

# When Almost All Sets Are Difference Dominated

Steven J Miller  
Williams College

Steven.J.Miller@williams.edu  
<http://www.williams.edu/go/math/sjmillier/>

Workshop on Combinatorial and Additive Number Theory  
(CANT 2009)  
CUNY Graduate Center, New York, May 2009

## Summary

- History of the problem.
- Examples.
- Main results and proofs.
- Describe open problems.

This is joint work with Peter Hegarty, Brooke Orosz and Dan Scheinerman.

## Introduction

## Statement

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

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

## Statement

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

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

### 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

## Examples

- Conway:  $\{0, 2, 3, 4, 7, 11, 12, 14\}$ .
- Marica (1969):  $\{0, 1, 2, 4, 7, 8, 12, 14, 15\}$ .
- Freiman and Pigarev (1973):  $\{0, 1, 2, 4, 5, 9, 12, 13, 14, 16, 17, 21, 24, 25, 26, 28, 29\}$ .
- Computer search of random subsets of  $\{1, \dots, 100\}$ :  
 $\{2, 6, 7, 9, 13, 14, 16, 18, 19, 22, 23, 25, 30, 31, 33, 37, 39, 41, 42, 45, 46, 47, 48, 49, 51, 52, 54, 57, 58, 59, 61, 64, 65, 66, 67, 68, 72, 73, 74, 75, 81, 83, 84, 87, 88, 91, 93, 94, 95, 98, 100\}$ .
- Recently infinite families (Hegarty, Nathanson).

## Infinite Families

### Key observation

If  $A$  is an arithmetic progression,  $|A + A| = |A - A|$ .

## Infinite Families

### Key observation

If  $A$  is an arithmetic progression,  $|A + A| = |A - A|$ .

Proof:

- WLOG,  $A = \{0, 1, \dots, n\}$  as  $A \rightarrow \alpha A + \beta$  doesn't change  $|A + A|$ ,  $|A - A|$ .

## Infinite Families

### Key observation

If  $A$  is an arithmetic progression,  $|A + A| = |A - A|$ .

Proof:

- WLOG,  $A = \{0, 1, \dots, n\}$  as  $A \rightarrow \alpha A + \beta$  doesn't change  $|A + A|$ ,  $|A - A|$ .
- $A + A = \{0, \dots, 2n\}$ ,  $A - A = \{-n, \dots, n\}$ , both of size  $2n + 1$ . □

## Previous Constructions

Most constructions perturb an arithmetic progression.

Example:

- MSTD set  $A = \{0, 2, 3, 4, 7, 11, 12, 14\}$ .
- $A = \{0, 2\} \cup \{3, 7, 11\} \cup (14 - \{0, 2\}) \cup \{4\}$ .

## Example (Nathanson)

### Theorem

$m, d, k \in \mathbb{N}$  with  $m \geq 4$ ,  $1 \leq d \leq m - 1$ ,  $d \neq m/2$ ,  $k \geq 3$  if  $d < m/2$  else  $k \geq 4$ . Let

- $B = [0, m - 1] \setminus \{d\}$ .
- $L = \{m - d, 2m - d, \dots, km - d\}$ .
- $a^* = (k + 1)m - 2d$ .
- $A^* = B \cup L \cup (a^* - B)$ .
- $A = A^* \cup \{m\}$ .

*Then  $A$  is an MSTD set.*

## New Construction: Notation

- $[a, b] = \{k \in \mathbb{Z} : a \leq k \leq b\}$ .
- $A$  is a  **$P_n$ -set** if its sumset and its difference set contain all but the first and last  $n$  possible elements (and of course it may or may not contain some of these fringe elements).

## New Construction

### Theorem (Miller-Scheinerman '09)

- $A = L \cup R$  be a  $P_n$ , MSTD set where  $L \subset [1, n]$ ,  $R \subset [n + 1, 2n]$ , and  $1, 2n \in A$ .
- Fix a  $k \geq n$  and let  $m$  be arbitrary.
- $M$  any subset of  $[n + k + 1, n + k + m]$  st no run of more than  $k$  missing elements. Assume  $n + k + 1 \notin M$ .
- Set  $A(M) = L \cup O_1 \cup M \cup O_2 \cup R'$ , where  $O_1 = [n + 1, n + k]$ ,  $O_2 = [n + k + m + 1, n + 2k + m]$ , and  $R' = R + 2k + m$ .

Then  $A(M)$  is an MSTD set, and  $\exists C > 0$  st the percentage of subsets of  $\{0, \dots, r\}$  that are in this family (and thus are MSTD sets) is at least  $C/r^4$ .

## Generalization: Miller-Orosz-Scheinerman

Can we find  $A$  so that:

$$|\epsilon_1 A + \dots + \epsilon_n A| > |\tilde{\epsilon}_1 A + \dots + \tilde{\epsilon}_n A|, \quad \epsilon_i, \tilde{\epsilon}_i \in \{-1, 1\}.$$

Consider the generalized sumset

$$f_{j_1, j_2}(A) = A + A + \dots + A - A - A - \dots - A,$$

where there are  $j_1$  pluses and  $j_2$  minuses, and set  $j = j_1 + j_2$ .

### $P_n^j$ -set

Let  $A \subset [1, k]$  with  $1, k, \in A$ . We say  $A$  is a  $P_n^j$ -set if any  $f_{j_1, j_2}(A)$  contains all but the first  $n$  and last  $n$  possible elements. (Note that a  $P_n^2$ -set is the same as what we called a  $P_n$ -set earlier.)

## Generalization: Miller-Orosz-Scheinerman

### Conjecture (MOS)

For any  $f_{j_1, j_2}$  and  $f_{j'_1, j'_2}$ , there exists a finite set of integers  $A$  which is (1) a  $P_n^j$ -set; (2)  $A \subset [1, 2n]$  and  $1, 2n \in A$ ; and (3)  $|f_{j_1, j_2}(A)| > |f_{j'_1, j'_2}(A)|$ .

- Problem is finding an  $A$  with  $|f_{j_1, j_2}(A)| > |f_{j'_1, j'_2}(A)|$ ; once we find such a set, we can mirror previous construction and construct infinitely many.
- Theorem: The conjecture is true for  $j \in \{2, 3\}$ .

## Proof of Generalization

- Needed input set for  $j = 3$ :  $A = \{1, 2, 5, 6, 16, 19, 22, 26, 32, 34, 35, 39, 43, 48, 49, 50\}$ .  
 Found by taking elements in  $\{2, \dots, 49\}$  to be in  $A$  with probability  $1/3$ ; it took about 300000 sets to find the first one satisfying our conditions. To be a  $P_{25}^3$ -set we need to have  $A + A + A \supset [n + 3, 6n - n] = [28, 125]$  and  $A + A - A \supset [-n + 2, 3n - 1] = [-23, 74]$ . Have  $A + A + A = [3, 150]$  (all possible elements), while  $A + A - A = [-48, 99] \setminus \{-34\}$  (i.e., all but -34). Thus  $A$  is a  $P_{25}^3$ -set satisfying  $|A + A + A| > |A + A - A|$ , and have the needed example.
- Could also take  $A = \{1, 2, 3, 4, 8, 12, 18, 22, 23, 25, 26, 29, 30, 31, 32, 34, 45, 46, 49, 50\}$ .

## Results

## Probability Review

$X$  random variable with density  $f(x)$  means

- $f(x) \geq 0$ ;
- $\int_{-\infty}^{\infty} f(x) = 1$ ;
- $\text{Prob}(X \in [a, b]) = \int_a^b f(x) dx$ .

Key quantities:

- Expected (Average) Value:  $\mathbb{E}[X] = \int xf(x) dx$ .
- Variance:  $\sigma^2 = \int (x - \mathbb{E}[X])^2 f(x) dx$ .

## Binomial model

### Binomial model, parameter $p(n)$

Each  $k \in \{0, \dots, n\}$  is in  $A$  with probability  $p(n)$ .

Consider uniform model ( $p(n) = 1/2$ ):

- Let  $A \in \{0, \dots, n\}$ . Most elements in  $\{0, \dots, 2n\}$  in  $A + A$  and in  $\{-n, \dots, n\}$  in  $A - A$ .
- $\mathbb{E}[|A + A|] = 2n - 11$ ,  $\mathbb{E}[|A - A|] = 2n - 7$ .

## Martin and O'Bryant '06

**Theorem**

*Let  $A$  be chosen from  $\{0, \dots, N\}$  according to the binomial model with constant parameter  $p$  (thus  $k \in A$  with probability  $p$ ). At least  $k_{\text{SD};p} 2^{N+1}$  subsets are sum dominated.*

## Martin and O'Bryant '06

**Theorem**

*Let  $A$  be chosen from  $\{0, \dots, N\}$  according to the binomial model with constant parameter  $p$  (thus  $k \in A$  with probability  $p$ ). At least  $k_{\text{SD};p} 2^{N+1}$  subsets are sum dominated.*

- $k_{\text{SD};1/2} \geq 10^{-7}$ , expect about  $10^{-3}$ .

## Martin and O'Bryant '06

### Theorem

Let  $A$  be chosen from  $\{0, \dots, N\}$  according to the binomial model with constant parameter  $p$  (thus  $k \in A$  with probability  $p$ ). At least  $k_{\text{SD};p} 2^{N+1}$  subsets are sum dominated.

- $k_{\text{SD};1/2} \geq 10^{-7}$ , expect about  $10^{-3}$ .
- Proof ( $p = 1/2$ ): Generically  $|A| = \frac{N}{2} + O(\sqrt{N})$ .
  - ◇ about  $\frac{N}{4} - \frac{|N-k|}{4}$  ways write  $k \in A + A$ .
  - ◇ about  $\frac{N}{4} - \frac{|k|}{4}$  ways write  $k \in A - A$ .
  - ◇ Almost all numbers that can be in  $A \pm A$  are.
  - ◇ Win by controlling fringes.

## Notation

- $X \sim f(N)$  means  $\forall \epsilon_1, \epsilon_2 > 0, \exists N_{\epsilon_1, \epsilon_2}$  st  $\forall N \geq N_{\epsilon_1, \epsilon_2}$   
 $\text{Prob}(X \notin [(1 - \epsilon_1)f(N), (1 + \epsilon_1)f(N)]) < \epsilon_2.$

## Notation

- $X \sim f(N)$  means  $\forall \epsilon_1, \epsilon_2 > 0, \exists N_{\epsilon_1, \epsilon_2}$  st  $\forall N \geq N_{\epsilon_1, \epsilon_2}$   
 $\text{Prob}(X \notin [(1 - \epsilon_1)f(N), (1 + \epsilon_1)f(N)]) < \epsilon_2.$
- $\mathcal{S} = |A + A|, \mathcal{D} = |A - A|,$   
 $\mathcal{S}^c = 2N + 1 - \mathcal{S}, \mathcal{D}^c = 2N + 1 - \mathcal{D}.$

## Notation

- $X \sim f(N)$  means  $\forall \epsilon_1, \epsilon_2 > 0, \exists N_{\epsilon_1, \epsilon_2}$  st  $\forall N \geq N_{\epsilon_1, \epsilon_2}$

$$\text{Prob}(X \notin [(1 - \epsilon_1)f(N), (1 + \epsilon_1)f(N)]) < \epsilon_2.$$

- $S = |A + A|, D = |A - A|,$   
 $S^c = 2N + 1 - S, D^c = 2N + 1 - D.$

New model: Binomial with parameter  $p(N)$ :

- $1/N = o(p(N))$  and  $p(N) = o(1)$ ;
- $\text{Prob}(k \in A) = p(N).$

### Conjecture (Martin-O'Bryant)

As  $N \rightarrow \infty, A$  is a.s. difference dominated.

## Main Result

### Theorem (Hegarty-Miller)

$p(N)$  as above,  $g(x) = 2 \frac{e^{-x} - (1-x)}{x}$ .

- $p(N) = o(N^{-1/2})$ :  $\mathcal{D} \sim 2\mathcal{S} \sim (Np(N))^2$ ;
- $p(N) = cN^{-1/2}$ :  $\mathcal{D} \sim g(c^2)N$ ,  $\mathcal{S} \sim g\left(\frac{c^2}{2}\right)N$   
( $c \rightarrow 0$ ,  $\mathcal{D}/\mathcal{S} \rightarrow 2$ ;  $c \rightarrow \infty$ ,  $\mathcal{D}/\mathcal{S} \rightarrow 1$ );
- $N^{-1/2} = o(p(N))$ :  $\mathcal{S}^c \sim 2\mathcal{D}^c \sim 4/p(N)^2$ .

Can generalize to binary linear forms, still have **critical threshold**.

## Inputs

Key input: recent strong concentration results of Kim and Vu (Applications: combinatorial number theory, random graphs, ...).

## Inputs

Key input: recent strong concentration results of Kim and Vu (Applications: combinatorial number theory, random graphs, ...).

**Example (Chernoff):**  $t_i$  iid binary random variables,  
 $Y = \sum_{i=1}^n t_i$ , then

$$\forall \lambda > 0 : \text{Prob} \left( |Y - \mathbb{E}[Y]| \geq \sqrt{\lambda n} \right) \leq 2e^{-\lambda/2}.$$

## Inputs

Key input: recent strong concentration results of Kim and Vu (Applications: combinatorial number theory, random graphs, ...).

**Example (Chernoff):**  $t_i$  iid binary random variables,  $Y = \sum_{i=1}^n t_i$ , then

$$\forall \lambda > 0 : \text{Prob} \left( |Y - \mathbb{E}[Y]| \geq \sqrt{\lambda n} \right) \leq 2e^{-\lambda/2}.$$

Need to allow dependent random variables.

## Inputs

Key input: recent strong concentration results of Kim and Vu (Applications: combinatorial number theory, random graphs, ...).

**Example (Chernoff):**  $t_i$  iid binary random variables,  $Y = \sum_{i=1}^n t_i$ , then

$$\forall \lambda > 0 : \text{Prob} \left( |Y - \mathbb{E}[Y]| \geq \sqrt{\lambda n} \right) \leq 2e^{-\lambda/2}.$$

Need to allow dependent random variables.

Sketch of proofs:  $\mathcal{X} \in \{\mathcal{S}, \mathcal{D}, \mathcal{S}^c, \mathcal{D}^c\}$ .

- 1 Prove  $\mathbb{E}[\mathcal{X}]$  behaves asymptotically as claimed;
- 2 Prove  $\mathcal{X}$  is strongly concentrated about mean.

# Proofs

## Setup

Note: only need strong concentration for  $N^{-1/2} = o(p(N))$ .

## Setup

Note: only need strong concentration for  $N^{-1/2} = o(p(N))$ .

Will assume  $p(N) = o(N^{-1/2})$  as proofs are elementary (i.e., Chebyshev:  $\text{Prob}(|Y - \mathbb{E}[Y]| \geq k\sigma_Y) \leq 1/k^2$ ).

## Setup

Note: only need strong concentration for  $N^{-1/2} = o(p(N))$ .

Will assume  $p(N) = o(N^{-1/2})$  as proofs are elementary (i.e., Chebyshev:  $\text{Prob}(|Y - \mathbb{E}[Y]| \geq k\sigma_Y) \leq 1/k^2$ ).

For convenience let  $p(N) = N^{-\delta}$ ,  $\delta \in (1/2, 1)$ .

IID binary indicator variables:

$$X_{n;N} = \begin{cases} 1 & \text{with probability } N^{-\delta} \\ 0 & \text{with probability } 1 - N^{-\delta}. \end{cases}$$

$$X = \sum_{i=1}^N X_{i;N}, \quad \mathbb{E}[X] = N^{1-\delta}.$$

## Proof

## Lemma

$$P_1(N) = 4N^{-(1-\delta)},$$

$$\mathcal{O} = \#\{(m, n) : m < n \in \{1, \dots, N\} \cap A\}.$$

With probability at least  $1 - P_1(N)$  have

$$\textcircled{1} \quad X \in \left[ \frac{1}{2}N^{1-\delta}, \frac{3}{2}N^{1-\delta} \right].$$

$$\textcircled{2} \quad \frac{\frac{1}{2}N^{1-\delta}(\frac{1}{2}N^{1-\delta}-1)}{2} \leq \mathcal{O} \leq \frac{\frac{3}{2}N^{1-\delta}(\frac{3}{2}N^{1-\delta}-1)}{2}.$$

## Proof

## Lemma

$$P_1(N) = 4N^{-(1-\delta)},$$

$$\mathcal{O} = \#\{(m, n) : m < n \in \{1, \dots, N\} \cap A\}.$$

With probability at least  $1 - P_1(N)$  have

$$\textcircled{1} \quad X \in \left[ \frac{1}{2}N^{1-\delta}, \frac{3}{2}N^{1-\delta} \right].$$

$$\textcircled{2} \quad \frac{\frac{1}{2}N^{1-\delta}(\frac{1}{2}N^{1-\delta}-1)}{2} \leq \mathcal{O} \leq \frac{\frac{3}{2}N^{1-\delta}(\frac{3}{2}N^{1-\delta}-1)}{2}.$$

Proof:

- (1) is Chebyshev:  $\text{Var}(X) = N\text{Var}(X_{n;N}) \leq N^{1-\delta}$ .
- (2) follows from (1) and  $\binom{r}{2}$  ways to choose 2 from  $r$ .

## Concentration

### Lemma

- $f(\delta) = \min\left(\frac{1}{2}, \frac{3\delta-1}{2}\right)$ ,  $g(\delta)$  any function st  $0 < g(\delta) < f(\delta)$ .
- $p(N) = N^{-\delta}$ ,  $\delta \in (1/2, 1)$ ,  $P_1(N) = 4N^{-(1-\delta)}$ ,  
 $P_2(N) = CN^{-(f(\delta)-g(\delta))}$ .

With probability at least  $1 - P_1(N) - P_2(N)$  have  $\mathcal{D}/\mathcal{S} = 2 + O(N^{-g(\delta)})$ .

## Concentration

### Lemma

- $f(\delta) = \min\left(\frac{1}{2}, \frac{3\delta-1}{2}\right)$ ,  $g(\delta)$  any function st  $0 < g(\delta) < f(\delta)$ .
- $p(N) = N^{-\delta}$ ,  $\delta \in (1/2, 1)$ ,  $P_1(N) = 4N^{-(1-\delta)}$ ,  
 $P_2(N) = CN^{-(f(\delta)-g(\delta))}$ .

With probability at least  $1 - P_1(N) - P_2(N)$  have  $\mathcal{D}/\mathcal{S} = 2 + O(N^{-g(\delta)})$ .

Proof: Show  $\mathcal{D} \sim 2\mathcal{O} + O(N^{3-4\delta})$ ,  $\mathcal{S} \sim \mathcal{O} + O(N^{3-4\delta})$ .

As  $\mathcal{O}$  is of size  $N^{2-2\delta}$  with high probability, need  $2 - 2\delta > 3 - 4\delta$  or  $\delta > 1/2$ .

## Analysis of $\mathcal{D}$

Contribution from ‘diagonal’ terms lower order, ignore.

## Analysis of $\mathcal{D}$

Contribution from 'diagonal' terms lower order, ignore.

Difficulty:  $(m, n)$  and  $(m', n')$  could yield same differences.

## Analysis of $\mathcal{D}$

Contribution from ‘diagonal’ terms lower order, ignore.

Difficulty:  $(m, n)$  and  $(m', n')$  could yield same differences.

**Notation:**  $m < n, m' < n', m \leq m'$ ,

$$Y_{m,n,m',n'} = \begin{cases} 1 & \text{if } n - m = n' - m' \\ 0 & \text{otherwise.} \end{cases}$$

## Analysis of $\mathcal{D}$

Contribution from 'diagonal' terms lower order, ignore.

Difficulty:  $(m, n)$  and  $(m', n')$  could yield same differences.

**Notation:**  $m < n, m' < n', m \leq m'$ ,

$$Y_{m,n,m',n'} = \begin{cases} 1 & \text{if } n - m = n' - m' \\ 0 & \text{otherwise.} \end{cases}$$

$\mathbb{E}[Y] \leq N^3 \cdot N^{-4\delta} + N^2 \cdot N^{-3\delta} \leq 2N^{3-4\delta}$ . As  $\delta > 1/2$ ,  
 expected number bad pairs  $\lll |\mathcal{O}|$ .

## Analysis of $\mathcal{D}$

Contribution from 'diagonal' terms lower order, ignore.

Difficulty:  $(m, n)$  and  $(m', n')$  could yield same differences.

**Notation:**  $m < n, m' < n', m \leq m'$ ,

$$Y_{m,n,m',n'} = \begin{cases} 1 & \text{if } n - m = n' - m' \\ 0 & \text{otherwise.} \end{cases}$$

$\mathbb{E}[Y] \leq N^3 \cdot N^{-4\delta} + N^2 \cdot N^{-3\delta} \leq 2N^{3-4\delta}$ . As  $\delta > 1/2$ ,  
 expected number bad pairs  $\lll |\mathcal{O}|$ .

**Claim:**  $\sigma_Y \leq N^{r(\delta)}$  with  $r(\delta) = \frac{1}{2} \max(3 - 4\delta, 5 - 7\delta)$ . This  
 and Chebyshev conclude proof of theorem.

## Proof of claim

Cannot use CLT as  $Y_{m,n,m',n'}$  are not independent.

## Proof of claim

Cannot use CLT as  $Y_{m,n,m',n'}$  are not independent.

Use  $\text{Var}(U + V) \leq 2\text{Var}(U) + 2\text{Var}(V)$ .

## Proof of claim

Cannot use CLT as  $Y_{m,n,m',n'}$  are not independent.

Use  $\text{Var}(U + V) \leq 2\text{Var}(U) + 2\text{Var}(V)$ .

Write

$$\sum Y_{m,n,m',n'} = \sum U_{m,n,m',n'} + \sum V_{m,n,n'}$$

with all indices distinct (at most one in common, if so must be  $n = m'$ ).

$$\text{Var}(U) = \sum \text{Var}(U_{m,n,m',n'}) + 2 \sum_{\substack{(m,n,m',n') \neq \\ (\tilde{m}, \tilde{n}, \tilde{m}', \tilde{n}')}} \text{CoVar}(U_{m,n,m',n'}, U_{\tilde{m}, \tilde{n}, \tilde{m}', \tilde{n}'})$$

## Analyzing $\text{Var}(U_{m,n,m',n'})$

At most  $N^3$  tuples.

Each has variance  $N^{-4\delta} - N^{-8\delta} \leq N^{-4\delta}$ .

Thus  $\sum \text{Var}(U_{m,n,m',n'}) \leq N^{3-4\delta}$ .

## Analyzing $\text{CoVar}(U_{m,n,m',n'}, U_{\tilde{m},\tilde{n},\tilde{m}',\tilde{n}'})$

- All 8 indices distinct: independent, covariance of 0.
- 7 indices distinct: At most  $N^3$  choices for first tuple, at most  $N^2$  for second, get

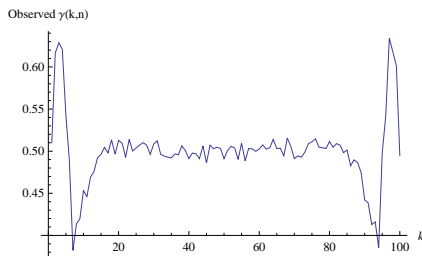
$$\mathbb{E}[U_{(1)}U_{(2)}] - \mathbb{E}[U_{(1)}]\mathbb{E}[U_{(2)}] = N^{-7\delta} - N^{-4\delta}N^{-4\delta} \leq N^{-7\delta}.$$

- Argue similarly for rest, get  $\ll N^{5-7\delta} + N^{3-4\delta}$ .

## Open Problems

## Probability $k$ in an MSTD set (uniform model)

$$\gamma(k, n) := \text{Prob}(k \in A : A \subset [1, n] \text{ is an MSTD set})$$



**Figure:** Observed  $\gamma(k, 100)$ , random sample 4458 MSTD sets.

### Conjecture

Fix a constant  $0 < \alpha < 1$ . Then  $\lim_{n \rightarrow \infty} \gamma(k, n) = 1/2$  for  $\lfloor \alpha n \rfloor \leq k \leq n - \lfloor \alpha n \rfloor$ .

## Generalization of main result

Theorem (Hegarty-M): Binomial model with parameter  $p(N)$  as before,  $u, v$  be non-zero integers with  $u \geq |v|$ ,  $\gcd(u, v) = 1$  and  $(u, v) \neq (1, 1)$ . Put  $f(x, y) := ux + vy$  and let  $\mathcal{D}_f$  denote the random variable  $|f(A)|$ . Then the following three situations arise:

- 1  $p(N) = o(N^{-1/2})$  : Then

$$\mathcal{D}_f \sim (N \cdot p(N))^2.$$

- 2  $p(N) = c \cdot N^{-1/2}$  for some  $c \in (0, \infty)$  : Define the function  $g_{u,v} : (0, \infty) \rightarrow (0, u + |v|)$  by

$$g_{u,v}(x) := (u + |v|) - 2|v| \left( \frac{1 - e^{-x}}{x} \right) - (u - |v|)e^{-x}.$$

Then

$$\mathcal{D}_f \sim g_{u,v} \left( \frac{c^2}{u} \right) N.$$

- 3  $N^{-1/2} = o(p(N))$  : Let  $\mathcal{D}_f^c := (u + |v|)N - \mathcal{D}_f$ . Then  $\mathcal{D}_f^c \sim \frac{2u|v|}{p(N)^2}$ .

## Generalization of main results (cont)

Let  $f, g$  be two binary linear forms. Say  $f$  **dominates**  $g$  for the parameter  $p(N)$  if, as  $N \rightarrow \infty$ ,  $|f(A)| > |g(A)|$  almost surely when  $A$  is a random subset (binomial model with parameter  $p(N)$ ).

Theorem (Hegarty-M):  $f(x, y) = u_1x + u_2y$  and  $g(x, y) = u_2x + g_2y$ , where  $u_i \geq |v_i| > 0$ ,  $\gcd(u_i, v_i) = 1$  and  $(u_2, v_2) \neq (u_1, \pm v_1)$ . Let

$$\alpha(u, v) := \frac{1}{u^2} \left( \frac{|v|}{3} + \frac{u - |v|}{2} \right) = \frac{3u - |v|}{6u^2}.$$

The following two situations can be distinguished :

- $u_1 + |v_1| \geq u_2 + |v_2|$  and  $\alpha(u_1, v_1) < \alpha(u_2, v_2)$ . Then  $f$  dominates  $g$  for all  $p$  such that  $N^{-3/5} = o(p(N))$  and  $p(N) = o(1)$ . In particular, every other difference form dominates the form  $x - y$  in this range.
- $u_1 + |v_1| > u_2 + |v_2|$  and  $\alpha(u_1, v_1) > \alpha(u_2, v_2)$ . Then there exists  $c_{f,g} > 0$  such that one form dominates for  $p(N) < cN^{-1/2}$  ( $c < c_{f,g}$ ) and other dominates for  $p(N) > cN^{-1/2}$  ( $c > c_{f,g}$ ).

## Open Problems

- One unresolved matter is the comparison of arbitrary difference forms in the range where  $N^{-3/4} = O(p)$  and  $p = O(N^{-3/5})$ . Note that the property of one binary form dominating another is not monotone, or even convex.
- A very tantalizing problem is to investigate what happens while crossing a sharp threshold.
- One can ask if the various concentration estimates can be improved (i.e., made explicit).

## Programs

## Mathematica Code: Computing Sum/Difference Set

```

setA = {1, 2, 5, 7, 11, 13, 17, 19};
sumset = {};
diffset = {};
n = Length[setA];
For[i = 1, i <= n, i++,
For[j = 1, j <= n, j++,
{
sum = setA[[i]] + setA[[j]];
diff = setA[[i]] - setA[[j]];
If[MemberQ[sumset, sum] == False, sumset = AppendTo[sumset, sum]];
If[MemberQ[diffset, diff] == False, diffset = AppendTo[diffset, diff]];
}]];
sumset = Sort[sumset];
diffset = Sort[diffset];
Print[sumset];
Print[diffset];
Print["Size of sumset = ", Length[sumset], " and size of difference set = ",
Length[diffset], "."];

```

## Bibliography

## Bibliography

- N. Alon and J. H. Spencer, *The Probabilistic Method*, Wiley, 1992.
- G. A. Freiman and V. P. Pigarev, *The relation between the invariants  $R$  and  $T$* , Numbertheoretic studies in the Markov spectrum and in the structural theory of set addition (Russian), Kalinin. Gos. Univ., Moscow, 1973, 172–174.
- A. P. Godbole, S. Janson, N. W. Locantore Jr. and R. Rapoport, *Random Sidon sequences*, J. Number Theory **75** (1999), no. 1, 7–22.
- P. V. Hegarty, *Some explicit constructions of sets with more sums than differences* (2007). To appear in Acta Arithmetica.
- P. V. Hegarty and S. J. Miller, *When almost all sets are difference dominated*, to appear in Random Structures and Algorithms.  
<http://arxiv.org/abs/0707.3417>
- J. H. Kim and V. H. Vu, *Concentration of multivariate polynomials and its applications*, Combinatorica **20** (2000), 417–434.
- J. Marica, *On a conjecture of Conway*, Canad. Math. Bull. **12** (1969), 233–234.

## Bibliography (cont)

- G. Martin and K. O'Bryant, *Many sets have more sums than differences*. To appear in : Proceedings of CRM-Clay Conference on Additive Combinatorics, Montréal 2006.
- S. J. Miller, B. Orosz and D. Scheinerman, *Explicit constructions of infinite families of MSTD sets*, preprint.  
<http://arxiv.org/abs/0809.4621>
- M. B. Nathanson, *Problems in additive number theory, 1*. To appear in : Proceedings of CRM-Clay Conference on Additive Combinatorics, Montréal 2006.
- M. B. Nathanson, *Sets with more sums than differences*, Integers : Electronic Journal of Combinatorial Number Theory **7** (2007), Paper A5 (24pp).
- M. B. Nathanson, K. O'Bryant, B. Orosz, I. Ruzsa and M. Silva, *Binary linear forms over finite sets of integers* (2007). To appear in Acta Arithmetica.
- 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 Tarsulat, 1978, 933–938.

## Bibliography (cont)

- I. Z. Ruzsa, *Sets of sums and differences*, Seminaire de Theorie des Nombres de Paris 1982-1983 (Boston), Birkhauser, 1984, 267–273.
- I. Z. Ruzsa, *On the number of sums and differences*, Acta Math. Sci. Hungar. **59** (1992), 439–447.
- V. H. Vu, *New bounds on nearly perfect matchings of hypergraphs: Higher codegrees do help*, Random Structures and Algorithms **17** (2000), 29–63.
- V. H. Vu, *Concentration of non-Lipschitz functions and Applications*, Random Structures and Algorithms **20** (2002), no. 3, 262-316.