MSTD Subsets and Properties of Divots in the Distribution of Missing Sums

Steven J Miller, Williams College
sjm1@williams.edu
Victor Xu, Carnegie Mellon University
vzx@andrew.cmu.edu
Xiaorong Zhang, Carnegie Mellon University
xiaoronz@andrew.cmu.edu

Combinatorial and Additive Number Theory, May 26, 2016

http://web.williams.edu/Mathematics/sjmiller/public_html/

MSTD Subsets

Statement

A finite set of integers, $|A|$ its size. Form

- Sumset: $A + A = \{a_i + a_j : a_j, a_j \in A\}$.
- Difference set: $A - A = \{a_i - a_j : a_j, a_j \in A\}$.

Definition

We say A is difference dominated if $|A - A| > |A + A|$, balanced if $|A - A| = |A + A|$ and sum dominated (or an MSTD set) if $|A + A| > |A - A|$.

Questions

Expect generic set to be difference dominated:

- addition is commutative, subtraction isn't:
- Generic pair (x, y) gives 1 sum, 2 differences.

Questions

Expect generic set to be difference dominated:

- addition is commutative, subtraction isn't:
- Generic pair (x, y) gives 1 sum, 2 differences.

Questions

- Do there exist sum-dominated sets?
- If yes, how many?

Examples

- Conway: $\{0, 2, 3, 4, 7, 11, 12, 14\}$.
- Computer search of random subsets of $\{1, \dots, 100\}$: $\{2, 6, 7, 9, 13, 14, 16, 18, 19, 22, 23, 25, 30, 31, 33, 37, 39, 41, 42, 45, 46, 47, 48, 49, 51, 52, 54, 57, 58, 59, 61, 64, 65, 66, 67, 68, 72, 73, 74, 75, 81, 83, 84, 87, 88, 91, 93, 94, 95, 98, 100\}$.
- Many infinite families (Hegarty, Miller - Orosz - Scheinerman, Nathanson, ...).
- If *A* chosen uniformly at random positive probability it is MSTD (Martin-O'Bryant).

Subsets and MSTD Sets: General Results

Theorem

Let $A := \{a_k\}_{k=1}^{\infty}$ be a sequence of natural numbers. If there exists a positive integer r such that

1. $a_k > a_{k-1} + a_{k-r}$ *for all* $k \geq r + 1$, *and*
2. a_k *does not contain any MSTD set S with* $|S| \leq 2r + 1$,

then A contains no MSTD set.

Immediate corollary: No subset of the Fibonacci numbers is an MSTD set.

Proof: MSTD set must have at least 8 elements, show gain more differences than sums as add elements.

Subsets and MSTD Sets: Preliminaries

Hardy-Littlewood Conjecture

Let $b_1, b_2, \dots, b_m$ be m distinct integers,
$P(x; b_1, b_2, \dots, b_m)$ the number of integers at most x st
$\{n + b_1, n + b_2, \dots, n + b_m\}$ consists wholly of primes,
v the number of distinct residues of $b_1, b_2, \dots b_m$ mod p,

$$G(b_1, b_2, \dots, b_m) \;=\; \prod_{p \geq 2} \left(\left(\frac{p}{p-1} \right)^{m-1} \frac{p-v}{p-1} \right).$$

Then as $x \to \infty$

$$P(x) \;\sim\; G(b_1, b_2, \dots, b_m) \int_2^x \frac{du}{(\log u)^m}.$$

Subsets and MSTD Sets: Primes

Theorem

The Hardy-Littlewood conjecture implies there are infinitely many MSTD subsets of the primes.

Proof (sketch):

- Smallest MSTD set is $S = \{0, 2, 3, 4, 7, 11, 12, 14\}$.
- $\{p, p+2s, p+3s, p+4s, p+7s, p+11s, p+12s, p+14s\}$ is an MSTD set for all positive integers p, s.
- Set $s = 30$. Hardy-Littlewood Conjecture implies $\{p, p+60, p+90, p+120, p+210, p+330, p+360, p+420\}$ are all primes for infinitely many prime p.

Distribution of Divots:
Introduction and Background

Definitions

- Let S be a subset of $I_n = \{0, \ldots, n-1\}$, let $S + S = \{x + y : x, y \in S\}$.

Definitions

- Let S be a subset of $I_n = \{0, \ldots, n-1\}$, let $S + S = \{x + y : x, y \in S\}$.
- For each $i \in I_n$, choose independently that $i \notin S$ with probability q.

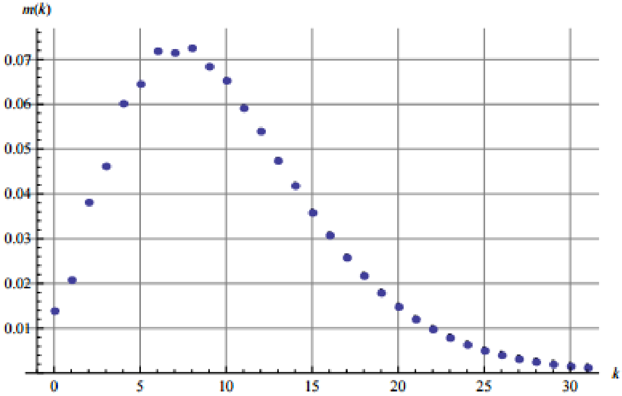
Definitions

- Let S be a subset of $I_n = \{0, \dots, n-1\}$, let $S + S = \{x + y : x, y \in S\}$.
- For each $i \in I_n$, choose independently that $i \notin S$ with probability q.
- What's the distribution of $|S + S|$? Instead, we can look at $2n - 1 - |S + S|$.

Definitions

- Let S be a subset of $I_n = \{0, \dots, n-1\}$, let $S + S = \{x + y : x, y \in S\}$.
- For each $i \in I_n$, choose independently that $i \notin S$ with probability q.
- What's the distribution of $|S+S|$? Instead, we can look at $2n - 1 - |S + S|$.
- Let $M = I_n \setminus S$.
- Let $T = (I_n + I_n) \setminus (S + S)$.

Previous Results



Distribution of Missing sums for $q = .5$.

Previous Results

Lazarev, Miller, O'Bryant (2012)

For $q = .5$, let $m(n)$ denote the probability that $|T| = n$, then $m(7) < m(6) < m(8)$.

Previous Results

Lazarev, Miller, O'Bryant (2012)

For $q = .5$, let $m(n)$ denote the probability that $|T| = n$, then $m(7) < m(6) < m(8)$.

- Used massive computation of 2^{28} sets to prove result.
- The "divot" in the probabilities is interesting.
- Recall $T = (I_n + I_n) \setminus (S + S)$.

Problem

- What about for different q, $q \neq .5$?

Problem

- What about for different q, $q \neq .5$?
- If q is close to 0, then S will have many elements and $|T|$ will usually be small.

Problem

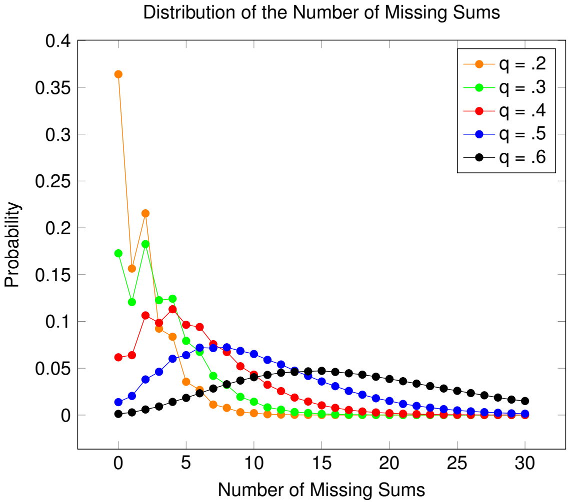
- What about for different q, $q \neq .5$?
- If q is close to 0, then S will have many elements and $|T|$ will usually be small.
- This seems easier than the general case.

Problem

- What about for different q, $q \neq .5$?
- If q is close to 0, then S will have many elements and $|T|$ will usually be small.
- This seems easier than the general case.
- Are there any divots for q close to 0?

Behavior of the Divot

Distribution of $|T|$



Computer simulation of $1,000,000$ subsets of $\{0, 1, \ldots, 255\}$.

Observations and Problems

- We show existence of a divot at 1 for $q < .034$; this result is very loose.

Observations and Problems

- We show existence of a divot at 1 for $q < .034$; this result is very loose.
- How does the position of the divot depend on q?

Observations and Problems

- We show existence of a divot at 1 for $q < .034$; this result is very loose.
- How does the position of the divot depend on q?
- Also, at $q = .3$ there appear two divots at 1 and 3; for what values of q are there more than one divot?

Observations and Problems

- We show existence of a divot at 1 for $q < .034$; this result is very loose.
- How does the position of the divot depend on q?
- Also, at $q = .3$ there appear two divots at 1 and 3; for what values of q are there more than one divot?
- Lastly, for $q = .6$ the divot disappears. Where is this phase transition point where the divot disappears?

Divot for Small q

The Divot for Small q

- There is a divot at 1 when q is small ($< .034$), for $n > 20$.

The Divot for Small q

- There is a divot at 1 when q is small ($< .034$), for $n > 20$.
- $T = \{0, 1, \dots, 2n - 2\} \setminus (S + S)$ is the set of missing sums.
- To show this, we can split up $T = B + C + E$ as follows:

The Divot for Small q

- There is a divot at 1 when q is small ($< .034$), for $n > 20$.

- $T = \{0, 1, \dots, 2n - 2\} \setminus (S + S)$ is the set of missing sums.
- To show this, we can split up $T = B + C + E$ as follows:

- $B = T \cap \{0, 1, \dots, \lfloor n/2 \rfloor - 1\}$.
- $C = T \cap \{\lfloor n/2 \rfloor, n + 1, \dots, 2n - 3 - \lfloor n/2 \rfloor\}$.
- $E = T \cap \{2n - 2 - \lfloor n/2 \rfloor, 2n - 1 - \lfloor n/2 \rfloor, \dots, 2n - 2\}$.

Intuition

Largest Sum Missing	$\|B\| = 1$	$\|B\| = 2$
0		
1	$\{1\}$	$\{0\}$
2		$\{1, 2\}$
3	$\{2, 3\}$	$\{1, 3\}$
4		$\{2, 3, 4\}$
5	$\{2, 4, 5\}, \{3, 4, 5\}$	
...	...	...

Intuition

Largest Sum Missing	$\|B\|$ = 1	$\|B\|$ = 2
0		
1	$\{1\}$	$\{0\}$
2		$\{1,2\}$
3	$\{2,3\}$	$\{1,3\}$
4		$\{2,3,4\}$
5	$\{2,4,5\},\{3,4,5\}$	
...	...	...

- Recall q is the probability that any element $i \notin S$.
- $\mathbb{P}[|B| = 1] \sim q + q^2 + O(q^3)$.
- $\mathbb{P}[|B| = 2] \sim q + 2q^2 + O(q^3)$.

Finding Bounds

Example: $\mathbb{P}\left(6 \in T\right)$

Sums	Probability
0+6	$\mathbb{P}\left(0 \notin S \vee 6 \notin S\right) < \mathbb{P}\left(0 \notin S\right) + \mathbb{P}\left(6 \notin S\right) < 2q$
1+5	$\mathbb{P}\left(1 \notin S \vee 5 \notin S\right) < \mathbb{P}\left(1 \notin S\right) + \mathbb{P}\left(5 \notin S\right) < 2q$
2+4	$\mathbb{P}\left(2 \notin S \vee 4 \notin S\right) < \mathbb{P}\left(2 \notin S\right) + \mathbb{P}\left(4 \notin S\right) < 2q$
3+3	$\mathbb{P}\left(3 \notin S \vee 3 \notin S\right) < \mathbb{P}\left(3 \notin S\right) + \mathbb{P}\left(3 \notin S\right) < 2q$

Finding Bounds

Example: $\mathbb{P}(6 \in T)$

Sums	Probability
0+6	$\mathbb{P}(0 \notin S \vee 6 \notin S) < \mathbb{P}(0 \notin S) + \mathbb{P}(6 \notin S) < 2q$
1+5	$\mathbb{P}(1 \notin S \vee 5 \notin S) < \mathbb{P}(1 \notin S) + \mathbb{P}(5 \notin S) < 2q$
2+4	$\mathbb{P}(2 \notin S \vee 4 \notin S) < \mathbb{P}(2 \notin S) + \mathbb{P}(4 \notin S) < 2q$
3+3	$\mathbb{P}(3 \notin S \vee 3 \notin S) < \mathbb{P}(3 \notin S) + \mathbb{P}(3 \notin S) < 2q$

- $\mathbb{P}[i \in T] < (2q)^{\lfloor \frac{i}{2} \rfloor + 1}$.
- For $k \leq n$,

$$\sum_{i=k}^{n} \mathbb{P}[i \in T] < \sum_{i=k}^{n} (2q)^{\lfloor \frac{i}{2} \rfloor + 1} < \frac{2(2q)^{\lfloor \frac{i}{2} \rfloor + 1}}{1 - 2q}.$$

Explanation

- We show that $|C| = 0$ is very likely.

Explanation

- We show that $|C| = 0$ is very likely.
- $|B|$ and $|E|$ have the same distribution.

Explanation

- We show that $|C| = 0$ is very likely.
- $|B|$ and $|E|$ have the same distribution.
- Then, we find bounds on $\mathbb{P}(|B| = 1)$ and $\mathbb{P}(|B| = 2)$ in terms of q.

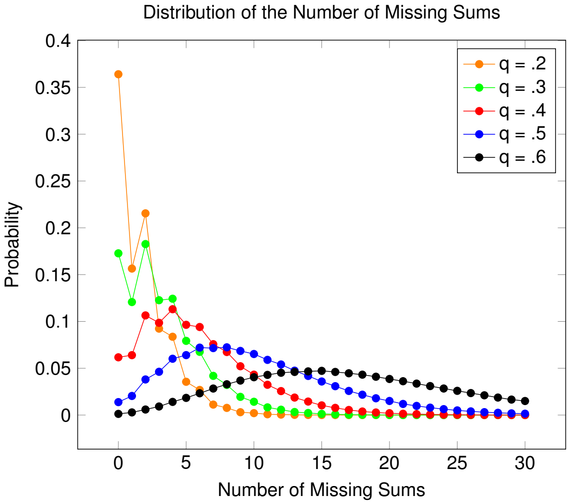
Explanation

- We show that $|C| = 0$ is very likely.
- $|B|$ and $|E|$ have the same distribution.
- Then, we find bounds on $\mathbb{P}(|B| = 1)$ and $\mathbb{P}(|B| = 2)$ in terms of q.
- Examining the cases for $|T| = 1$ and $|T| = 2$ leads to

$$\mathbb{P}(|T| = 0) > \mathbb{P}(|T| = 1) < \mathbb{P}(|T| = 2)$$

for $q < .034$ and $n > 20$.

Distribution of $|T|$



Computer simulation of $1,000,000$ subsets of $\{0, 1, \ldots, 255\}$.

Questions?