He’s just going through a phase: Miller’s SMALL students and Phase Transitions

Steven J. Miller, Williams College

sjml@williams.edu, Steven.Miller.MC.96@aya.yale.edu
http://web.williams.edu/Mathematics/sjmiller/public_html/

Williams College, October 10, 2014
Introduction to Continued Fractions and Maclaurin’s Inequality
Joint with Francesco Cellarosi (UIUC), Doug Hensley (Texas A & M), Jake Wellens (SMALL ’13)
Plan of Part I: Continued Fractions and Maclaurin’s Inequality

- Classical ergodic theory of continued fractions.
  - Almost surely geometric mean $\sqrt{a_1 \cdots a_n} \to K_0$.
  - Almost surely arithmetic mean $\left( a_1 + \cdots + a_n \right) / n \to \infty$.

- Symmetric averages and Maclaurin’s inequalities.
  - $S(x, n, k) := \binom{n}{k}^{-1} \sum_{1 \leq i_1 < i_2 < \cdots < i_k \leq n} x_{i_1} x_{i_2} \cdots x_{i_k}$.
  - $\text{AM} = S(x, n, 1)^{1/1} \geq S(x, n, 2)^{1/2} \geq \cdots \geq S(x, n, n)^{1/n} = \text{GM}$.

- Results / conjectures on typical / periodic continued fraction averages.

- Elementary proofs of weak results, sketch of stronger results.

Every real number $\alpha \in (0, 1)$ can be expressed as

$$x = \frac{1}{a_1 + \frac{1}{a_2 + \frac{1}{a_3 + \cdots}}} = [a_1, a_2, a_3, \ldots], \quad a_i \in \{1, 2, \ldots \}.$$
Continued Fractions

- Every real number $\alpha \in (0, 1)$ can be expressed as

$$x = \frac{1}{a_1 + \frac{1}{a_2 + \frac{1}{a_3 + \ddots}}} = [a_1, a_2, a_3, \ldots], \quad a_i \in \{1, 2, \ldots\}.$$ 

- The sequence $\{a_i\}_i$ is finite iff $\alpha \in \mathbb{Q}$. 
Continued Fractions

- Every real number $\alpha \in (0, 1)$ can be expressed as

  $$ x = \cfrac{1}{a_1 + \cfrac{1}{a_2 + \cfrac{1}{a_3 + \cdots}}} = [a_1, a_2, a_3, \ldots], \quad a_i \in \{1, 2, \ldots\}. $$

- $x = \frac{p}{q} \in \mathbb{Q}$ then $a_i$’s the partial quotients of Euclidean Alg.

  $$ \frac{333}{106} = [3, 7, 15] \quad 333 = 3 \cdot 106 + 15 $$

  $$ \frac{106}{333} = [3, 7, 15] \quad 106 = 7 \cdot 15 + 1 $$

  $$ 15 = 15 \cdot 1 + 0. $$
Continued Fractions

- Every real number $\alpha \in (0, 1)$ can be expressed as

$$x = \frac{1}{a_1 + \frac{1}{a_2 + \frac{1}{a_3 + \ldots}}} = [a_1, a_2, a_3, \ldots], \quad a_i \in \{1, 2, \ldots\}.$$  

- $\{a_i\}_i$ preperiodic iff $\alpha$ a quadratic irrational;  
  ex: $\sqrt{3} - 1 = [1, 2, 1, 2, 1, 2, \ldots]$.  


Gauss Map: Definition

- The Gauss map $T : (0, 1] \rightarrow (0, 1]$, $T(x) = \{ \frac{1}{x} \} = \frac{1}{x} - [\frac{1}{x}]$ generates the continued fraction digits

  $$a_1 = \lfloor 1/T^0(\alpha) \rfloor, \quad a_{i+1} = \lfloor 1/T^i(\alpha) \rfloor, \quad \ldots$$

  corresponding to the Markov partition

  $$\begin{equation}
  (0, 1] = \bigcup_{k=1}^{\infty} \left( \frac{1}{k+1}, \frac{1}{k} \right).
  \end{equation}$$

- $T$ preserves the measure $d\mu = \frac{1}{\log 2} \frac{1}{1+x} dx$ and it is mixing.
Gauss Map: Example: $\sqrt{3} - 1 = [1, 2, 1, 2, 1, 2, \ldots]$

$T : (0, 1] \to (0, 1], T(x) = \{\frac{1}{x}\} = \frac{1}{x} - \lfloor \frac{1}{x} \rfloor$ generates digits

$a_1 = \lfloor 1/T^0(\alpha) \rfloor, \quad a_{i+1} = \lfloor 1/T^i(\alpha) \rfloor, \quad \ldots$

$\alpha = \sqrt{3} - 1 = [1, 2, 1, 2, \ldots]:$ Note $a_1 = \lfloor \frac{1}{\sqrt{3} - 1} \rfloor = 1
Gauss Map: Example: $\sqrt{3} - 1 = [1, 2, 1, 2, 1, 2, \ldots]$

$$T : (0, 1] \rightarrow (0, 1], \ T(x) = \{ \frac{1}{x} \} = \frac{1}{x} - \lfloor \frac{1}{x} \rfloor$$ generates digits

$$a_1 = \lfloor 1/T^0(\alpha) \rfloor, \quad a_{i+1} = \lfloor 1/T^i(\alpha) \rfloor, \quad \ldots$$

$\alpha = \sqrt{3} - 1 = [1, 2, 1, 2, \ldots]$: Note $a_1 = \lfloor \frac{1}{\sqrt{3} - 1} \rfloor = 1$ and

$$T^1(\sqrt{3} - 1) = \frac{1}{\sqrt{3} - 1} - \left\lfloor \frac{1}{\sqrt{3} - 1} \right\rfloor = \frac{\sqrt{3} + 1}{3 - 1} - 1 = \frac{\sqrt{3} - 1}{2}$$

$$a_2 = \left\lfloor \frac{2}{\sqrt{3} - 1} \right\rfloor = 2.$$
Gauss Map: Example: \( \sqrt{3} - 1 = [1, 2, 1, 2, 1, 2, \ldots] \)

\[ T : (0, 1] \rightarrow (0, 1], \quad T(x) = \left\{ \frac{1}{x} \right\} = \frac{1}{x} - \left\lfloor \frac{1}{x} \right\rfloor \text{ generates digits} \]

\[ a_1 = \left\lfloor \frac{1}{T^0(\alpha)} \right\rfloor, \quad a_{i+1} = \left\lfloor \frac{1}{T^i(\alpha)} \right\rfloor, \quad \ldots \]

\( \alpha = \sqrt{3} - 1 = [1, 2, 1, 2, \ldots] \): Note \( a_1 = \left\lfloor \frac{1}{\sqrt{3} - 1} \right\rfloor = 1 \) and

\[ T^1(\sqrt{3} - 1) = \frac{1}{\sqrt{3} - 1} - \left\lfloor \frac{1}{\sqrt{3} - 1} \right\rfloor = \frac{\sqrt{3} + 1}{3 - 1} - 1 = \frac{\sqrt{3} - 1}{2} \]

\[ a_2 = \left\lfloor \frac{2}{\sqrt{3} - 1} \right\rfloor = 2. \]

\[ T^2(\sqrt{3} - 1) = \frac{2}{\sqrt{3} - 1} - \left\lfloor \frac{2}{\sqrt{3} - 1} \right\rfloor = \frac{2\sqrt{3} + 2}{2} - 2 = \sqrt{3} - 1 \]

\[ a_3 = \left\lfloor \frac{1}{\sqrt{3} - 1} \right\rfloor = 1. \]
**Gauss Map: Example:** \( \sqrt{3} - 1 = [1, 2, 1, 2, 1, 2, \ldots] \)

\[
T : (0, 1] \rightarrow (0, 1], \quad T(x) = \{\frac{1}{x}\} = \frac{1}{x} - \lfloor \frac{1}{x} \rfloor \text{ generates digits}
\]

\[
a_1 = \lfloor 1 / T^0(\alpha) \rfloor, \quad a_{i+1} = \lfloor 1 / T^i(\alpha) \rfloor, \quad \ldots
\]

\[
\alpha = \sqrt{3} - 1 = [1, 2, 1, 2, \ldots]: \text{Note } a_1 = \lfloor \frac{1}{\sqrt{3} - 1} \rfloor = 1 \text{ and}
\]

\[
T^1(\sqrt{3} - 1) = \frac{1}{\sqrt{3} - 1} - \left\lfloor \frac{1}{\sqrt{3} - 1} \right\rfloor = \frac{\sqrt{3} + 1}{3 - 1} - 1 = \frac{\sqrt{3} - 1}{2}
\]

\[
a_2 = \left\lfloor \frac{2}{\sqrt{3} - 1} \right\rfloor = 2.
\]

\[
T^2(\sqrt{3} - 1) = \frac{2}{\sqrt{3} - 1} - \left\lfloor \frac{2}{\sqrt{3} - 1} \right\rfloor = \frac{2\sqrt{3} + 2}{2} - 2 = \sqrt{3} - 1
\]

\[
a_3 = \left\lfloor \frac{1}{\sqrt{3} - 1} \right\rfloor = 1.
\]
The digits $a_i$ follow the Gauss-Kuzmin distribution:

$$\lim_{n \to \infty} \mathbb{P}(a_n = k) = \log_2 \left( 1 + \frac{1}{k(k+2)} \right)$$

(note the expectation is infinite).
Statistics of Continued Fraction Digits 1/3

The digits $a_i$ follow the Gauss-Kuzmin distribution:

$$
\lim_{n \to \infty} \mathbb{P}(a_n = k) = \log_2 \left( 1 + \frac{1}{k(k+2)} \right)
$$

(note the expectation is infinite).

The function $x \mapsto f(x) = \lfloor 1/T(x) \rfloor$ on $(0, 1]$ is not integrable wrt $\mu$. However, $\log f \in L^1(\mu)$. 
Statistics of Continued Fraction Digits 1/3

- The digits $a_i$ follow the Gauss-Kuzmin distribution:

$$\lim_{n \to \infty} P(a_n = k) = \log_2 \left( 1 + \frac{1}{k(k+2)} \right)$$

(note the expectation is infinite).

- The function $x \mapsto f(x) = \lfloor 1/T(x) \rfloor$ on $(0,1]$ is not integrable wrt $\mu$. However, $\log f \in L^1(\mu)$.

- Pointwise ergodic theorem (applied to $f$ and $\log f$) reads

$$\lim_{n \to \infty} \frac{a_1 + a_2 + \cdots + a_n}{n} = \infty \quad \text{almost surely}$$

$$\lim_{n \to \infty} (a_1 a_2 \cdots a_n)^{1/n} = e^{\int \log f \, d\mu} \quad \text{almost surely.}$$
Statistics of Continued Fraction Digits 2/3

- Geometric mean converges a.s. to Khinchin’s constant:

\[
\lim_{n \to \infty} \left( a_1 a_2 \cdots a_n \right)^{1/n} = \prod_{k=1}^{\infty} \left( 1 + \frac{1}{k(k+2)} \right)^{\log_2 k} = K_0 \approx 2.6854.
\]
Statistics of Continued Fraction Digits 2/3

- Geometric mean converges a.s. to Khinchin’s constant:

\[
\lim_{n \to \infty} (a_1 a_2 \cdots a_n)^{1/n} = \prod_{k=1}^{\infty} \left(1 + \frac{1}{k(k+2)}\right)^{\log_2 k} = K_0 \approx 2.6854.
\]

- Hölder means: For \( p < 1 \), almost surely

\[
\lim_{n \to \infty} \left(\frac{1}{n} \sum_{i=1}^{n} a_i^p\right)^{1/p} = K_p = \left(\sum_{k=1}^{\infty} -k^p \log_2 \left(1 - \frac{1}{(k+1)^2}\right)\right)^{1/p}.
\]
Statistics of Continued Fraction Digits 2/3

- Geometric mean converges a.s. to Khinchin’s constant:

$$\lim_{n \to \infty} (a_1 a_2 \cdots a_n)^{1/n} = \prod_{k=1}^{\infty} \left(1 + \frac{1}{k(k+2)}\right)^{\log_2 k} = K_0 \approx 2.6854.$$

- Hölder means: For $p < 1$, almost surely

$$\lim_{n \to \infty} \left(\frac{1}{n} \sum_{i=1}^{n} a_i^p\right)^{1/p} = K_p = \left(\sum_{k=1}^{\infty} -k^p \log_2 \left(1 - \frac{1}{(k+1)^2}\right)\right)^{1/p}.$$

- Example: The harmonic mean $K_{-1} = 1.74540566 \ldots$
Geometric mean converges a.s. to Khinchin’s constant:

$$\lim_{n \to \infty} (a_1 a_2 \cdots a_n)^{1/n} = \prod_{k=1}^{\infty} \left( 1 + \frac{1}{k(k+2)} \right)^{\log_2 k} = K_0 \approx 2.6854.$$ 

Hölder means: For $p < 1$, almost surely

$$\lim_{n \to \infty} \left( \frac{1}{n} \sum_{i=1}^{n} a_i^p \right)^{1/p} = K_p = \left( \sum_{k=1}^{\infty} -k^p \log_2 \left( 1 - \frac{1}{(k+1)^2} \right) \right)^{1/p}. $$

Example: The harmonic mean $K_{-1} = 1.74540566 \ldots$

$$\lim_{p \to 0} K_p = K_0.$$
Khinchin also proved: For \( a'_m = a_m \) if \( a_m < m(\log m)^{4/3} \) and 0 otherwise:

\[
\lim_{n \to \infty} \frac{\sum_{i=1}^{n} a'_i}{n \log n} = \frac{1}{\log 2} \quad \text{in measure.}
\]
Khinchin also proved: For $a'_m = a_m$ if $a_m < m(\log m)^{4/3}$ and 0 otherwise:

$$\lim_{n \to \infty} \frac{\sum_{i=1}^{n} a'_i}{n \log n} = \frac{1}{\log 2} \quad \text{in measure.}$$

Diamond and Vaaler (1986) showed that

$$\lim_{n \to \infty} \frac{\sum_{i=1}^{n} a_i - \max_{1 \leq i \leq n} a_i}{n \log n} = \frac{1}{\log 2} \quad \text{almost surely.}$$
Maclaurin Inequalities
Definitions and Maclaurin’s Inequalities

- Both $\frac{1}{n} \sum_{i=1}^{n} x_i$ and $\left( \prod_{i=1}^{n} x_i \right)^{1/n}$ are defined in terms of elementary symmetric polynomials in $x_1, \ldots, x_n$.
- Define $k^{\text{th}}$ elementary symmetric mean of $x_1, \ldots, x_n$ by

$$S(x, n, k) := \frac{1}{\binom{n}{k}} \sum_{1 \leq i_1 < i_2 < \ldots < i_k \leq n} x_{i_1} x_{i_2} \cdots x_{i_k}.$$
Definitions and Maclaurin’s Inequalities

- Both $\frac{1}{n} \sum_{i=1}^{n} x_i$ and $(\prod_{i=1}^{n} x_i)^{1/n}$ are defined in terms of elementary symmetric polynomials in $x_1, \ldots, x_n$.

- Define $k^{th}$ elementary symmetric mean of $x_1, \ldots, x_n$ by

$$S(x, n, k) := \frac{1}{\binom{n}{k}} \sum_{1 \leq i_1 < i_2 < \cdots < i_k \leq n} x_{i_1} x_{i_2} \cdots x_{i_k}.$$ 

Maclaurin’s Inequalities

For positive $x_1, \ldots, x_n$ we have

$$\text{AM} := S(x, n, 1)^{1/1} \geq S(x, n, 2)^{1/2} \geq \cdots \geq S(x, n, n)^{1/n} := \text{GM}$$

(and equalities hold iff $x_1 = \cdots = x_n$).
IV. A second Letter from Mr. Colin MacLaurin, Professor of Mathematicks in the University of Edinburgh and F. R. S. to Martin Folkes, Esq; concerning the Roots of Equations, with the Demonstration of other Rules in Algebra; being the Continuation of the Letter published in the Philosophical Transactions, No. 394.

Edinburgh, April 19th, 1729.

S I R,

In the Year 1725, I wrote to you that I had a Method of demonstrating Sir Isaac Newton's Rule concerning the Impossible Roots of Equations, deduced from this obvious Principle, that the Squares of the Differences of real Quantities must always be positive; and some time after, I sent you the first Principles of that Method, which were published in the Philosophical Transactions for the Month of May, 1726. The
Proof

Standard proof through Newton’s inequalities.

Define the $k^{\text{th}}$ elementary symmetric function by

$$s_k(x) = \sum_{1 \leq i_1 < i_2 < \cdots < i_k \leq n} x_{i_1} x_{i_2} \cdots x_{i_k},$$

and the $k^{\text{th}}$ elementary symmetric mean by

$$E_k(x) = \frac{s_k(x)}{\binom{n}{k}}.$$

Newton’s inequality: $E_k(x)^2 \geq E_{k-1}(x)E_{k+1}(x)$.

New proof by Iddo Ben-Ari and Keith Conrad:

Bernoulli’s inequality: $t > -1$: $(1 + t)^n \geq 1 + nt$ or 
$1 + \frac{1}{n}x \geq (1 + x)^{1/n}$.

Generalized Bernoulli: $x > -1$:

$$1 + \frac{1}{n}x \geq \left(1 + \frac{2}{n}x\right)^{1/2} \geq \left(1 + \frac{3}{n}x\right)^{1/3} \geq \cdots \geq \left(1 + \frac{n}{n}x\right)^{1/n}.$$
Bernoulli’s inequality: \( t > -1: (1 + t)^n \geq 1 + nt \) or 
\[
1 + \frac{1}{n}x \geq (1 + x)^{1/n}.
\]

Generalized Bernoulli: \( x > -1: \)
\[
1 + \frac{1}{n}x \geq \left(1 + \frac{2}{n}x\right)^{1/2} \geq \left(1 + \frac{3}{n}x\right)^{1/3} \geq \cdots \geq \left(1 + \frac{n}{n}x\right)^{1/n}.
\]

Proof: Equivalent to \( \frac{1}{k} \log \left(1 + \frac{k}{n}x\right) \geq \frac{1}{k+1} \log \left(1 + \frac{k+1}{n}x\right), \)
which follows by \( \log t \) is strictly concave:
\[
\lambda = \frac{1}{k+1}, 1 + \frac{k}{n}x = \lambda \cdot 1 + (1 - \lambda) \cdot (1 + \frac{k+1}{n}x).
\]
Sketch of Ben-Ari and Conrad’s Proof

Proof of Maclaurin’s Inequalities:

Trivial for \( n \in \{1, 2\} \), wlog assume \( x_1 \leq x_2 \leq \cdots \leq x_n \).

Set \( E_k := s_k(x) / \binom{n}{k} \), \( \epsilon_k := E_k(x_1, \ldots, x_{n-1}) \).

Have
\[
E_k(x_1, \ldots, x_n) = \left(1 - \frac{k}{n}\right) E_k(x_1, \ldots, x_{n-1}) + \frac{k}{n} E_k(x_1, \ldots, x_{n-1}) x_n.
\]

Proceed by induction in number of variables, use Generalized Bernoulli.
Main Results
(Elementary Techniques)
Symmetric Averages and Maclaurin’s Inequalities

- Recall: $S(x, n, k) = \frac{1}{\binom{n}{k}} \sum_{1 \leq i_1 < \cdots < i_k \leq n} x_{i_1} \cdots x_{i_k}$

and $S(x, n, 1)^{1/1} \geq S(x, n, 2)^{1/2} \geq \cdots \geq S(x, n, n)^{1/n}$. 
Symmetric Averages and Maclaurin’s Inequalities

Recall: $S(x, n, k) = \frac{1}{\binom{n}{k}} \sum_{1 \leq i_1 < \cdots < i_k \leq n} x_{i_1} \cdots x_{i_k}$

and $S(x, n, 1)^{1/1} \geq S(x, n, 2)^{1/2} \geq \cdots \geq S(x, n, n)^{1/n}$.

Khinchin’s results: almost surely as $n \to \infty$

$S(\alpha, 1, 1)^{1/1} \to \infty$ and $S(\alpha, n, n)^{1/n} \to K_0$.

We study the intermediate means $S(\alpha, n, k)^{1/k}$ as $n \to \infty$ when $k = k(n)$, with

$$S(\alpha, n, k(n))^{1/k(n)} = S(\alpha, n, \lceil k(n) \rceil)^{1/\lceil k(n) \rceil}.$$
Our results on typical continued fraction averages

Recall: \( S(\alpha, n, k) = \frac{1}{\binom{n}{k}} \sum_{1 \leq i_1 < \ldots < i_k \leq n} a_{i_1} \cdots a_{i_k} \)

and \( S(\alpha, n, 1)^{1/1} \geq S(\alpha, n, 2)^{1/2} \geq \cdots \geq S(\alpha, n, n)^{1/n} \).

**Theorem 1**

Let \( f(n) = o(\log \log n) \) as \( n \to \infty \). Then, almost surely,

\[
\lim_{n \to \infty} S(\alpha, n, f(n))^{1/f(n)} = \infty.
\]

**Theorem 2**

Let \( f(n) = o(n) \) as \( n \to \infty \). Then, almost surely,

\[
\lim_{n \to \infty} S(\alpha, n, n - f(n))^{1/(n-f(n))} = K_0.
\]

Note: Theorems do not cover the case \( f(n) = cn \) for \( 0 < c < 1 \).
Sketch of Proofs of Theorems 1 and 2

**Theorem 1:** For $f(n) = o(\log \log n)$ as $n \to \infty$:

Almost surely \( \lim_{n \to \infty} S(\alpha, n, f(n))^{1/f(n)} = \infty. \)

Uses Niculescu’s strengthening of Maclaurin (2000):

\[
S(n, tj + (1 - t)k) \geq S(n, j)^t \cdot S(n, k)^{1-t}.
\]
Sketch of Proofs of Theorems 1 and 2

**Theorem 1:** For $f(n) = o(\log \log n)$ as $n \to \infty$:

Almost surely $\lim_{n \to \infty} S(\alpha, n, f(n))^{1/f(n)} = \infty$.

Uses Niculescu’s strengthening of Maclaurin (2000):

$$S(n, tj + (1 - t)k) \geq S(n, j)^t \cdot S(n, k)^{1-t}.$$ 

**Theorem 2:** For $f(n) = o(n)$ as $n \to \infty$:

Almost surely $\lim_{n \to \infty} S(\alpha, n, n - f(n))^{1/(n - f(n))} = K_0$.

Use (a.s.) $K_0 \leq \limsup_{n \to \infty} S(\alpha, n, cn)^{1/cn} \leq K_0^{1/c} < \infty$, $0 < c < 1$. 
Proof of Theorem 1: Preliminaries

**Lemma**

Let \( X \) be a sequence of positive real numbers. Suppose \( \lim_{n \to \infty} S(X, n, k(n))^{1/k(n)} \) exists. Then, for any \( f(n) = o(k(n)) \) as \( n \to \infty \), we have

\[
\lim_{n \to \infty} S(X, n, k(n) + f(n))^{1/(k(n) + f(n))} = \lim_{n \to \infty} S(n, k(n))^{1/k(n)}.
\]

**Proof:** Assume \( f(n) \geq 0 \) for large enough \( n \), and for display purposes write \( k \) and \( f \) for \( k(n) \) and \( f(n) \).

From Newton’s inequalities and Maclaurin’s inequalities, we get

\[
\left( S(X, n, k)^{1/k} \right)^{\frac{k}{k+f}} = S(X, n, k)^{1/(k+f)} \leq S(X, n, k+f)^{1/(k+f)} \leq S(X, n, k)^{1/k}.
\]
Proof of Theorem 1: \( f(n) = o(\log \log n) \)

Each entry of \( \alpha \) is at least 1.
Let \( f(n) = o(\log \log n) \). Set \( t = 1/2 \) and \( (j, k) = (1, 2f(n) - 1) \), so that \( tj + (1 - t)k = f(n) \). Niculescu’s result yields

\[
S(\alpha, n, f(n)) \geq \sqrt{S(\alpha, n, 1) \cdot S(\alpha, n, 2f(n) - 1)} > \sqrt{S(\alpha, n, 1)}.
\]

Square both sides, raise to the power \( 1/f(n) \):

\[
S(\alpha, n, f(n))^{2/f(n)} \geq S(\alpha, n, 1)^{1/f(n)}.
\]

From Khinchin almost surely if \( g(n) = o(\log n) \)

\[
\lim_{n \to \infty} \frac{S(\alpha, n, 1)}{g(n)} = \infty.
\]

Let \( g(n) = \log n / \log \log n \). Taking logs:

\[
\log \left( S(\alpha, n, 1)^{1/f(n)} \right) > \frac{\log g(n)}{f(n)} > \frac{\log \log n}{2f(n)}.
\]
Proof of Theorem 2

**Theorem 2**: Let $f(n) = o(n)$ as $n \to \infty$. Then, almost surely,

$$
\lim_{n \to \infty} S(\alpha, n, n - f(n))^{1/(n - f(n))} = K_0.
$$

**Proof**: Follows immediately from:

For any constant $0 < c < 1$ and almost all $\alpha$ have

$$
K_0 \leq \limsup_{n \to \infty} S(\alpha, n, cn)^{1/cn} \leq K_0^{1/c} < \infty.
$$

To see this, note

$$
S(\alpha, n, cn)^{1/cn} = \left( \prod_{i=1}^{n} a_i(\alpha)^{1/n} \right)^{n/cn} \left( \sum_{i_1 < \cdots < i_{(1-c)n} \leq n} 1/(a_{i_1}(\alpha) \cdots a_{i_{(1-c)n}}(\alpha)) \right)^{1/cn}.
$$
Limiting Behavior

Recall $S(\alpha, n, k) = \frac{1}{\binom{n}{k}} \sum_{1 \leq i_1 < \cdots < i_k \leq n} a_{i_1} \cdots a_{i_k}$

and $S(\alpha, n, 1)^{1/1} \geq S(\alpha, n, 2)^{1/2} \geq \cdots \geq S(\alpha, n, n)^{1/n}$.

**Proposition**

For $0 < c < 1$ and for almost every $\alpha$

$$K_0 \leq \limsup_{n \to \infty} S(\alpha, n, cn)^{1/cn} \leq K_0^{1/c}(K_{-1})^{1-1/c}.$$

**Conjecture**

Almost surely $F_+^\alpha(c) = F_-^\alpha(c) = F(c)$ for all $0 < c < 1$, with

$$F_+^\alpha(c) = \limsup_{n \to \infty} S(\alpha, n, cn)^{1/cn},$$

$$F_-^\alpha(c) = \liminf_{n \to \infty} S(\alpha, n, cn)^{1/cn}.$$
Limiting Behavior

Recall

\[ F_\alpha^+(c) = \limsup_{n \to \infty} S(\alpha, n, cn)^{1/cn} \]
\[ F_\alpha^-(c) = \liminf_{n \to \infty} S(\alpha, n, cn)^{1/cn}, \]

and we conjecture \( F_\alpha^+(c) = F_\alpha^-(c) = F(c) \) a.s.

Assuming conjecture, can show that the function \( c \mapsto F(c) \) is continuous.

Assuming conjecture is false, we can show that for every \( 0 < c < 1 \) the set of limit points of the sequence \( \{ S(\alpha, n, cn)^{1/cn} \}_{n \in \mathbb{N}} \) is a non-empty interval inside \([K, K^{1/c}]\).
Evidence for Conjecture 1

\[ n \mapsto S(\alpha, n, cn)^{1/cn} \text{ for } c = \frac{1}{4}, \frac{1}{2}, \frac{3}{4} \text{ and } \alpha = \pi - 3, \gamma, \sin(1). \]
Our results on periodic continued fraction averages 1/2

For $\alpha = \sqrt{3} - 1 = [1, 2, 1, 2, 1, 2, \ldots]$,

$$\lim_{n \to \infty} S(\alpha, n, 1)^{1/1} = \frac{3}{2} \neq \infty$$

$$\lim_{n \to \infty} S(\alpha, n, n)^{1/n} = \sqrt{2} \neq K_0$$
Our results on periodic continued fraction averages 1/2

- For $\alpha = \sqrt{3} - 1 = [1, 2, 1, 2, 1, 2, \ldots]$, 
  \[
  \lim_{n \to \infty} S(\alpha, n, 1)^{1/1} = \frac{3}{2} \neq \infty \\
  \lim_{n \to \infty} S(\alpha, n, n)^{1/n} = \sqrt{2} \neq K_0
  \]

- What can we say about $\lim_{n \to \infty} S(\alpha, n, cn)^{1/cn}$?
Our results on periodic continued fraction averages 1/2

- For $\alpha = \sqrt{3} - 1 = [1, 2, 1, 2, 1, 2, \ldots]$,
  
  $$\lim_{n \to \infty} S(\alpha, n, 1)^{1/1} = \frac{3}{2} \neq \infty$$
  
  $$\lim_{n \to \infty} S(\alpha, n, n)^{1/n} = \sqrt{2} \neq K_0$$

- What can we say about $\lim_{n \to \infty} S(\alpha, n, cn)^{1/cn}$?
- Consider the quadratic irrational $\alpha = [x, y, x, y, x, y, \ldots]$. 
Our results on periodic continued fraction averages 1/2

- For $\alpha = \sqrt{3} - 1 = [1, 2, 1, 2, 1, 2, \ldots]$, 

$$
\lim_{n \to \infty} S(\alpha, n, 1)^{1/1} = \frac{3}{2} \neq \infty
$$

$$
\lim_{n \to \infty} S(\alpha, n, n)^{1/n} = \sqrt{2} \neq K_0
$$

- What can we say about $\lim_{n \to \infty} S(\alpha, n, cn)^{1/cn}$?
- Consider the quadratic irrational $\alpha = [x, y, x, y, x, y, \ldots]$.
- Let us look at $S(\alpha, n, cn)^{1/cn}$ for $c = 1/2$.

$$
S(\alpha, n, \lceil \frac{n}{2} \rceil) = \begin{cases} 
S(\alpha, n, \frac{n}{2}) & \text{if } n \equiv 0 \text{ mod } 2; \\
S(\alpha, n, \frac{n+1}{2}) & \text{if } n \equiv 1 \text{ mod } 2.
\end{cases}
$$
Our results on periodic continued fraction averages 1/2

- For $\alpha = \sqrt{3} - 1 = [1, 2, 1, 2, 1, 2, \ldots]$, 
  \[
  \lim_{n \to \infty} S(\alpha, n, 1)^{1/1} = \frac{3}{2} \neq \infty
  \]
  \[
  \lim_{n \to \infty} S(\alpha, n, n)^{1/n} = \sqrt{2} \neq K_0
  \]

- What can we say about $\lim_{n \to \infty} S(\alpha, n, cn)^{1/cn}$?
- Consider the quadratic irrational $\alpha = [x, y, x, y, x, y, \ldots]$.
- Let us look at $S(\alpha, n, cn)^{1/cn}$ for $c = 1/2$.

$$S(\alpha, n, \lceil \frac{n}{2} \rceil) = \begin{cases} 
S(\alpha, n, \frac{n}{2}) & \text{if } n \equiv 0 \mod 2; \\
S(\alpha, n, \frac{n+1}{2}) & \text{if } n \equiv 1 \mod 2.
\end{cases}$$

- We find the limit $\lim_{n \to \infty} S(\alpha, n, \lceil \frac{n}{2} \rceil)^{1/\lceil \frac{n}{2} \rceil}$ in terms of $x, y$. 
Our results on periodic continued fraction averages 2/2

**Theorem 3**

Let \( \alpha = [x, y] \). Then \( S(\alpha, n, \lceil \frac{n}{2} \rceil)^{1/\lceil \frac{n}{2} \rceil} \) converges as \( n \to \infty \) to the \( \frac{1}{2} \)-Hölder mean of \( x \) and \( y \):

\[
\lim_{n \to \infty} S(\alpha, n, \lceil \frac{n}{2} \rceil)^{1/\lceil \frac{n}{2} \rceil} = \left( \frac{x^{1/2} + y^{1/2}}{2} \right)^2.
\]
Our results on periodic continued fraction averages 2/2

**Theorem 3**

Let \( \alpha = [x, y] \). Then \( S(\alpha, n, \lceil \frac{n}{2} \rceil)^{1/\lceil \frac{n}{2} \rceil} \) converges as \( n \to \infty \) to the \( \frac{1}{2} \)-Hölder mean of \( x \) and \( y \):

\[
\lim_{n \to \infty} S(\alpha, n, \lceil \frac{n}{2} \rceil)^{1/\lceil \frac{n}{2} \rceil} = \left( \frac{x^{1/2} + y^{1/2}}{2} \right)^2.
\]

Suffices to show for \( n \equiv 0 \mod 2 \), say \( n = 2k \).

In this case we have that \( S(\alpha, 2k, k)^{1/k} \to \left( \frac{x^{1/2} + y^{1/2}}{2} \right)^2 \)
monotonically as \( k \to \infty \).
On the proof of Theorem 3, 1/2

Goal: \( \alpha = [x, y] \Rightarrow \lim_{n \to \infty} S(\alpha, n, \lceil \frac{n}{2} \rceil)^{1/[\frac{n}{2}]} = \left( \frac{x^{1/2} + y^{1/2}}{2} \right)^2. \)

The proof uses an asymptotic formula for Legendre polynomials \( P_k \) (with \( t = \frac{x}{y} < 1 \) and \( u = \frac{1+t}{1-t} > 1 \)):

\[
P_k(u) = \frac{1}{2^k} \sum_{j=0}^{k} \binom{k}{j}^2 (u - 1)^{k-j}(u + 1)^j
\]

\[
S(\alpha, 2k, k) = \frac{1}{(2k)^k} \sum_{j=0}^{k} \binom{k}{j}^2 x^j y^{k-j} = \frac{y^k}{(2k)^k} \sum_{j=0}^{k} \binom{k}{j}^2 t^j
\]

\[
= \frac{y^k}{(2k)^k} (1 - t)^k P_k(u).
\]
On the proof of Theorem 3, 2/2

Goal: \( \alpha = [x, y] \Rightarrow \lim_{n \to \infty} S(\alpha, n, \lceil n/2 \rceil)^{1/\lceil n/2 \rceil} = \left( \frac{x^{1/2} + y^{1/2}}{2} \right)^2. \)

Using the *generalized Laplace-Heine asymptotic formula* for \( P_k(u) \) for \( u > 1 \) and \( t = \frac{x}{y} < 1 \) and \( u = \frac{1+t}{1-t} > 1 \) gives

\[
S(\alpha, 2k, k)^{1/k} = y(1 - t) \left( \frac{P_k(u)}{\binom{2k}{k}} \right)^{1/k}
\]

\[
\quad \to y(1 - t) \frac{u + \sqrt{u^2 - 1}}{4} = y \left( \frac{1 + \sqrt{t}}{2} \right)^2
\]

\[
= \left( \frac{x^{1/2} + y^{1/2}}{2} \right)^2.
\]
A conjecture on periodic continued fraction averages 1/3

Expect the same result of Theorem 3 to hold for every quadratic irrational \( \alpha \) and for every \( c \).

Conjecture 2

For every \( \alpha = [x_1, \ldots, x_L] \) and every \( 0 \leq c \leq 1 \) the limit

\[
\lim_{n \to \infty} S(\alpha, n, \lfloor cn \rfloor)^{1/\lceil cn \rceil} =: F(\alpha, c)
\]

exists and it is a continuous function of \( c \).

Notice \( c \mapsto F(\alpha, c) \) is automatically decreasing by Maclaurin’s inequalities.
**Conjecture 2** for period 2 and period 3, $0 \leq c \leq 1$. 
Main Results
(Sketch of More Technical Arguments)
Explicit Formula for $F(c)$

Result of Halász and Székely yields conjecture and $F(c)$.

**Theorem 4**

If $\lim_{n \to \infty} \frac{k}{n} = c \in (0, 1]$, then for almost all $\alpha \in [0, 1]$

$$\lim_{n \to \infty} S(\alpha, n, k)^{1/k} =: F(c)$$

exists, and $F(c)$ is continuous and given explicitly by

$$c(1-c)^{\frac{1-c}{c}} \exp \left\{ \frac{1}{c} \left( (c - 1) \log r_c - \sum_{k=1}^{\infty} \log (r_c + k) \log_2 \left( 1 - \frac{1}{(k+1)^2} \right) \right) \right\},$$

where $r_c$ is the unique nonnegative solution of the equation

$$\sum_{k=1}^{\infty} \frac{r}{r + k} \log_2 \left( 1 - \frac{1}{(k+1)^2} \right) = c - 1.$$
Proof: Work of Halász and Székely

Halász and Székely calculate asymptotic properties of iid rv $\xi_1, \ldots, \xi_n$ when
- $c = \lim_{n \to \infty} k/n \in [0, 1]$.
- $\xi_j$ non-negative.
- $\mathbb{E}[\log \xi_j] < \infty$ if $c = 1$.
- $\mathbb{E}[\log(1 + \xi_j)] < \infty$ if $0 < c < 1$.
- $\mathbb{E}[\xi_j] < \infty$ if $c = 0$.

Prove $\lim_{n \to \infty} \sqrt[k]{S(\xi, n, k)} / \binom{n}{k}$ exists with probability 1 and determine it.
Proof: Work of Halász and Székely

Random variables $a_i(\alpha)$ not independent, but Halász and Székely only use independence to conclude sum of the form

$$\frac{1}{n} \sum_{k=1}^{n} f(T^k(\alpha))$$

(where $T$ is the Gauss map and $f$ is some function integrable with respect to the Gauss measure) converges a.e. to $\mathbb{E}f$ as $n \to \infty$.

Arrive at the same conclusion by appealing to the pointwise ergodic theorem.
References
References


Work supported by AMS-Simons Travel grant, NSF grants DMS0850577, DMS0970067, DMS1265673 and DMS1363227, and Williams College.
Introduction to MSTD
Joint with Peter Hegarty (Chambers), Oleg Lazarev (SMALL ’12), Kevin O’Bryant (CUNY), ...
A finite set of integers, $|A|$ its size. Form

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

Arise in Goldbach’s Problem, Twin Primes, Fermat’s Last Theorem, ....
A finite set of integers, $|A|$ its size. Form

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

Arise in Goldbach’s Problem, Twin Primes, Fermat’s Last Theorem, ....

**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.
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: \( \{0, 2, 3, 4, 7, 11, 12, 14\} \).

- 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 \( \{1, \ldots, 100\} \):
  \( \{2, 6, 7, 9, 13, 14, 16, 18, 19, 22, 23, 25, 30, 31, 33, 37, 39, 41, 42, 45, 46, 47, 48, 49, 51, 52, 54, 57, 58, 59, 61, 64, 65, 66, 67, 68, 72, 73, 74, 75, 81, 83, 84, 87, 88, 91, 93, 94, 95, 98, 100\} \).

- Recently infinite families (Hegarty, Nathanson).
Infinite Families

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

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

Proof:

- WLOG, $A = \{0, 1, \ldots, n\}$ as $A \rightarrow \alpha A + \beta$ doesn’t change $|A + A|, |A - A|$. 
- $A + A = \{0, \ldots, 2n\}, A - A = \{-n, \ldots, n\}$, 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**

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

Let

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

Then \( A \) is an MSTD set.

Note: gives \textit{exponentially} low density of MSTD sets.
New Explicit Constructions: Results and Notation

Previous best explicit sub-family of \( \{1, \ldots, n\} \) gave density of \( C_1 n^d / 2^{n/2} \).

Our new family gives \( C_2 / n^2 \), almost a positive percent.

Current record by Zhao: \( C_3 / n \).

Notation:

- \( [a, b] = \{k \in \mathbb{Z} : a \leq k \leq b\} \).

- \( A \) is a \( P_n \)-set if its sumset and difference sets contain all but the first and last \( n \) possible elements (may or may not contain some of these fringe elements).
New Construction

Theorem (Miller-Orosz-Scheinerman ’09)

- \( A = L \cup R \) be a \( P_n \), MSTD set where \( L \subseteq [1, n] \), \( R \subseteq [n + 1, 2n] \), 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 \( \{0, \ldots, r\} \) that are in this family (and thus are MSTD sets) is at least \( C/r^2 \).
Phase Transition
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: \( \mathbb{E}[X] = \int x f(x) \, dx. \)
- Variance: \( \sigma^2 = \int (x - \mathbb{E}[X])^2 f(x) \, dx. \)
Binomial model

Binomial model, parameter $p(n)$

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

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

- Let $A \in \{0, \ldots, n\}$. Most elements in $\{0, \ldots, 2n\}$ in $A + A$ and in $\{-n, \ldots, n\}$ in $A - A$.

- $\mathbb{E}[|A + A|] = 2n - 11$, $\mathbb{E}[|A - A|] = 2n - 7$. 
Theorem

Let $A$ be chosen from $\{0, \ldots, N\}$ according to the binomial model with constant parameter $p$ (thus $k \in A$ with probability $p$). At least $k_{SD;p}2^{N+1}$ subsets are sum dominated.
Martin and O’Bryant ’06

**Theorem**

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

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

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

- $k_{SD;1/2} \geq 10^{-7}$, 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 $\forall N \geq N_{\epsilon_1, \epsilon_2}$

$$\text{Prob} \left( X \not\in \left[ (1 - \epsilon_1)f(N), (1 + \epsilon_1)f(N) \right] \right) < \epsilon_2.$$
\[ X \sim f(N) \text{ means } \forall \epsilon_1, \epsilon_2 > 0, \exists N_{\epsilon_1, \epsilon_2} \text{ st } \forall N \geq N_{\epsilon_1, \epsilon_2} \]

\[ \text{Prob} (X \not\in [(1 - \epsilon_1)f(N), (1 + \epsilon_1)f(N)]) < \epsilon_2. \]

\[ S = |A + A|, \ D = |A - A|, \]
\[ S^c = 2N + 1 - S, \ D^c = 2N + 1 - D. \]
Notation

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

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

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

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

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

Can generalize to binary linear forms or arbitrarily many summands, still have **critical threshold**.
Key input: recent strong concentration results of Kim and Vu (Applications: combinatorial number theory, random graphs, ...).

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

Need to allow dependent random variables.

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

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

Only need new 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

Only need new 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:

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

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

Lemma

\[ P_1(N) = 4N^{-(1-\delta)}, \quad \mathcal{O} = \#\{(m, n) : m < n \in \{1, \ldots, N\} \cap A\}. \]

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

1. \(X \in \left[\frac{1}{2}N^{1-\delta}, \frac{3}{2}N^{1-\delta}\right].\)

2. \(\frac{1}{2}N^{1-\delta}\left(\frac{1}{2}N^{1-\delta} - 1\right) \leq \mathcal{O} \leq \frac{3}{2}N^{1-\delta}\left(\frac{3}{2}N^{1-\delta} - 1\right).\)
Proof

**Lemma**

\[
P_1(N) = 4N^{-1-\delta}, \quad O = \#\{(m, n) : m < n \in \{1, \ldots, N\} \cap A\}.
\]

*With probability at least* \(1 - P_1(N)\) *have*

1. \(X \in \left[\frac{1}{2} N^{1-\delta}, \frac{3}{2} N^{1-\delta}\right].\)
2. \(\frac{1}{2} N^{1-\delta} \left(\frac{1}{2} N^{1-\delta} - 1\right) \leq O \leq \frac{3}{2} N^{1-\delta} \left(\frac{3}{2} N^{1-\delta} - 1\right).\)

**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

\begin{itemize}
  \item \( f(\delta) = \min \left( \frac{1}{2}, \frac{3\delta - 1}{2} \right) \), \( g(\delta) \) satisfies \( 0 < g(\delta) < f(\delta) \).
  \item \( p(N) = N^{-\delta}, \ \delta \in (1/2, 1), \ P_1(N) = 4N^{-(1-\delta)}, \ P_2(N) = CN^{-(f(\delta)-g(\delta))} \).
\end{itemize}

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

Lemma

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

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.

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

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

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

$$\mathbb{E}[Y] \leq N^3 \cdot N^{-4\delta} + N^2 \cdot N^{-3\delta} \leq 2N^{3-4\delta}. \text{ As } \delta > 1/2,$$

$$\#\{\text{bad pairs}\} \ll O.$$ 

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.

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 $n = m'$).

$$\text{Var}(U) = \sum \text{Var}(U_{m,n,m',n'}) + 2 \sum_{(m,n,m',n') \neq (\tilde{m},\tilde{n},\tilde{m}',\tilde{n}')} \text{CoVar}(U_{m,n,m',n'}, U_{\tilde{m},\tilde{n},\tilde{m}',\tilde{n}'})$$
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 $\ll N^{5-7\delta} + N^{3-4\delta}$. 
Ongoing Research
Current and Open Problems

- Similar results for arbitrary finite groups (with Kevin Vissuet).

- Generalize phase transition results for more summands (SMALL ’13 hopefully).

- Generalize to subsets of $\mathbb{Z}^+ \times \mathbb{Z}^+$ (SMALL ’13 hopefully).

- Study the dependence of the pivot on $p(N)$.
Let $m(k)$ be the probability a uniformly drawn subset $A$ of $[0, n]$ has $A + A$ missing exactly $k$ summands as $n \to \infty$.

**Figure:** Experimental values of $m(k)$, vertical bars error (often smaller than dot!).

What happens if draw $A$ from binomial with parameter $p(N)$?
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 $D_f$ denote the random variable $|f(A)|$. Then the following three situations arise:

1. $p(N) = o(N^{-1/2})$ : Then
   
   $D_f \sim (N \cdot p(N))^2$.

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

   Then
   
   $D_f \sim g_{u,v} \left( \frac{c^2}{u} \right) N$.

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


Generalization of Hegarty-Miller

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

Theorem (Hegarty-M): $f(x, y) = u_1 x + u_2 y$ and $g(x, y) = u_2 x + g_2 y$, 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 $p(N) > cN^{-1/2}$ ($c > c_{f,g}$).
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).
Bibliography

Bibliography (cont)


