$, have **divot** at k if
$\mathbb{P}(|B| = k-1) > \mathbb{P}(|B| = k) < \mathbb{P}(|B| = k+1)$.

Example of Divot at 3

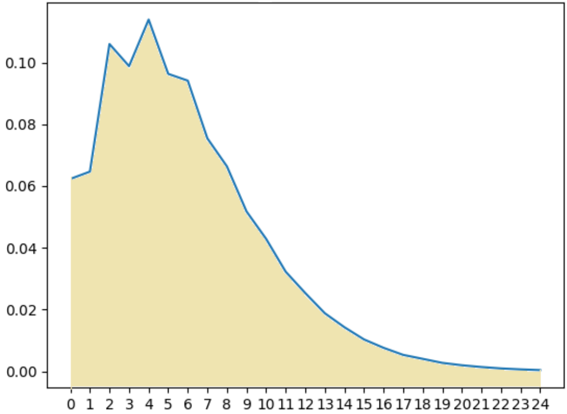


Figure: Frequency of the number of missing sums for subsets of $\{0, 1, 2, ..., 400\}$ by simulating 1,000,000 subsets with $p = 0.6$.

Numerical Analysis for $p = 1/2$

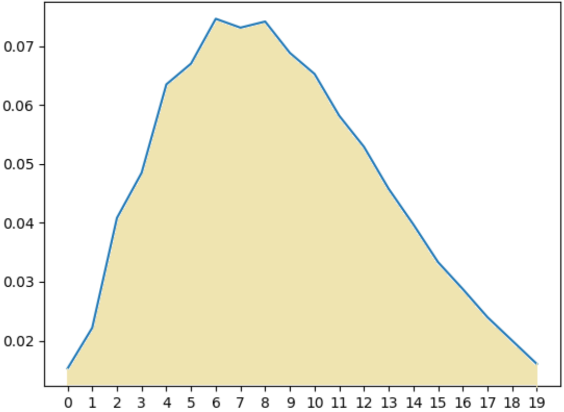


Figure: Frequency of the number of missing sums for all subsets of $\{0, 1, 2, ..., 25\}$.

Lazarev-Miller-O'Bryant '11

Divot at 7

For $p = 1/2$, there is a divot at 7:
$\mathbb{P}(|B| = 6) > \mathbb{P}(|B| = 7) < \mathbb{P}(|B| = 8)$.

Question

Existence of Divots

For a fixed different value of p, are there other divots?

Question

Existence of Divots

For a fixed different value of p, are there other divots?

Answer: Yes!

Numerical analysis for different $p \in (0, 1)$: $p = 0.6$

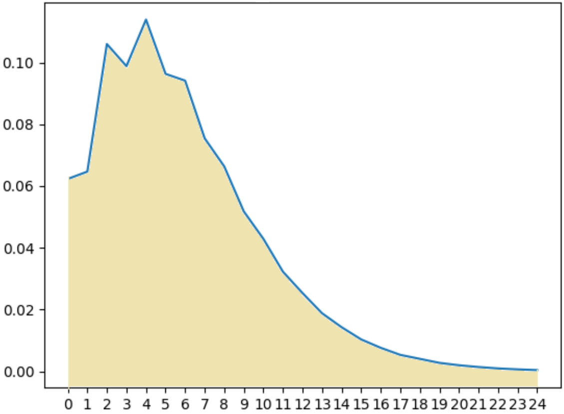


Figure: Distribution of $|B| = k$ by simulating 1,000,000 subsets of $\{0, 1, 2, \ldots, 400\}$ with $p = 0.6$.

Numerical analysis for $p = 0.7$: divots at 1 and 3

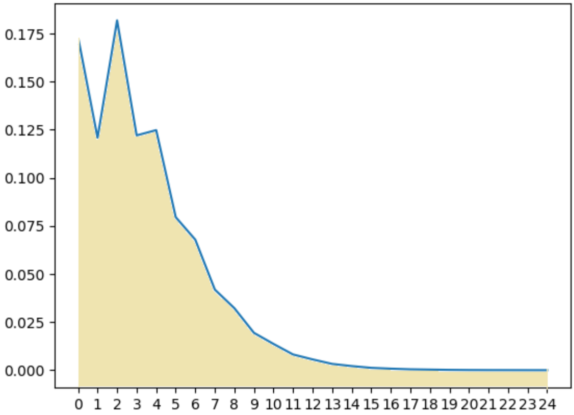


Figure: Distribution of $|B| = k$ by simulating 1,000,000 subsets of $\{0, 1, 2, \ldots, 400\}$ with $p = 0.7$.

Numerical analysis for different $p = 0.8$: divot at 1

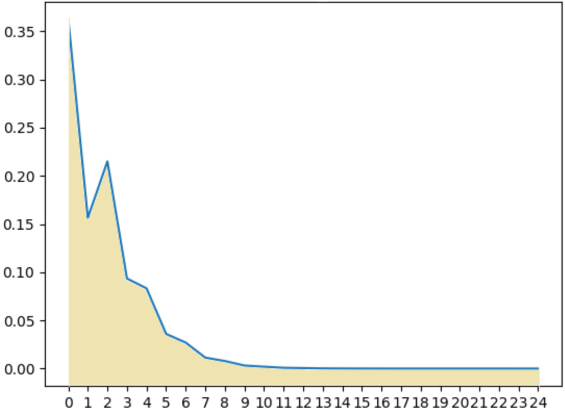


Figure: Distribution of $|B| = k$ by simulating 1,000,000 subsets of $\{0, 1, 2, \ldots, 400\}$ with $p = 0.8$.

Numerical analysis for different $p = 0.9$: divot at 1

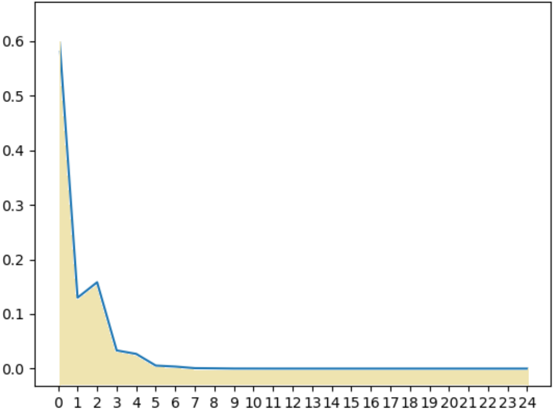


Figure: Distribution of $|B| = k$ by simulating 1,000,000 subsets of $\{0, 1, 2, \ldots, 400\}$ with $p = 0.9$.

Main Result

Divot at 1 [CLMSX'18]

For $p \geq 0.68$, there is a divot at 1:
$\mathbb{P}(|B| = 0) > \mathbb{P}(|B| = 1) < \mathbb{P}(|B| = 2)$. Empirical evidence predicts the value of p such that the divot at 1 starts to exist is between 0.6 and 0.7.

Sketch of Proof

Key Ideas

- Want $\mathbb{P}(|B| = 0) > \mathbb{P}(|B| = 1) < \mathbb{P}(|B| = 2)$.

Key Ideas

- Want $\mathbb{P}(|B| = 0) > \mathbb{P}(|B| = 1) < \mathbb{P}(|B| = 2)$.
- Establish an upper bound T^1 for $\mathbb{P}(|B| = 1)$.

Key Ideas

- Want $\mathbb{P}(|B| = 0) > \mathbb{P}(|B| = 1) < \mathbb{P}(|B| = 2)$.
- Establish an upper bound T^1 for $\mathbb{P}(|B| = 1)$.
- Establish lower bounds T_0 and T_2 for $\mathbb{P}(|B| = 0)$ and $\mathbb{P}(|B| = 2)$, respectively.

Key Ideas

- Want $\mathbb{P}(|B| = 0) > \mathbb{P}(|B| = 1) < \mathbb{P}(|B| = 2)$.
- Establish an upper bound T^1 for $\mathbb{P}(|B| = 1)$.
- Establish lower bounds T_0 and T_2 for $\mathbb{P}(|B| = 0)$ and $\mathbb{P}(|B| = 2)$, respectively.
- Find values of p such that $T_2 > T^1 < T_0$.

Fringe Analysis

- Most of the missing sums come from the fringe: many more ways to form middle elements than fringe elements.

Fringe Analysis

- Most of the missing sums come from the fringe: many more ways to form middle elements than fringe elements.
- Fringe analysis is enough to find good lower bounds and upper bounds for $\mathbb{P}(|B| = k)$.

Setup

- Consider $S \subseteq \{0, 1, 2, ..., n-1\}$ with probability p of each element being picked.
- Analyze fringe of size 30.
- Write $S = L \cup M \cup R$, where $L \subseteq [0, 29]$, $M \subseteq [30, n-31]$ and $R \subseteq [n-30, n-1]$.

Notation

- Write $S = L \cup M \cup R$, where $L \subseteq [0, 29]$, $M \subseteq [30, n-31]$ and $R \subseteq [n-30, n-1]$.
- L_k : the event that $L + L$ misses k sums in $[0, 29]$.
- L_k^a: the event that $L + L$ misses k sums in $[0, 29]$ and contains $[30, 48]$.
- Similar notations applied for R.

Upper Bound

Given $0 \leq k \leq 30$,

$$\mathbb{P}(|B| = k) \;\leq\; \sum_{i=0}^{k} \mathbb{P}(L_i)\mathbb{P}(L_{k-i}) + \frac{2(2q-q^2)^{15}(3q-q^2)}{(1-q)^2}.$$

Lower Bound

Given $0 \leq k \leq 30$,

$$\mathbb{P}(|B| = k) \geq \sum_{i=0}^{k} \Bigg[1 - (a-2)(q^{\tau(L_i^a)} + q^{\tau(L_{k-i}^a)}) - \frac{1+q}{(1-q)^2}(q^{\min L_i^a} + q^{\min L_{k-i}^a}) \Bigg] \mathbb{P}(L_i^a)\mathbb{P}(L_{k-i}^a).$$

Our Bounds Are Fairly Sharp ($p \geq 0.7$)

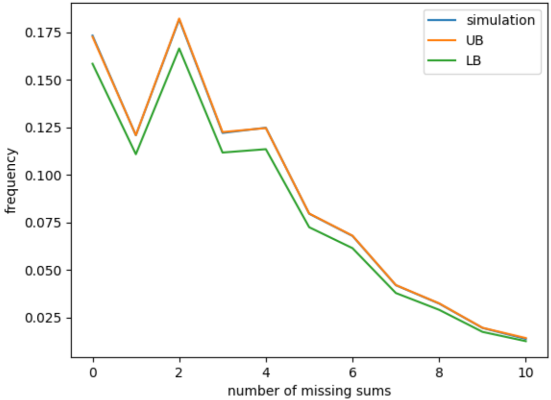
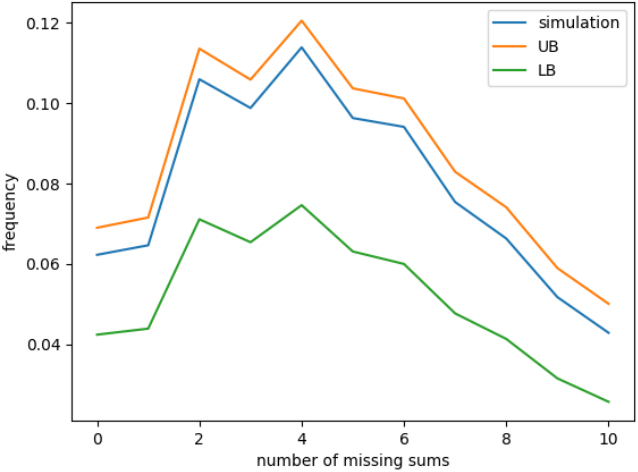


Figure: We cannot see the blue line because our upper bound is so sharp that the orange line lies on the blue line.

Our Bounds Are Bad ($p \leq 0.6$)



Divot at 1

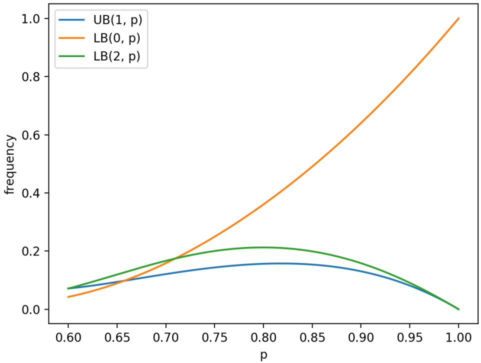


Figure: For $p \geq 0.68$, the lower bounds for $\mathbb{P}(|B| = 0)$ and $\mathbb{P}(|B| = 2)$ are higher than the upper bound for $\mathbb{P}(|B| = 1)$. There is a divot at 1.

A Powerful Family of MSTD sets

Why Powerful?

- Have appeared in the proof of many important results in previous works.
- Give many sets with large $\log|A+A|/\log|A-A|$.
- Economically way to construct sets with fixed $|A+A|-|A-A|$ (save more than four times of what previous construction has).
- A is restricted-sum-dominant (RSD) if its restricted sum set is bigger than its difference set. Improve the lower bound for the proportion of RSD sets from 10^{-37} to 10^{-25}.

A different notation

- We use a different notation to write a set; was first introduced by Spohn (1973).
- Given a set $S = \{a_1, a_2, \dots, a_n\}$, we arrange its elements in increasing order and find the differences between two consecutive numbers to form a sequence.
- For example, $S = \{2, 3, 5, 9, 10\}$. We write $S = (2|1, 2, 4, 1)$.

$\mathcal{F}$ family

$\mathcal{F}$ family

Let M^k denote $1, \underbrace{4, \dots, 4}_{k\text{-times}}, 3$. Our family is

$$\mathcal{F} := \{1, 1, 2, 1, M^{k_1}, M^{k_2}, \dots, M^{k_\ell}, M_1 : \ell, k_1, \dots, k_\ell \in \mathbb{N}\},$$

where M_1 is either $1, 1$ or $1, 1, 2$ or $1, 1, 2, 1$.

Conjecture

All sets in $\mathcal{F}$ are MSTD.

We proved that the conjecture holds for a periodic family.

Periodic Family [CLMS'18]

$$S_{k,\ell} = (0|1,1,2,1,\underbrace{\underbrace{4,\dots,4}_{k\text{-times}},3,\dots,1,\underbrace{4,\dots,4}_{k\text{-times}}}_{\ell\text{-times}},3,1,1,2,1)$$

has $|S_{k,\ell}+S_{k,\ell}|-|S_{k,\ell}-S_{k,\ell}|=2\ell$.

$$S'_{k,\ell} = (0|1,1,2,1,\underbrace{\underbrace{4,\dots,4}_{k\text{-times}},3,\dots,1,\underbrace{4,\dots,4}_{k\text{-times}}}_{\ell\text{-times}},3,1,1,2)$$

has $|S'_{k,\ell}+S'_{k,\ell}|-|S'_{k,\ell}-S'_{k,\ell}|=2\ell-1$.

First Application

Sets A with fixed $|A+A| - |A-A|$

Given $x \in \mathbb{N}$, there exists a set $A \subseteq [0, 12+4x]$ such that $|A+A| - |A-A| = x$. (Previous was $[0, 17x]$).

We save more than four times!

Method: Explicit constructions using $S_{k,\ell}$ and $S'_{k,\ell}$.

Second Application

Lower bound for restricted-sum-dominant sets

For $n \geq 81$, the proportion of RSD subsets of $\{0, 1, 2, \ldots, n-1\}$ is at least $4.135 \cdot 10^{-25}$. (Previous was about 10^{-37}).

Method: $S_{k,\ell}$ reduces the needed fringe size from 120 to 81.

Future Research

Distribution of Missing Sums

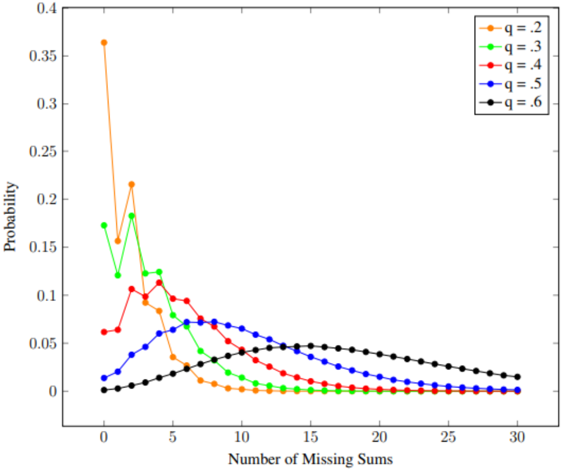


Figure: Shift of Divots....

Future Research

- Prove there are no divots at even numbers.
- Is there a value of p such that there are no divots?
- What about missing differences?
- What if probability of choosing depends on n?

Work supported by NSF Grants DMS1561945 and DMS1659037, the Finnerty Fund, the University of Michigan, Washington and Lee, and Williams College.

Bibliography

Bibliography

- O. Lazarev, S. J. Miller, K. O'Bryant, *Distribution of Missing Sums in Sumsets* (2013), Experimental Mathematics **22**, no. 2, 132–156.
- P. V. Hegarty, *Some explicit constructions of sets with more sums than differences* (2007), Acta Arithmetica **130** (2007), no. 1, 61–77.
- P. V. Hegarty and S. J. Miller, *When almost all sets are difference dominated*, Random Structures and Algorithms **35** (2009), no. 1, 118–136.
- 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.

Bibliography

- J. Marica, *On a conjecture of Conway*, Canad. Math. Bull. **12** (1969), 233–234.
- 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.
- 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.
- S. J. Miller, B. Orosz and D. Scheinerman, *Explicit constructions of infinite families of MSTD sets*, Journal of Number Theory **130** (2010), 1221–1233.