When almost all sets are difference dominated in $\mathbb{Z}/n\mathbb{Z}$

Adam Lott, alott@u.rochester.edu
(Joint work with Anand Hemmady and Steven J. Miller)

Department of Mathematics, University of Rochester, Rochester, NY 14627

Integers Conference
University of West Georgia
October 7, 2016
Given a set $A \subset \mathbb{Z}$, define the \textit{sumset} and \textit{difference set}

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

\begin{definition}
If $|A + A| > |A - A|$, $A$ is said to be \textit{sum-dominated}.
If $|A + A| = |A - A|$, $A$ is said to be \textit{balanced}.
If $|A + A| < |A - A|$, $A$ is said to be \textit{difference-dominated}.
\end{definition}
Background

- Addition commutes, subtraction doesn’t.

“Even though there exist sets $A$ which 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 $|A - A| > |A + A|$.”

–Melvyn Nathanson
Theorem (Martin and O’Bryant, 2006)

A positive proportion of sets of integers are sum-dominated, in the sense that the quantity

$$\liminf_{n \to \infty} \frac{\text{\# of sum-dominated subsets of } \{1, \ldots, n\}}{2^n}$$

is positive.

Equivalent: if we pick a subset of \( \{1, \ldots, n\} \) uniformly at random, the probability of being sum-dominated is nonzero as \( n \to \infty \).
Known results

What’s going on?

- “Fringe” elements are most important.
  - Large numbers and small numbers have fewer representations as sums than numbers in the middle.
  - Think of rolling two dice – more ways to get 7 than 12.
- If $A$ is big, then almost every possible sum and difference is realized.
Known results

- What if we pick random subsets in a different way?
- Construct $A \subseteq \{1, \ldots, n\} \subset \mathbb{Z}$ randomly by picking each element independently with probability $p(n)$.
  - Uniform case corresponds to $p(n) = 1/2$ constant.
  - Let $p(n)$ decay to 0 as $n \to \infty$ (smaller sets are more likely to be picked).
Known results

- What if we pick random subsets in a different way?
- Construct $A \subseteq \{1, \ldots, n\} \subset \mathbb{Z}$ randomly by picking each element independently with probability $p(n)$.
  - Uniform case corresponds to $p(n) = 1/2$ constant.
  - Let $p(n)$ decay to 0 as $n \to \infty$ (smaller sets are more likely to be picked).

**Theorem (Hegarty and Miller, 2009)**

Let $A \subseteq \{1, \ldots, n\} \subset \mathbb{Z}$ be chosen randomly in this way where $p(n) = o(1)$. Then

$$\text{Prob (} A \text{ is difference-dominated) } \to 1 \text{ as } n \to \infty.$$
New setting

- Look at subsets $A \subseteq \mathbb{Z}/n\mathbb{Z}$ (i.e., take sums and differences modulo $n$).
  - No fringe elements!

- Construct randomly according to decaying probability $p(n)$.
  - Try to avoid sumsets and difference sets being full.
Let $X(n)$ and $Y(n)$ be random variables depending on $n$. We write
$X(n) \sim Y(n)$ if, for every $\epsilon > 0$,

$$\text{Prob} \left( \left| \frac{X(n)}{Y(n)} - 1 \right| < \epsilon \right) \to 1 \text{ as } n \to \infty.$$
Our result (full statement)

**Theorem (HLM, 2016)**

Let $A \subseteq \mathbb{Z}/n\mathbb{Z}$ be chosen randomly according to a binomial parameter $p(n) = o(1)$.

- **(Fast decay)** If $p(n) = o(n^{-1/2})$, then $|A + A| \sim \frac{1}{2}(np(n))^2$ and $|A - A| \sim (np(n))^2$.

- **(Critical decay)** If $p(n) = c \cdot n^{-1/2}$, then $|A + A| \sim (1 - \exp(-c^2/2))n$ and $|A - A| \sim (1 - \exp(-c^2))n$.

- **(Slow decay)** If $\sqrt{\log n} \cdot n^{-1/2} = o(p(n))$ and $n$ is prime, then $|A + A| \sim |A - A| \sim n$. 

Our result (qualitative statement)

Theorem (HLM, 2016)

Let $A \subseteq \mathbb{Z}/n\mathbb{Z}$ be chosen randomly according to a binomial parameter $p(n) = o(1)$.

- (Fast/critical decay) If $p(n) = O(n^{-1/2})$, then
  
  $\Pr(A \text{ is difference-dominated}) \to 1$ as $n \to \infty$.

- (Slow decay) If $n^{-1/2} \sqrt{\log n} = o(p(n))$ and $n$ is prime, then
  
  $\Pr(A \text{ is balanced}) \to 1$ as $n \to \infty$. 
Fast/critical decay ($p(n) = O(n^{-1/2})$)

- Expect $|A| \sim np(n)$.
- Control number of times a sum or difference is realized more than once.
  - Compute mean number of repeats and bound the variance.
  - Modify techniques of Hegarty and Miller.
- In slow decay case, get
  \[
  |A + A| \sim \binom{|A|}{2} = \frac{1}{2}|A||A| - 1 \sim \frac{1}{2}(np(n))^2 \\
  |A - A| \sim |A||A| - 1 \sim (np(n))^2.
  \]
Fast/critical decay \((p(n) = O(n^{-1/2}))\)

- Expect \(|A| \sim np(n)\).
- Control number of times a sum or difference is realized more than once.
  - Compute mean number of repeats and bound the variance.
  - Modify techniques of Hegarty and Miller.
- In slow decay case, get
  \[
  |A + A| \sim \left(\frac{|A|}{2}\right) = \frac{1}{2}|A|(|A| - 1) \sim \frac{1}{2}(np(n))^2
  \]
  \[
  |A - A| \sim |A|(|A| - 1) \sim (np(n))^2.
  \]
- Critical decay case is similar, but a bit more delicate.
Slow decay \( (\sqrt{\log n \cdot n^{-1/2}} = o(p(n))) \)

- No control over number of repeats.
  - When \( p(n) \gg n^{-1/2} \), expect \( |A| \sim np(n) \gg n^{1/2} \).
  - Number of pairs \( \sim |A|^2 \gg n \), but only \( n \) possible sums!
Slow decay (\(\sqrt{\log n \cdot n^{-1/2}} = o(p(n))\))

- No control over number of repeats.
  - When \(p(n) \gg n^{-1/2}\), expect \(|A| \sim np(n) \gg n^{1/2}\).
  - Number of pairs \(\sim |A|^2 \gg n\), but only \(n\) possible sums!

- Compute number of \textit{missing} sums and differences instead.
  - Show they are both 0 with high probability.
Idea of proof

- $S^c :=$ number of missing sums.
- $D^c :=$ number of missing differences.
- Show $\mathbb{E}[S^c], \mathbb{E}[D^c], \text{Var}(S^c),$ and $\text{Var}(D^c)$ all tend to 0 as $n \to \infty$.
  - By Chebyshev’s inequality, this implies $\text{Prob}(S^c = D^c = 0) \to 1$ as $n \to \infty$. 
Comparison with $\mathbb{Z}$

- In $\mathbb{Z}$, $\mathbb{E}[S^c]$ and $\mathbb{E}[D^c]$ don’t tend to 0 (Hegarty & Miller).
  - Qualitatively different behavior in $\mathbb{Z}/n\mathbb{Z}$.
- In $\mathbb{Z}$, need heavy machinery from probability to prove strong concentration.
  - More elementary arguments in $\mathbb{Z}/n\mathbb{Z}$. 
Computing $\mathbb{E}[S^c]$ 

Write

$$\mathbb{E}[S^c] = \sum_{k \in \mathbb{Z}/n\mathbb{Z}} \text{Prob}(k \not\in A + A).$$
Computing $E[S^c]$

- Each $k \in \mathbb{Z}/n\mathbb{Z}$ can be written as a sum in exactly $(n + 1)/2$ disjoint ways.
  - This is what separates $\mathbb{Z}/n\mathbb{Z}$ from $\mathbb{Z}$. 

\[
E[S^c] = n(1 - p^2)\frac{(n+1)}{2} \sim n\frac{(1 - p^2)}{2n}.
\]

Note: doesn't tend to 0 unless $\sqrt{\log n} \cdot n - \frac{1}{2} = o(p(n))$. 

\[
\operatorname{Prob}(k \not\in A + A) = (1 - p^2)^2 \frac{(n+1)}{2} \text{ independently of } k.
\]
Computing $\mathbb{E} [S^c]$

- Each $k \in \mathbb{Z}/n\mathbb{Z}$ can be written as a sum in exactly $(n + 1)/2$ disjoint ways.
  - This is what separates $\mathbb{Z}/n\mathbb{Z}$ from $\mathbb{Z}$.
- $\text{Prob } (k \not\in A + A) = (1 - p^2)^{(n+1)/2}$ independently of $k$.
- $\mathbb{E} [S^c] = n(1 - p^2)^{(n+1)/2} \sim n(1 - p^2)^{n/2}$.
  - Note: doesn’t tend to 0 unless $\sqrt{\log n \cdot n^{-1/2}} = o(p(n))$. 
Each $k \in \mathbb{Z}/n\mathbb{Z}$ can be written as a difference in exactly $n$ different ways.

Pairs aren’t disjoint, so we can’t count them independently like we did for sums.
Each $k \in \mathbb{Z}/n\mathbb{Z}$ can be written as a difference in exactly $n$ different ways.

- Pairs aren't disjoint, so we can't count them independently like we did for sums.

- Translate to graph theory.
Graph theoretic framework

- Modeling Prob \( k \not\in A - A \).
- Each element of \( \mathbb{Z}/n\mathbb{Z} \) is a vertex, connect \( a \) to \( b \) if \( a - b \equiv k \) (mod \( n \)).
- Example (\( n = 7, k = 2 \)): 

![Graph diagram]

\[\begin{array}{c}
\text{0} & \text{6} & \text{1} & \text{2} & \text{5} & \text{4} & \text{3} & \text{1} & \text{6} & \text{2} & \text{4} & \text{5} & \text{0} \\
\end{array}\]
Computing $\mathbb{E} [D^c]$ 

- $\text{Prob}(k \notin A - A)$ is the same as the probability that no two adjacent vertices are in $A$.
- Equivalent: pick a random subset of $\{1, \ldots, n\}$, probability that it doesn’t contain any consecutive elements.
Counting problem – probability is

\[ \sum_{r=1}^{\lfloor n/2 \rfloor} \left[ \binom{n-r+1}{r} - \binom{n-r-1}{r-2} \right] p^r (1-p)^{n-r} \]

\[ \sim \sum_{r=1}^{\lfloor n/2 \rfloor} \binom{n-r}{r} p^r (1-p)^{n-r} \]

So

\[ \mathbb{E}[D^c] \sim n \sum_{r=1}^{\lfloor n/2 \rfloor} \binom{n-r}{r} p^r (1-p)^{n-r}. \]
Computing variances

- Define indicator random variables

\[ X_k := \begin{cases} 
1 & k \notin A + A \\
0 & k \in A + A.
\end{cases} \]

- \( S^c = \sum_{k \in \mathbb{Z}/n\mathbb{Z}} X_k. \)
Computing variances

- Define indicator random variables

\[ X_k := \begin{cases} 
1 & k \not\in A + A \\
0 & k \in A + A.
\end{cases} \]

- \( S^c = \sum_{k\in\mathbb{Z}/n\mathbb{Z}} X_k. \)
Computing variances

- Define indicator random variables

\[ X_k := \begin{cases} 
1 & k \not\in A + A \\
0 & k \in A + A. 
\end{cases} \]

- \( S^c = \sum_{k \in \mathbb{Z}/n\mathbb{Z}} X_k. \)

- \( X_k \) are not independent, so

\[
\text{Var} (S^c) = \sum_{k \in \mathbb{Z}/n\mathbb{Z}} \text{Var} (X_k) + \sum_{i \neq j \in \mathbb{Z}/n\mathbb{Z}} \text{Cov} (X_i, X_j). 
\]
Computing variances

- Covariance terms rely on evaluating
  \[ \text{Prob} \left( i \notin A + A \text{ and } j \notin A + A \right). \]

- Graph theory works again!
Graph theoretic framework

- $n, i, j$ fixed.
- Connect $a$ and $b$ with an edge if $a + b \equiv i$ or $a + b \equiv j \mod n$.
- Example $(n = 7, i = 2, j = 5)$:
Computing variances

- Translate to same counting problem.
- So

\[ \text{Prob} \left( i \notin A + A \text{ and } j \notin A + A \right) \sim \sum_{r=1}^{\left\lfloor n/2 \right\rfloor} \binom{n-r}{r} p^r (1-p)^{n-r}. \]

- In variance expression, this term dominates, giving

\[ \text{Var} \left( S^c \right) \sim n^2 \sum_{r=1}^{\left\lfloor n/2 \right\rfloor} \binom{n-r}{r} p^r (1-p)^{n-r}. \]

- \( \text{Var} \left( D^c \right) \) handled similarly.
Getting a good estimate

Key Lemma

Let

\[ F(n) := \sum_{r=1}^{\lfloor n/2 \rfloor} \binom{n-r}{r} p^r (1-p)^{n-r}. \]

Then \( F(n) = o(1/n^3) \).
Getting a good estimate

- By comparing to a binomial distribution and using Stirling’s formula, we can get the bound

\[ n^3 F(n) \leq 2n^4 (e^p - pe^p)^n. \]

- Take log and use power series expansion:

\[ \log(n^3 F(n)) \ll \log n - \frac{1}{2} np^2 + O(np^3). \]

- Tends to \(-\infty\) provided

\[ \log n = o(np^2) \iff \sqrt{\log n} \cdot n^{-1/2} = o(p(n)). \]
"Correspondence" principle

- When $p(n)$ decays rapidly, subsets of $\mathbb{Z}/n\mathbb{Z}$ behave like subsets of $\mathbb{Z}$ (as $n \to \infty$).
- When $p(n)$ decays slowly, subsets of $\mathbb{Z}/n\mathbb{Z}$ behave as if $p(n)$ were constant (as $n \to \infty$).
Open questions

What happens when \( n^{-1/2} \ll p(n) \ll \sqrt{\log n} \cdot n^{-1/2} \)?

Can we extend slow decay analysis to non-prime \( n \)?
Thanks

- Anand Hemmady, Steven J. Miller
- University of West Georgia
- Bruce Landman, Florian Luca, Melvyn Nathanson, Jarik Nesetril, Richard Nowakowski
- Williams College
- NSF/SMALL REU
- Williams College Finnerty Fund
