More Sums than Differences Sets and Beyond

Yufei Zhao

Massachusetts Institute of Technology

AMS/MAA Joint Math Meetings
January 14, 2010

Sum sets and difference sets

For a finite set $S \subset \mathbb{Z}$, let

$$S + S = \{a + b : a, b \in S\}$$
$$S - S = \{a - b : a, b \in S\}$$

Question

Which set is bigger?

Sum sets and difference sets

For a finite set $S \subset \mathbb{Z}$, let

$$S + S = \{a + b : a, b \in S\}$$
$$S - S = \{a - b : a, b \in S\}$$

Question

Which set is bigger?

Example: $S = \{0, 1, 3, 8\}$

$S + S = \{0, 1, 2, 3, 4, 6, 8, 9, 11, 16\}$ 10 elements

$S - S = \{-8, -7, -5, -3, -2, -1, 0, 1, 2, 3, 5, 7, 8\}$ 13 elements

Sum sets and difference sets

For a finite set $S \subset \mathbb{Z}$, let

$$S + S = \{a + b : a, b \in S\}$$
$$S - S = \{a - b : a, b \in S\}$$

Question

Which set is bigger?

Example: $S = \{0, 1, 3, 8\}$

$$S + S = \{0, 1, 2, 3, 4, 6, 8, 9, 11, 16\} \qquad \text{10 elements}$$
$$S - S = \{-8, -7, -5, -3, -2, -1, 0, 1, 2, 3, 5, 7, 8\} \qquad \text{13 elements}$$

- Since addition is commutative while subtraction is not, two distinct elements generate one sum but two differences.
- So we should expect there to be more differences.

A counterexample

- It was thought that perhaps $|S+S| \leq |S-S|$ for every finite $S \subset \mathbb{Z}$.
- However, in the 1960's, Conway found the following counterexample:

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

We have

$$S + S = [0, 28] \setminus \{1, 20, 27\} \qquad \text{26 elements}$$
$$S - S = [-14, 14] \setminus \{-13, -6, 6, 13\} \qquad \text{25 elements}$$

- So it began the search for more of such sets ...

MSTD sets

In 2006, Nathanson revived the topic, can called sets S with $|S+S| > |S-S|$ MSTD (more sums than differences) sets.

MSTD sets

In 2006, Nathanson revived the topic, can called sets S with $|S+S| > |S-S|$ MSTD (more sums than differences) sets.

> *Even though there exist sets A that have more sums than differences, such sets should be rare, and it must be true with the right way of counting that the vast majority of sets satisfies $|A-A| > |A+A|$. — Nathanson, 2006*

MSTD sets

In 2006, Nathanson revived the topic, can called sets S with $|S+S|>|S-S|$ MSTD (more sums than differences) sets.

> *Even though there exist sets A that have more sums than differences, such sets should be rare, and it must be true with the right way of counting that the vast majority of sets satisfies $|A-A|>|A+A|$. — Nathanson, 2006*

Let's look at subsets of $\{0, 1, \ldots, n\}$.

- [Martin, O'Bryant 2007] MSTD sets are abundant.
- [Hegarty, Miller 2009] MSTD sets are rare.

MSTD sets

In 2006, Nathanson revived the topic, can called sets S with $|S+S| > |S-S|$ MSTD (more sums than differences) sets.

> *Even though there exist sets A that have more sums than differences, such sets should be rare, and it must be true with the right way of counting that the vast majority of sets satisfies $|A-A| > |A+A|$. — Nathanson, 2006*

Let's look at subsets of $\{0, 1, \ldots, n\}$.

- [Martin, O'Bryant 2007] MSTD sets are abundant. $\leftarrow$ uniform model
- [Hegarty, Miller 2009] MSTD sets are rare. $\leftarrow$ sparse model

MSTD sets

In 2006, Nathanson revived the topic, can called sets S with $|S+S|>|S-S|$ MSTD (more sums than differences) sets.

> *Even though there exist sets A that have more sums than differences, such sets should be rare, and it must be true with the right way of counting that the vast majority of sets satisfies* $|A-A|>|A+A|$. *— Nathanson, 2006*

Let's look at subsets of $\{0, 1, \ldots, n\}$.

- [Martin, O'Bryant 2007] MSTD sets are abundant. ← uniform model
- [Hegarty, Miller 2009] MSTD sets are rare. ← sparse model

Later, Nathanson wrote:

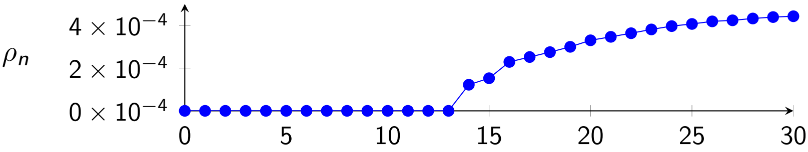
> *A difficult and subtle problem is to decide what is the appropriate method of counting (or, equivalently, the appropriate probability measure) to apply to MSTD sets.*

Proportion of MSTD sets

Theorem (Martin and O'Bryant, 2006)

Let $\rho_n 2^{n+1}$ be the number MSTD subsets of $\{0, 1, \ldots, n\}$. Then $\rho_n \geq 2 \times 10^{-7}$ for $n \geq 14$.

Conjecture: ρ_n has a limit; estimated at 4.5×10^{-4} using Monte Carlo.

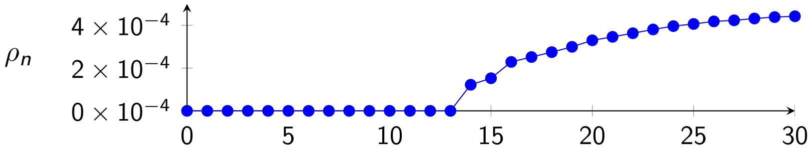


Proportion of MSTD sets

Theorem (Martin and O'Bryant, 2006)

Let $\rho_n 2^{n+1}$ be the number MSTD subsets of $\{0, 1, \ldots, n\}$. Then $\rho_n \geq 2 \times 10^{-7}$ for $n \geq 14$.

Conjecture: ρ_n has a limit; estimated at 4.5×10^{-4} using Monte Carlo.



My result

ρ_n converges to a limit $\rho > 4 \times 10^{-4}$.

Furthermore, we have a deterministic algorithm that could, in principle, compute ρ up to arbitrary precision.

Intuition Behind MSTD Sets

- Fringe is important. Middle matters less.
- With high probability, most of the middle sums and differences will be present.

Intuition Behind MSTD Sets

- Fringe is important. Middle matters less.
- With high probability, most of the middle sums and differences will be present.

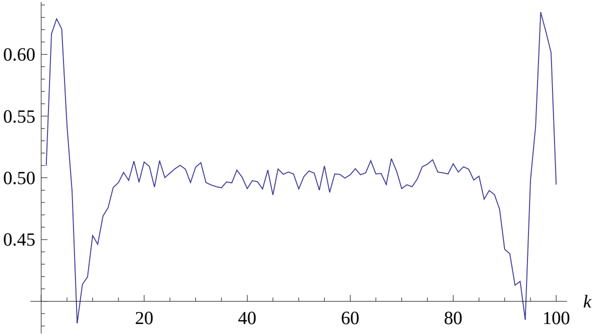
- This intuition helped to prove many results about MSTD sets.
- However, there has been no description on what "most" MSTD looks like.
- We address this question by giving a rigorous formulation of the intuition.

Behavior of the middle portion?

For uniform random subset $S \subset [1, n]$, let

$$\gamma(k, n) = \mathbb{P}(k \in S \mid S \text{ is MSTD})$$

Estimated values of $\gamma(k, 100)$:



Miller et al. [MOS] conjectured that, for any constant $0 < c < 1/2$, if $cn < k < n - cn$, then $\gamma(k, n) \to 1/2$ as $n \to \infty$.

What does a typical MSTD set look like?

Answer: A well-controlled fringe and an almost unrestricted middle.

Notation: $[a, b] = \{a, a+1, \ldots, b\}$.

What does a typical MSTD set look like?

Answer: A well-controlled fringe and an almost unrestricted middle.

Notation: $[a, b] = \{a, a+1, \dots, b\}$.

S

Theorem

- $\alpha_n \in \mathbb{Z}$ satisfying $0 < \alpha_n < n/2$ and $\alpha_n \to \infty$ as $n \to \infty$
- S a uniform random subset of $[0, n]$
- E an event that depends only on $S \cap [\alpha_n + 1, n - \alpha_n - 1]$

Then, as $n \to \infty$,

$$|\mathbb{P}(E \mid S \text{ is MSTD}) - \mathbb{P}(E)| = O\left((3/4)^{\alpha_n}\right).$$

What does a typical MSTD set look like?

Answer: A well-controlled fringe and an almost unrestricted middle.

Notation: $[a, b] = \{a, a+1, \dots, b\}$.

S

Theorem

- $\alpha_n \in \mathbb{Z}$ satisfying $0 < \alpha_n < n/2$ and $\alpha_n \to \infty$ as $n \to \infty$
- S a uniform random subset of $[0, n]$
- E an event that depends only on $S \cap [\alpha_n + 1, n - \alpha_n - 1]$
- F an event that depends only on $S \cap ([0, \alpha_n] \cup [n - \alpha_n, n])$

Then, as $n \to \infty$,

$$|\mathbb{P}(E \cap F \mid S \text{ is MSTD}) - \mathbb{P}(E)\mathbb{P}(F \mid S \text{ is MSTD})| = O\left((3/4)^{\alpha_n}\right).$$

Consequences

The theorem resolves Miller et at.'s conjecture that middle elements appear in MSTD sets with probability approaching $1/2$.

Consequences

The theorem resolves Miller et at.'s conjecture that middle elements appear in MSTD sets with probability approaching 1/2.

Other consequences of the theorem: as $n \to \infty$, for a uniform random subset $S \subset [1, n]$

- $\mathbb{E}[|S| \mid S \text{ is MSTD}] = \frac{n}{2} + O(\log n)$

Consequences

The theorem resolves Miller et at.'s conjecture that middle elements appear in MSTD sets with probability approaching $1/2$.

Other consequences of the theorem: as $n \to \infty$, for a uniform random subset $S \subset [1, n]$

- $\mathbb{E}[|S| \mid S \text{ is MSTD}] = \frac{n}{2} + O(\log n)$
- $\mathrm{Var}\left(|S| \mid S \text{ is MSTD}\right) = \frac{n}{4} + O((\log n)^2)$

Consequences

The theorem resolves Miller et at.'s conjecture that middle elements appear in MSTD sets with probability approaching $1/2$.

Other consequences of the theorem: as $n \to \infty$, for a uniform random subset $S \subset [1, n]$

- $\mathbb{E}[|S| \mid S \text{ is MSTD}] = \frac{n}{2} + O(\log n)$
- $\mathrm{Var}\left(|S| \mid S \text{ is MSTD}\right) = \frac{n}{4} + O((\log n)^2)$
- Central limit theorem: for any $t \in \mathbb{R}$,

$$\mathbb{P}\left(|S| < \frac{n + t\sqrt{n}}{2} \,\middle|\, S \text{ is MSTD}\right) \to \Phi(t)$$

where $\Phi(t)$ is the standard normal distribution.

Consequences

The theorem resolves Miller et at.'s conjecture that middle elements appear in MSTD sets with probability approaching $1/2$.

Other consequences of the theorem: as $n \to \infty$, for a uniform random subset $S \subset [1, n]$

- $\mathbb{E}[|S| \mid S \text{ is MSTD}] = \frac{n}{2} + O(\log n)$
- $\operatorname{Var}(|S| \mid S \text{ is MSTD}) = \frac{n}{4} + O((\log n)^2)$
- Central limit theorem: for any $t \in \mathbb{R}$,

$$\mathbb{P}\left(|S| < \frac{n + t\sqrt{n}}{2} \,\middle|\, S \text{ is MSTD}\right) \to \Phi(t)$$

  where $\Phi(t)$ is the standard normal distribution.
- So the size distribution of MSTD sets is very similar to the unrestricted binomial distribution.

Beyond MSTD sets

What proportion of subsets $S \subset \{0, 1, \ldots, n\}$ . . .

Beyond MSTD sets

What proportion of subsets $S \subset \{0, 1, \ldots, n\}$ ...

- have more differences than sums, i.e., $|S + S| < |S - S|$?

Beyond MSTD sets

What proportion of subsets $S \subset \{0, 1, \ldots, n\}$ ...

- have more differences than sums, i.e., $|S + S| < |S - S|$?
- have the equal number differences and sums, i.e., $|S + S| = |S - S|$?

Beyond MSTD sets

What proportion of subsets $S \subset \{0, 1, \ldots, n\}$ ...

- have more differences than sums, i.e., $|S + S| < |S - S|$?
- have the equal number differences and sums, i.e., $|S + S| = |S - S|$?
- are missing exactly s sums and d differences, i.e., $|S + S| = 2n + 1 - s$, $|S - S| = 2n + 1 - d$?

Beyond MSTD sets

What proportion of subsets $S \subset \{0, 1, \ldots, n\}$ ...

- have more differences than sums, i.e., $|S+S| < |S-S|$?
- have the equal number differences and sums, i.e., $|S+S| = |S-S|$?
- are missing exactly s sums and d differences, i.e., $|S+S| = 2n+1-s$, $|S-S| = 2n+1-d$?
- have exactly x more sums than differences, i.e. $|S+S| - |S-S| = x$?

Beyond MSTD sets

What proportion of subsets $S \subset \{0, 1, \dots, n\}$ ...

- have more differences than sums, i.e., $|S+S| < |S-S|$?
- have the equal number differences and sums, i.e., $|S+S| = |S-S|$?
- are missing exactly s sums and d differences, i.e., $|S+S| = 2n+1-s$, $|S-S| = 2n+1-d$?
- have exactly x more sums than differences, i.e. $|S+S| - |S-S| = x$?

Our method can be modified to give similar answers to each of these questions.

The Number of Missing Sums and Differences

For a subset $S \subset \{0, 1, \dots, n\}$, let

$$\lambda(S) = (\underbrace{2n + 1 - |S + S|}_{\#\text{missing sums}}, \underbrace{2n + 1 - |S - S|}_{\#\text{missing differences}})$$

The Number of Missing Sums and Differences

For a subset $S \subset \{0, 1, \ldots, n\}$, let

$$\lambda(S) = (\underbrace{2n + 1 - |S + S|}_{\#\text{missing sums}}, \underbrace{2n + 1 - |S - S|}_{\#\text{missing differences}})$$

Let

$$\Lambda \subset \mathbb{Z}_{\geq 0} \times \mathbb{Z}_{\geq 0}.$$

Assume that Λ has at least one element (s, d) with d even. We are interested in

$$\{S \subset \{0, 1, \ldots, n\} : \lambda(S) \in \Lambda\}$$

E.g., $\Lambda = \{(s, d) : s < d\}$ gives MSTD sets

General Results

Let $\Lambda \subset \mathbb{Z}_{\geq 0} \times \mathbb{Z}_{\geq 0}$ contain at least one element (s, d) with d even.
$\lambda(S) = (2n + 1 - |S + S|, 2n + 1 - |S - S|)$.
Let S be a uniform random subset of $\{0, 1, \ldots, n\}$. As $n \to \infty$,

- $\mathbb{P}(\lambda(S) \in \Lambda)$ approaches some positive limit
- We have a deterministic algorithm for computing this limit up to arbitrary precision
- if $\alpha_n \in \mathbb{Z}$ satisfying $0 < \alpha_n < n/2$ and $\alpha_n \to \infty$, E an event that depends only on $S \cap [\alpha_n + 1, n - \alpha_n - 1]$, and F an event that depends only on $S \cap ([0, \alpha_n] \cup [n - \alpha_n, n])$, then

$$\left|\mathbb{P}(E \cap F \mid \lambda(S) \in \Lambda) - \mathbb{P}(E)\mathbb{P}(F \mid \lambda(S) \in \Lambda)\right| = O\left((3/4)^{\alpha_n}\right)$$

General Results

Let $\Lambda \subset \mathbb{Z}_{\geq 0} \times \mathbb{Z}_{\geq 0}$ contain at least one element (s, d) with d even.
$\lambda(S) = (2n + 1 - |S + S|, 2n + 1 - |S - S|)$.
Let S be a uniform random subset of $\{0, 1, \ldots, n\}$. As $n \to \infty$,

- $\mathbb{P}(\lambda(S) \in \Lambda)$ approaches some positive limit
- We have a deterministic algorithm for computing this limit up to arbitrary precision
- if $\alpha_n \in \mathbb{Z}$ satisfying $0 < \alpha_n < n/2$ and $\alpha_n \to \infty$, E an event that depends only on $S \cap [\alpha_n + 1, n - \alpha_n - 1]$, and F an event that depends only on $S \cap ([0, \alpha_n] \cup [n - \alpha_n, n])$, then

$$|\mathbb{P}(E \cap F \mid \lambda(S) \in \Lambda) - \mathbb{P}(E)\mathbb{P}(F \mid \lambda(S) \in \Lambda)| = O\left((3/4)^{\alpha_n}\right)$$

Acknowledgments: This work was done at the Joe Gallian's REU at University of Minnesota at Duluth, with funding from NSF, DoD, NSA, and the MIT Math Department.