More Sums Than Differences Sets

Thao Do, Stony Brook University
Archit Kulkarni, Carnegie Mellon University
David Moon, Williams College
Jake Wellens, California Institute of Technology

Advisor: Steven J. Miller
SMALL REU, Williams College

SMALLTalks, 2013
Williams College

Introduction

Introduction

Given $A \subset \mathbb{Z}$, let

$$A + A = \{a_1 + a_2 : a_1, a_2 \in A\},$$
$$A - A = \{a_1 - a_2 : a_1, a_2 \in A\}.$$

Introduction

Given $A \subset \mathbb{Z}$, let

$$A + A = \{a_1 + a_2 : a_1, a_2 \in A\},$$
$$A - A = \{a_1 - a_2 : a_1, a_2 \in A\}.$$

Theorem

For n large, the proportion of sets $A \subset \{0, \ldots, n\}$ with $|A + A| > |A - A|$ is greater than 2×10^{-7}. (Martin and O'Bryant 2006)

Introduction

Given $A \subset \mathbb{Z}$, let

$$A + A = \{a_1 + a_2 : a_1, a_2 \in A\},$$
$$A - A = \{a_1 - a_2 : a_1, a_2 \in A\}.$$

Theorem

For n large, the proportion of sets $A \subset \{0, \dots, n\}$ with $|A + A| > |A - A|$ is greater than 2×10^{-7}. (Martin and O'Bryant 2006)

Such sets are called *More Sums Than Differences (MSTD) sets*, or *sum-dominant sets*.

Fringe Elements

Fringe Elements

Fringe elements are important.

Fringe Elements

Fringe elements are important.

Example

Given $A \subset [0, n]$ and $k \leq n$,

$$(A + A) \cap [0, k] \ \subset \ (A \cap [0, k]) + (A \cap [0, k]).$$

Fringe Elements

Fringe elements are important.

Example

Given $A \subset [0, n]$ and $k \leq n$,

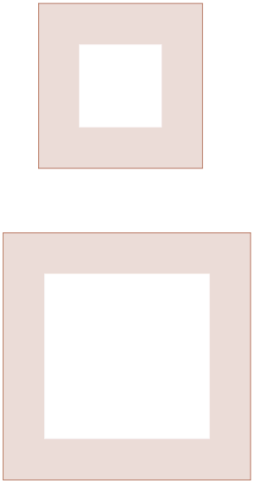
$$(A + A) \cap [0, k] \ \subset \ (A \cap [0, k]) + (A \cap [0, k]).$$

Martin and O'Bryant: for some fixed k, carefully choose fringe of A so that

$$\begin{aligned} &|\,(A + A) \cap ([0, k] \cup [2n - k, 2n])\,| \\ > &|\,(A - A) \cap ([-n, -n + k] \cup [n - k, n])\,| \end{aligned}$$

With positive probability, $[k + 1, 2n - k - 1] \subset A + A$ and A is MSTD.

Fringe in Higher Dimensions



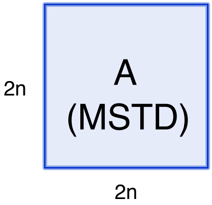
Positive Percentage in d Dimensions

Theorem

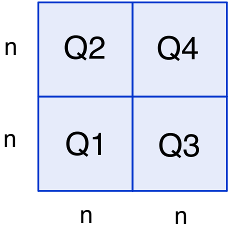
For any integer $d > 0$, there exists some constant $c_d > 0$ such that, for n large, the proportion of MSTD subsets A of $\{0, \ldots, n\}^d$ is greater than c_d.

Proof of positive proportion of MSTD sets is probabilistic.

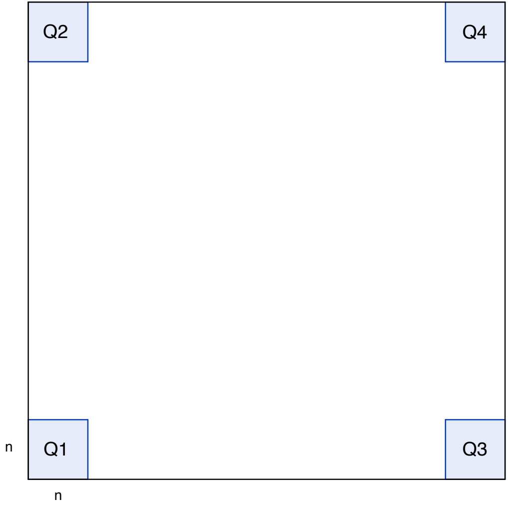
Construction of Infinite Families of MSTD Sets



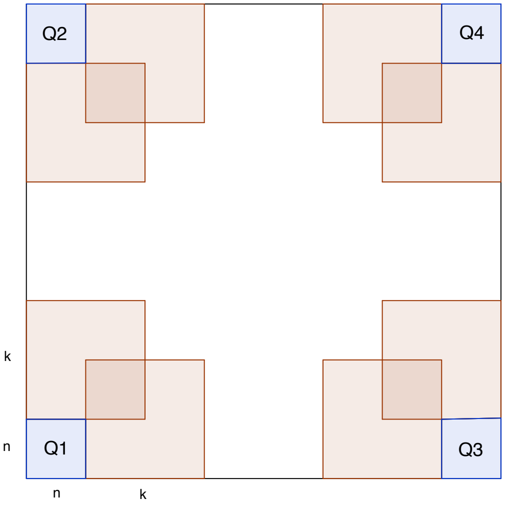
Construction of Infinite Families of MSTD Sets



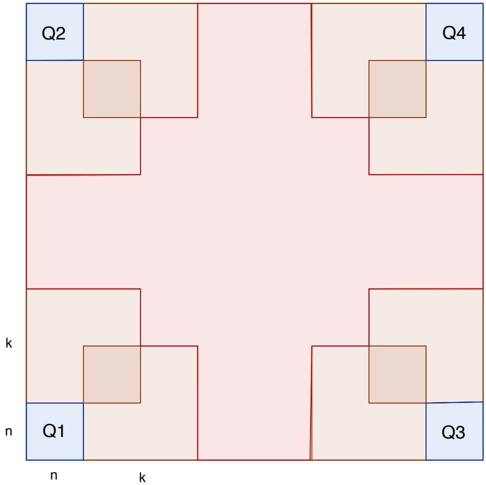
Construction of Infinite Families of MSTD Sets



Construction of Infinite Families of MSTD Sets



Construction of Infinite Families of MSTD Sets



Correlated Random Pairs

All of the literature to date has looked at sums and differences of a set *with itself*. We investigate sums and differences of *pairs* of subsets $(A, B) \subset \{0, \dots, n\}$. We select such pairs according to the dependent random process:

$$P(a \in A) = p; \quad P(a \in B | a \in A) = \rho_1; \quad P(a \in B | a \notin A) = \rho_2$$

Correlated Random Pairs

All of the literature to date has looked at sums and differences of a set *with itself*. We investigate sums and differences of *pairs* of subsets $(A, B) \subset \{0, \dots, n\}$. We select such pairs according to the dependent random process:

$$P(a \in A) = p; \quad P(a \in B | a \in A) = \rho_1; \quad P(a \in B | a \notin A) = \rho_2$$

Let $\vec{\rho} = (p, \rho_1, \rho_2)$. We call a pair of subsets selected by this process a $\vec{\rho}$-*correlated pair*. Note that when $(\rho_1, \rho_2) = (1, 0)$, this is the old case of (A, A). When $(\rho_1, \rho_2) = (0, 1)$, this is (A, A^c). When $\rho_1 = \rho_2$, A and B are independent.

Correlated Random Pairs

Let $P(\vec{\rho}, n)$ be the probability that a $\vec{\rho}$-correlated pair (A, B) with $A, B \subset \{0, \dots, n\}$ is MSTD, that is

$$|A + B| > |\pm(A - B)| = |(A - B) \cup (B - A)|$$

Correlated Random Pairs

Let $P(\vec{\rho}, n)$ be the probability that a $\vec{\rho}$-correlated pair (A, B) with $A, B \subset \{0, \dots, n\}$ is MSTD, that is

$$|A + B| > |\pm(A - B)| = |(A - B) \cup (B - A)|$$

Theorem

For any $\vec{\rho} \in [0, 1]^3$, the limit

$$\lim_{n \to \infty} P(\vec{\rho}, n) =: P(\vec{\rho})$$

exists. Moreover, as long as $p \notin \{0, 1\}$ and $(\rho_1, \rho_2) \neq (0, 0), (1, 1)$, then $P(\vec{\rho})$ is strictly positive.

The function $P(\vec{\rho})$

Theorem

The function $P(\vec{\rho})$ is continuous on $[0,1]^3$.

Maximizing the probability of sum dominance

As $P(\vec{\rho})$ is a continuous function on a compact set $[0,1]^3$, it must attain a maximum.

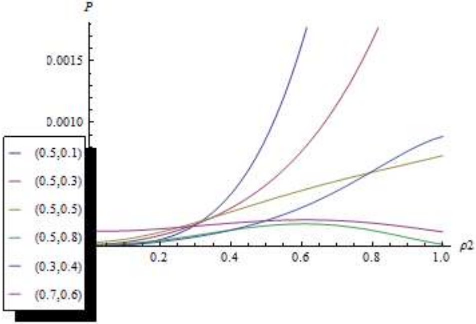
Maximizing the probability of sum dominance

As $P(\vec{\rho})$ is a continuous function on a compact set $[0, 1]^3$, it must attain a maximum.

Here we fix $n = 9$ and investigate how the percentage P changes when we vary p, ρ_1, ρ_2 and see where it is maximized.

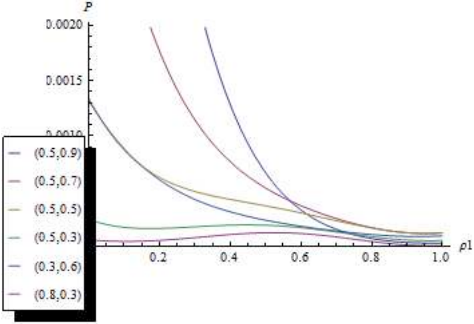
We find the maximum percentage at point $(0.5, 0, 1)$.

Fix (p, ρ_1)



Conjecture 1: For any fixed (p, ρ_1) with ρ_1 not too big ($\rho_1 \leq .4$) then P as a function of ρ_2 is strictly increasing in $[0, 1]$ and reaches its maximum at $\rho_2 = 1$.

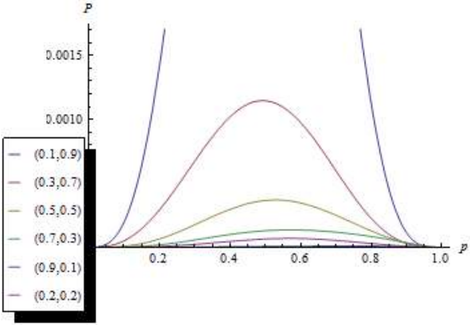
Fix (p, ρ_2)



Conjecture 2: For any fixed (p, ρ_2) with ρ_2 not too small $(\rho_2 \geq .5)$ then P as a function of ρ_1 is strictly decreasing in $[0, 1]$ and reaches its maximum at $\rho_1 = 0$.

Fix (ρ_1, ρ_2)

$n = 9$: If we fix (ρ_1, ρ_2), P as a function of p has a shape similar to parabola with a maximum at a point around $1/2$.



Conjecture 3: The maximum of function $P(p, \rho_1, \rho_2)$ is at $P(1/2, 0, 1) \approx 0.03$.

The minimal MSTD pair

Hegarty (2007) proved the smallest MSTD set has size 8.

The minimal MSTD pair

Hegarty (2007) proved the smallest MSTD set has size 8. We prove

Theorem

The smallest MSTD pair has size $(3, 5)$ or $(4, 4)$.

Examples:

$$A = \{1, 2, 5, 7\}, \quad B = \{1, 3, 6, 7\}$$
$$A = \{3, 4, 6\}, \quad B = \{1, 2, 5, 7, 8\}$$
$$A = \{3, 5, 6\}, \quad B = \{1, 2, 4, 7, 8\}.$$

References

P. Hegarty, Some explicit constructions of sets with more sums than differences. Acta Arithmetica 130 (2007), no.1, 61-77.

G. Martin and K. O'Bryant, Many sets have more sums than differences, Additive Combinatorics, CRM Proc. Lecture Notes, vol. 43, Amer. Math. Soc., Providence, RI, 2007, pp. 287-305

S. J. Miller, B. Orosz, and D. Scheinerman, Explicit constructions of infinite families of MSTD sets, J. Number Theory **130** (2010), no. 5, 1221-1233.

Y. Zhao, Sets Characterized by Missing Sums and Differences, *Journal of Number Theory*, 131 (2010), pp. 2107-2134.

Distribution of Summands in Generalized Zeckendorf Decompositions

Philippe Demontigny, Williams College
Archit Kulkarni, Carnegie Mellon University
Umang Varma, Kalamazoo College

(with Thao Do and David Moon)

Advisor: Steven J Miller
SMALL REU, Williams College

Zeckendorf's Theorem

- Our research is inspired by an elegant theorem of Zeckendorf.

Theorem

Write the Fibonacci numbers as $F_1 = 1$, $F_2 = 2$, $F_n = F_{n-1} + F_{n-2}$ for $n > 2$. All natural numbers can be uniquely written as a sum of non-consecutive Fibonacci numbers.

Example

$$2013 = 1597 + 377 + 34 + 5 = F_{16} + F_{13} + F_8 + F_4$$

Going the other way

Previous work:
linear recurrence sequence $\rightarrow$ notion of legal decomposition.

Our work:
notion of legal decompostition $\rightarrow$ linear recurrence sequence.

f-Decompositions

- We focused on constructing sequences from notions of legal decomposition.

f-Decompositions

- We focused on constructing sequences from notions of legal decomposition.
- Many notions of "legal" decompositions can be encoded as f-decompositions.

Definition

Let $f : \mathbb{N} \to \mathbb{N}$. A sum $\sum_{i=0}^{k} a_{n_i}$ of terms of $\{a_n\}$ is a *legal f-decomposition using* $\{a_n\}$ if for every a_{n_i}, the previous $f(n_i)$ terms are not in the f-decomposition.

Result

Theorem

If $f(n+1) \leq 1 + f(n)$ for all $n \in \mathbb{N}$, there exists a sequence $\{a_n\}$ such that every positive integer has a unique legal f-decomposition using $\{a_n\}$.

Example

The Zeckendorf condition is that consecutive terms may not be chosen.
This is equivalent to saying $f(n) = 1$ for all $n \in \mathbb{N}$.
This condition yields the Fibonacci numbers.

$$\{F_n\} = 1, 2, 3, 5, 8, 13, 21, 34, \ldots$$

Bins

- Base b representation can be interpretted as f-decompositions. For example, consider base 5:

$$\{a_n\} = \underbrace{1, 2, 3, 4}, \underbrace{5, 10, 15, 20}, \underbrace{25, 50, 75, 100}, \dots$$

Here, $a_n = 5a_{n-4}$.

- We build upon this notion of legal decomposition by adding the Zeckendorf condition. A legal decomposition is one that contains no consecutive terms and at most one term from each bin.

$$\{a_n\} = \underbrace{1, 2, 3}, \underbrace{4, 7, 11}, \underbrace{15, 26, 41}, \underbrace{56, 97, 153}, \underbrace{209, 362, 571}, \dots$$

Here, $a_n = 4a_{n-3} - a_{n-6}$. We analyze this case in detail.

Number of summands

Our goal is to show that the number of summands for integers in $[0, a_{bn})$ converges to the Gaussian distribution as $n \to \infty$.

- Let $p_{n,k}$ be the number of integers that can be legally written as the sum of exactly k summands from n bins.

Number of summands

Our goal is to show that the number of summands for integers in $[0, a_{bn})$ converges to the Gaussian distribution as $n \to \infty$.

- Let $p_{n,k}$ be the number of integers that can be legally written as the sum of exactly k summands from n bins.
- We prove that

$$p_{n,k} = p_{n-1,k} + bp_{n-1,k-1} - p_{n-2,k-2}$$

Number of summands

Our goal is to show that the number of summands for integers in $[0, a_{bn})$ converges to the Gaussian distribution as $n \to \infty$.

- Let $p_{n,k}$ be the number of integers that can be legally written as the sum of exactly k summands from n bins.
- We prove that

$$p_{n,k} = p_{n-1,k} + bp_{n-1,k-1} - p_{n-2,k-2}$$

- Define $g_n(y) = \sum_{k \geq 0} p_{n,k} y^k$. We were able to show that

$$g_n(y) = \frac{\left(by+1+\sqrt{(b^2-4)y^2+2by+1}\right)^{n+1} - \left(by+1-\sqrt{(b^2-4)y^2+2by+1}\right)^{n+1}}{2^{n+1}\sqrt{(b^2-4)y^2+2by+1}}$$

Mean and Variance

- We can use $g_n(y) = \sum_{k\geq 0} p_{n,k} y^k$ to compute mean and variance of the random variable X_n, the number of summands for integers in $[a_0, a_{bn})$.

Mean and Variance

- We can use $g_n(y) = \sum_{k\geq 0} p_{n,k} y^k$ to compute mean and variance of the random variable X_n, the number of summands for integers in $[a_0, a_{bn})$.
- The mean, μ_n, is simply

$$\mu_n = \frac{g_n'(1)}{g_n(1)} = \frac{\left(b^2+b-4+b\sqrt{b^2+2b-3}\right)}{\sqrt{b^2+2b-3}\left(1+b+\sqrt{b^2+2b-3}\right)} n + O(1)$$

Mean and Variance

- We can use $g_n(y) = \sum_{k\geq 0} p_{n,k} y^k$ to compute mean and variance of the random variable X_n, the number of summands for integers in $[a_0, a_{bn})$.
- The mean, μ_n, is simply

$$\mu_n = \frac{g_n'(1)}{g_n(1)} = \frac{\left(b^2+b-4+b\sqrt{b^2+2b-3}\right)}{\sqrt{b^2+2b-3}\left(1+b+\sqrt{b^2+2b-3}\right)} n + O(1)$$

- The variance is

$$\sigma_n^2 = \frac{\frac{d}{dy}\left[y g_n'(y)\right]\Big|_{y=1}}{g(1)} - \mu^2 = \frac{(b^2+b-4)n}{(b^2+2b-3)^{3/2}} + O(1)$$

Moment Generating Function

- If we normalize X_n to $Y_n = (X_n - \mu_n)/\sigma_n$, the moment generating function of Y_n is

$$M_{Y_n}(t) = \mathbb{E}(e^{tY_n}) = \sum_{k\geq 0} \frac{p_{n,k}e^{\frac{t(k-\mu_n)}{\sigma_n}}}{\sum_{k\geq 0} p_{n,k}} = \frac{g_n(e^{t/\sigma_n})e^{-t\mu_n/\sigma_n}}{g_n(1)}$$

Moment Generating Function

- If we normalize X_n to $Y_n = (X_n - \mu_n)/\sigma_n$, the moment generating function of Y_n is

$$M_{Y_n}(t) = \mathbb{E}(e^{tY_n}) = \sum_{k\geq 0} \frac{p_{n,k} e^{\frac{t(k-\mu_n)}{\sigma_n}}}{\sum_{k\geq 0} p_{n,k}} = \frac{g_n(e^{t/\sigma_n})e^{-t\mu_n/\sigma_n}}{g_n(1)}$$

- After multiple Taylor Series expansions, we get

$$\log(M_{Y_n}(t)) = \frac{t^2}{2} + O\left(\frac{t^3}{\sqrt{n}}\right)$$

as $n \to \infty$.

Moment Generating Function

- If we normalize X_n to $Y_n = (X_n - \mu_n)/\sigma_n$, the moment generating function of Y_n is

$$M_{Y_n}(t) = \mathbb{E}(e^{tY_n}) = \sum_{k\geq 0} \frac{p_{n,k}\, e^{\frac{t(k-\mu_n)}{\sigma_n}}}{\sum_{k\geq 0} p_{n,k}} = \frac{g_n(e^{t/\sigma_n})e^{-t\mu_n/\sigma_n}}{g_n(1)}$$

- After multiple Taylor Series expansions, we get

$$\log(M_{Y_n}(t)) = \frac{t^2}{2} + O\left(\frac{t^3}{\sqrt{n}}\right)$$

as $n \to \infty$.

- Hence, the distribution of Y_n converges to the standard normal distribution.

Far-Difference Representations

Definition

A far-difference representation is a sum of numbers and their additive inverses.

Far-Difference Representations

Definition

A far-difference representation is a sum of numbers and their additive inverses.

Example: $65 = 89 - 21 - 3 = F_{10} - F_7 - F_3$

- Alpert [Al] proved that a Fibonacci far-difference representation exists for all integers and is unique.

Far-Difference Representations

Definition

A far-difference representation is a sum of numbers and their additive inverses.

Example: $65 = 89 - 21 - 3 = F_{10} - F_7 - F_3$

- Alpert [Al] proved that a Fibonacci far-difference representation exists for all integers and is unique.
- Miller-Wang [MW], a paper from a previous SMALL summer with countably infinite pages, proved Gaussianity for Alpert's far-difference representations.

Preliminary Definitions

The k-Skipponaccis are recurrence relations of the form

$$S_{n+1} = S_n + S_{n-k}$$

for some $k \geq 0$.

Preliminary Definitions

The k-Skipponaccis are recurrence relations of the form

$$S_{n+1} = S_n + S_{n-k}$$

for some $k \geq 0$.

Alpert [Al] proved the following result for the Fibonaccis (also called the 1-Skipponaccis).

Alpert's Theorem

Every $x \in \mathbb{Z}$ has a unique far-difference representation for the Fibonaccis such that all terms of the same sign are at least 4 apart in index, and all terms of opposite sign are at least 3 apart in index.

Our First Result

Example: $119 = 144 - 34 + 8 + 1 = \underbrace{F_{11} - F_8}_{\text{3 apart}} + \underbrace{F_5 + F_1}_{\text{4 apart}}$

Our First Result

Example: $119 = 144 - 34 + 8 + 1 = \underbrace{F_{11} - F_8}_{\text{3 apart}} + \underbrace{F_5 + F_1}_{\text{4 apart}}$

Theorem 1

Every $x \in \mathbb{Z}$ has a unique far-difference representation for the k-Skipponaccis such that all terms of the same sign are at least $2k + 2$ apart in index and all terms of opposite sign are at least $k + 2$ apart in index.

Our Second Result

$$\text{Let } R(n) = \sum_{0 < n - b(2k+2) \leq n} S_{n-b(2k+2)} = S_n + S_{n-2k-2} + S_{n-4k-4} + \dots$$

Our Second Result

Let $R(n) = \sum_{0 < n - b(2k+2) \leq n} S_{n-b(2k+2)} = S_n + S_{n-2k-2} + S_{n-4k-4} + \dots$

Theorem 2

Let $\mathcal{K}_n$ and $\mathcal{L}_n$ be random variables denoting the number of positive and negative summands in the far-difference representation of integers on the interval $(R_{n-1}, R_n]$. As $n \to \infty$, the joint density of $\mathcal{K}_n$ and $\mathcal{L}_n$ converges to a bivariate Gaussian.

Our Second Result

$$\text{Let } R(n) = \sum_{0<n-b(2k+2)\leq n} S_{n-b(2k+2)} = S_n + S_{n-2k-2} + S_{n-4k-4} + \dots$$

Theorem 2

Let $\mathcal{K}_n$ and $\mathcal{L}_n$ be random variables denoting the number of positive and negative summands in the far-difference representation of integers on the interval $(R_{n-1}, R_n]$. As $n \to \infty$, the joint density of $\mathcal{K}_n$ and $\mathcal{L}_n$ converges to a bivariate Gaussian.

- This theorem expands upon the range of recurrences handled by Miller-Wang [MW].

Further Research and Open Questions

- We have further generalized the k-Skipponaccis to recurrence relations of the form $S_n = S_{n-1} + S_{n-x} + S_{n-y}$, where x is the distance between same-sign summands and y is the distance between opposite-sign summands.

Further Research and Open Questions

- We have further generalized the k-Skipponaccis to recurrence relations of the form $S_n = S_{n-1} + S_{n-x} + S_{n-y}$, where x is the distance between same-sign summands and y is the distance between opposite-sign summands.

  Example: The Fibonacci recurrence can be written as:
  $F_n \;=\; F_{n-1} + F_{n-2} \;=\; F_{n-1} + (F_{n-3} + F_{n-4})$

Further Research and Open Questions

- We have further generalized the k-Skipponaccis to recurrence relations of the form $S_n = S_{n-1} + S_{n-x} + S_{n-y}$, where x is the distance between same-sign summands and y is the distance between opposite-sign summands.

  Example: The Fibonacci recurrence can be written as:
  $F_n \;=\; F_{n-1} + F_{n-2} \;=\; F_{n-1} + (F_{n-3} + F_{n-4})$

- We believe it can be shown that every far-difference restriction (x, y) uniquely defines a sequence of numbers.

Further Research and Open Questions

- We have further generalized the k-Skipponaccis to recurrence relations of the form $S_n = S_{n-1} + S_{n-x} + S_{n-y}$, where x is the distance between same-sign summands and y is the distance between opposite-sign summands.

  Example: The Fibonacci recurrence can be written as:
  $F_n \;=\; F_{n-1} + F_{n-2} \;=\; F_{n-1} + (F_{n-3} + F_{n-4})$

- We believe it can be shown that every far-difference restriction (x, y) uniquely defines a sequence of numbers.

- We want to prove that the number of summands in *every* (x, y) far-difference representations approaches a Gaussian.

References

- H. Alpert, *Differences of Multiple Fibonacci Numbers*, Integers: Electronic Journal of Combinatorial Number Theory **9** (2009), 745-749.
- Kologlu, Kopp, Miller and Wang: *Gaussianity for Fibonacci case*, Fibonacci Quarterly **49** (2011), no. 2, 116–130.
- S. J. Miller and Y. Wang, *From Fibonacci numbers to Central Limit Type Theorems*, Journal of Combinatorial Theory, Series A **119** (2012), no. 7, 1398–1413.

Primes and L-functions

Alan Chang, Princeton University
Kyle Pratt, Brigham Young University
Minh-Tam Trinh, Princeton University

Advisors:
Steven J. Miller
Julio Cesar Bueno de Andrade

SMALL 2013 REU, Williams College

Motivation: Random Matrix Theory

- *L*-functions are functions on the complex plane that generalize the Riemann zeta function:

$$L(s) = \sum_{n=1}^{\infty} a_n n^{-s}$$

Motivation: Random Matrix Theory

- L-functions are functions on the complex plane that generalize the Riemann zeta function:

$$L(s) = \sum_{n=1}^{\infty} a_n n^{-s}$$

- $L(s)$ satsifies a Riemann Hypothesis iff all its zeros in the region $\Re s \in [0, 1]$ live on the line $\Re s = 1/2$.

Motivation: Random Matrix Theory

- L-functions are functions on the complex plane that generalize the Riemann zeta function:

$$L(s) = \sum_{n=1}^{\infty} a_n n^{-s}$$

- $L(s)$ satsifies a Riemann Hypothesis iff all its zeros in the region $\Re s \in [0, 1]$ live on the line $\Re s = 1/2$.

- Montgomery-Dyson, Katz-Sarnak: Spacing statistics of zeros match spacing statistics of angles of eigenvalues of a random matrix.

Zero Statistics

- Possible statistics: Correlated density of zeros (n-level density), distribution of spacings between zeros (pair correlation), moments along the critical line.

Zero Statistics

- Possible statistics: Correlated density of zeros (*n*-level density), distribution of spacings between zeros (pair correlation), moments along the critical line.

- Katz-Sarnak: *L*-functions form "families" $\mathcal{F} = \bigsqcup_N \mathcal{F}_N$, and the average zero statistics over $\mathcal{F}_N$ approach the eigenangle statistics in some matrix group $G_{\mathcal{F}}$ as $N \to \infty$.

Zero Statistics

- Possible statistics: Correlated density of zeros (n-level density), distribution of spacings between zeros (pair correlation), moments along the critical line.

- Katz-Sarnak: L-functions form "families" $\mathcal{F} = \bigsqcup_N \mathcal{F}_N$, and the average zero statistics over $\mathcal{F}_N$ approach the eigenangle statistics in some matrix group $G_{\mathcal{F}}$ as $N \to \infty$.

$$G = \mathrm{O}(N), \mathrm{U}(N), \mathrm{USp}(2N), \mathrm{SO(even)}(2N), \mathrm{SO(odd)}(2N+1)$$

Testing Katz-Sarnak Density Conjectures

- To study zero statistics, “bombard” the zeros with a test function ϕ whose Fourier transform has compact support.

Testing Katz-Sarnak Density Conjectures

- To study zero statistics, “bombard” the zeros with a test function ϕ whose Fourier transform has compact support.
- Often, too hard unless we assume $\operatorname{supp} \hat{\phi} \subseteq [-\sigma, +\sigma]$ for some fixed σ.

Testing Katz-Sarnak Density Conjectures

- To study zero statistics, “bombard” the zeros with a test function ϕ whose Fourier transform has compact support.
- Often, too hard unless we assume $\operatorname{supp}\hat{\phi} \subseteq [-\sigma, +\sigma]$ for some fixed σ.
- Need to estimate very hard sums over primes to increase the support.

Iwaniec, Luo, and Sarnak (ILS)

- Studied family of cuspidal weight-k, level-N holomorphic newforms as $N \to \infty$.

Iwaniec, Luo, and Sarnak (ILS)

- Studied family of cuspidal weight-k, level-N holomorphic newforms as $N \to \infty$.
- By getting support in the range $(-2, +2)$, able to distinguish L-functions according to sign.

Iwaniec, Luo, and Sarnak (ILS)

- Studied family of cuspidal weight-k, level-N holomorphic newforms as $N \to \infty$.
- By getting support in the range $(-2, +2)$, able to distinguish L-functions according to sign.
- To extend support of test function beyond $(-2, +2)$, need to assume a conjecture called Hypothesis S.

Hypothesis S

Hypothesis S

Let c be a positive integer, and let a be coprime to c. Then for $A \geq 0$ and some $\alpha \in [1/2, 3/4)$, we have

$$\sum_{\substack{p \leq X \\ p \equiv a \bmod c}} e^{2\pi i \left(\frac{2\sqrt{p}}{c}\right)} \ll c^A X^\alpha.$$

- Remark: $f(x) \ll g(x)$ means that $|f(x)| \leq k|g(x)|$ for some k and for all $x \geq x_0$.

Hypothesis S

Hypothesis S

Let c be a positive integer, and let a be coprime to c. Then for $A \geq 0$ and some $\alpha \in [1/2, 3/4)$, we have

$$\sum_{\substack{p \leq X \\ p \equiv a \bmod c}} e^{2\pi i\left(\frac{2\sqrt{p}}{c}\right)} \ll c^A X^\alpha.$$

- Remark: $f(x) \ll g(x)$ means that $|f(x)| \leq k|g(x)|$ for some k and for all $x \geq x_0$.
- Question: Why should we expect this sum to be smaller than X?

Philosophy of "Square-Root Cancellation"

- Take sequence of real numbers $a_1, \ldots, a_N$.

Philosophy of “Square-Root Cancellation”

- Take sequence of real numbers $a_1, \ldots, a_N$.
- If numbers are “random” we expect fractional parts are evenly distributed mod 1.

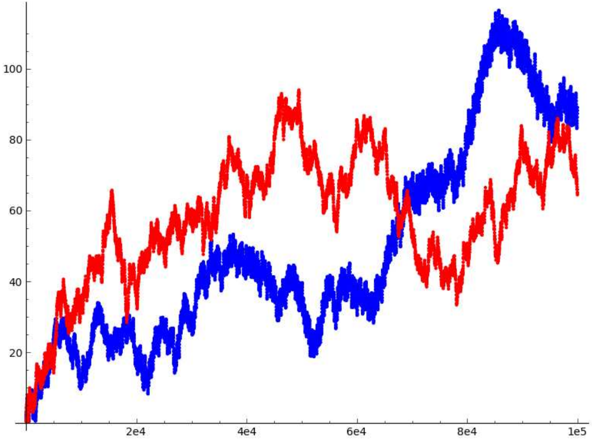
Philosophy of "Square-Root Cancellation"

- Take sequence of real numbers $a_1, \ldots, a_N$.
- If numbers are "random" we expect fractional parts are evenly distributed mod 1.
- If we have $\sum_{i=1}^{N} e^{2\pi i a_i}$, expect terms to point randomly in different directions, so that $\sum_{i=1}^{N} e^{2\pi i a_i} \ll N^{1/2}$ (essentially Central Limit Theorem).

Philosophy of "Square-Root Cancellation"

- Take sequence of real numbers $a_1, \dots, a_N$.
- If numbers are "random" we expect fractional parts are evenly distributed mod 1.
- If we have $\sum_{i=1}^N e^{2\pi i a_i}$, expect terms to point randomly in different directions, so that $\sum_{i=1}^N e^{2\pi i a_i} \ll N^{1/2}$ (essentially Central Limit Theorem).
- Expect similar thing to happen for Hypothesis S.

Pretty Picture



- Sum over Primes, Sum over Random Real Numbers

Cracking Hypothesis S

- Plan: relate information about primes to information about zeros of L-functions.

Cracking Hypothesis S

- Plan: relate information about primes to information about zeros of L-functions.
- When $c = 1$ the L-function is Riemann zeta function; same ideas work for all c.

Cracking Hypothesis S

- Plan: relate information about primes to information about zeros of L-functions.
- When $c = 1$ the L-function is Riemann zeta function; same ideas work for all c.
- Assume Riemann Hypothesis to get control over zeros.

First Steps

- Do fancy tricks to make sum nicer to work with.

First Steps

- Do fancy tricks to make sum nicer to work with.

Riemann-von Mangoldt Explicit Formula

Let Λ be the von Mangoldt function (like indicator for prime powers). Then

$$\sum_{n\leq x}\Lambda(n) = x - \sum_{|\gamma|<T}\frac{x^\rho}{\rho} + O\left(\frac{x\log^2(xT)}{T} + \log x\right),$$

where sum is over zeros of Riemann zeta function.

It's (Probably) True

Theorem (A-C-M-P-T)

Assume RH and that $\frac{2}{\pi e}\gamma_n \log(\frac{\gamma_n}{2\pi e})$ *is well distributed mod 1 (*γ_n *is ordinate of nth zeta zero). Then Hypothesis S is true for* $c = 1$ *with* $\alpha = 3/4 - \epsilon$ *with* ϵ *very small. That is,*

$$\sum_{p \leq X} e^{2\pi i(2\sqrt{p})} \ll X^{3/4-\epsilon}.$$

Hypothesis T

- Need 2-dimensional analogue of Hypothesis S for extending support in 2-level density (just a zero statistic).

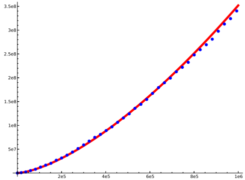
Hypothesis T

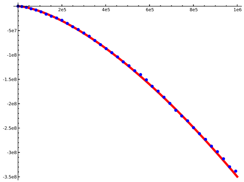
- Need 2-dimensional analogue of Hypothesis S for extending support in 2-level density (just a zero statistic).
- Expect sum to have a **main term** (not just being bounded) to give agreement with random matrix theory

Hypothesis T

- Need 2-dimensional analogue of Hypothesis S for extending support in 2-level density (just a zero statistic).
- Expect sum to have a **main term** (not just being bounded) to give agreement with random matrix theory
- Difficult to even find conjecture about what to prove; working on this

More Pretty Pictures

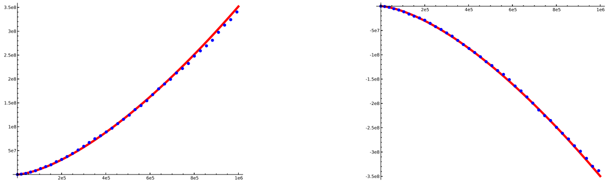




- Real and imaginary parts of

$$\sum_{p_1, p_2 \leq X} e^{2\pi i \cdot 2\sqrt{p_1 p_2}} \qquad \text{(here, } X_1 = X_2 = X\text{)}$$

More Pretty Pictures



- Real and imaginary parts of

$$\sum_{p_1,p_2 \le X} e^{2\pi i \cdot 2\sqrt{p_1 p_2}} \qquad \text{(here, } X_1 = X_2 = X\text{)}$$

- Not as random as the one-dimensional sum!

- Thank You!
- Next up, we have . . .

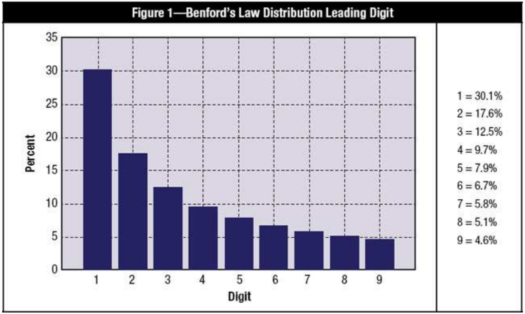
Benford Behavior of Dependent Random Variables

Taylor Corcoran - University of Arizona
Jaclyn Porfilio - Williams College
Jirapat Samranvedhya - Williams College

Benford's Law

Definition

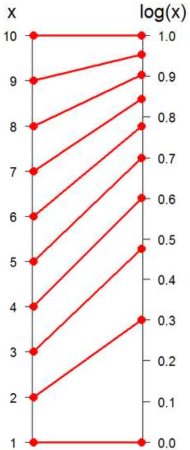
A dataset is said to follow **Benford's Law** (base b) if the probability of observing a first digit of d is $\log_b \frac{1+d}{d}$.



Logarithms and Benford's Law

$\mathbb{P}(\text{leading digit } d) = \log(d+1) - \log(d)$

Benford's law $\leftrightarrow$ mantissa of logarithms of data are uniformly distributed



Stick Decomposition

Fixed Proportion Decomposition Process

Decomposition Process

1. Consider a stick of length $\mathcal{L}$.

Fixed Proportion Decomposition Process

Decomposition Process

1. Consider a stick of length $\mathcal{L}$.
2. Uniformly choose a proportion $p \in (0, 1)$.

Fixed Proportion Decomposition Process

Decomposition Process

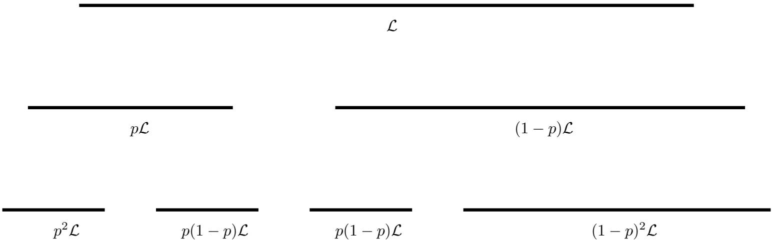
1. Consider a stick of length $\mathcal{L}$.
2. Uniformly choose a proportion $p \in (0, 1)$.
3. Break the stick into two pieces: lengths $p\mathcal{L}$ and $(1 - p)\mathcal{L}$.

Fixed Proportion Decomposition Process

Decomposition Process

1. Consider a stick of length $\mathcal{L}$.
2. Uniformly choose a proportion $p \in (0, 1)$.
3. Break the stick into two pieces: lengths $p\mathcal{L}$ and $(1 - p)\mathcal{L}$.
4. Repeat N times (using the same proportion).

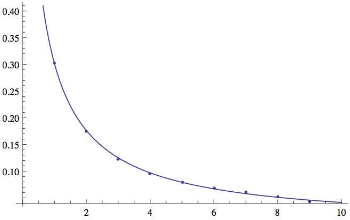
Fixed Proportion Decomposition Process



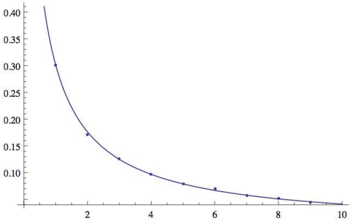
Fixed Proportion Conjecture

Joy Jing's Conjecture

The above decomposition process results in stick lengths that obey Benford's Law as $N \to \infty$ for any $p \in (0,1)$, $p \neq \frac{1}{2}$.



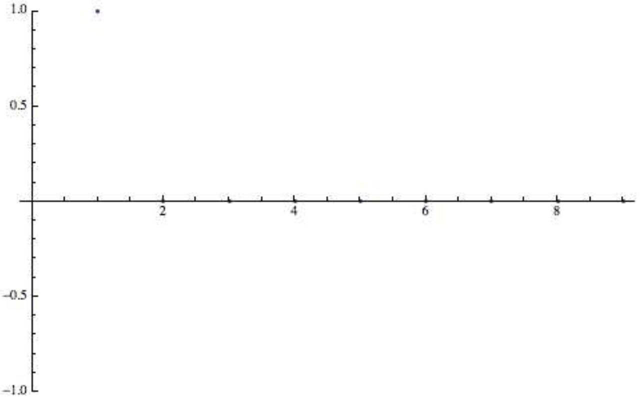
(B) $p = 0.51$ and $N = 10000$.



(B) $p = 0.99$ and $N = 50000$. Benford distribution overlaid.

Counterexample: $p = \frac{1}{11},\ 1 - p = \frac{10}{11}$.

```
BenfordFixedCut2[1 / 11, Floor[SetAccuracy[Random[], 50] * 10 ^ 4]]
```



Benford Analysis

After Nth interation,

- 2^N sticks
- $N + 1$ distinct lengths.

Distinct lengths are given by

$$x_{j+1} = \left(\frac{1-p}{p}\right) x_j, \;\; x_0 = p^N.$$

Benford Analysis

After Nth interation,

- 2^N sticks
- $N+1$ distinct lengths.

Distinct lengths are given by

$$x_{j+1} \; = \; \left(\frac{1-p}{p}\right) x_j, \;\; x_0 = p^N.$$

Let $\frac{1-p}{p} = 10^y$.

$\frac{1-p}{p} = 10^y$, y $\in \mathbb{Q}$

Theorem

Let $\frac{1-p}{p} = 10^y$. If $y \in \mathbb{Q}$, the described decomposition process results in stick lengths that do not obey Benford's Law.

Let $y = \frac{r}{q}$.

Leading digit of x_j repeats every q indices. Thus,

$$\sum_k P(x_{j+kq}) = \sum_k \binom{N}{j+kq}.$$

Series Multisection

Multisection Formula

If $f = \sum_{n=-\infty}^{\infty} a_n x^n$,

$$\sum_{k=-\infty}^{\infty} a_{kq+j} x^{kq+j} = \frac{1}{q} \sum_{p=0}^{q-1} \omega^{-jp} f(\omega^p x)$$

where ω is the primitive qth root of unity $e^{2\pi i/q}$.

$\frac{1-p}{p} = 10^y,\ y \in \mathbb{Q}$

$$\sum_k P(x_{j+kq}) = \frac{1}{q}\left(1 + \mathscr{E}\left[(q-1)\left(\cos\frac{\pi}{q}\right)^N\right]\right)$$

$\frac{1-p}{p} = 10^y$, $y \in \mathbb{Q}$

$$\sum_k P(x_{j+kq}) = \frac{1}{q}\left(1 + \mathscr{E}\left[(q-1)\left(\cos\frac{\pi}{q}\right)^N\right]\right)$$

Digit frequencies are multiples of $\frac{1}{q}$.

Benford frequencies are irrational, so *not* perfect Benford.

$\frac{1-p}{p} = 10^y$, y $\notin \mathbb{Q}$: Outline

Theorem

Let $\frac{1-p}{p} = 10^y$. If $y \notin \mathbb{Q}$, the described decomposition process results in stick lengths that obey Benford's Law.

$\frac{1-p}{p} = 10^y$, y $\notin \mathbb{Q}$: Outline

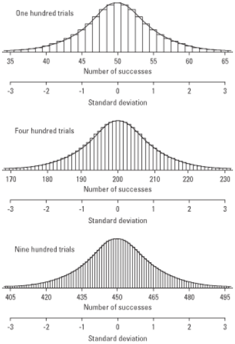
Theorem

Let $\frac{1-p}{p} = 10^y$. If $y \notin \mathbb{Q}$, the described decomposition process results in stick lengths that obey Benford's Law.

$\{x_j\} \sim Bin(N, \frac{1}{2})$
mean: $\frac{N}{2}$
standard deviation: $\frac{\sqrt{N}}{2}$

Outline of proof strategy:

1. Truncation
2. Break into intervals
   - Roughly equal probability
   - Equidistribution



$\frac{1-p}{p} = 10^y$, $y \notin \mathbb{Q}$: Truncation

For $\epsilon > 0$, Chebyshev's Inequality gives

$$\begin{aligned} P\left(\left|x - \frac{N}{2}\right| \geq N^{\frac{1}{2}+\epsilon}\right) &= P\left(\left|x - \frac{N}{2}\right| \geq N^{\epsilon} N^{\frac{1}{2}}\right) \\ &\leq \frac{1}{N^{2\epsilon}}. \end{aligned}$$

$\frac{1-p}{p} = 10^y$, $y \notin \mathbb{Q}$: Truncation

For $\epsilon > 0$, Chebyshev's Inequality gives

$$\begin{aligned} P\left(\left|x - \frac{N}{2}\right| \geq N^{\frac{1}{2}+\epsilon}\right) &= P\left(\left|x - \frac{N}{2}\right| \geq N^{\epsilon} N^{\frac{1}{2}}\right) \\ &\leq \frac{1}{N^{2\epsilon}}. \end{aligned}$$

So we can limit our analysis to

- One standard deviation
- Right half of binomial

$\frac{1-p}{p} = 10^y$, y $\notin \mathbb{Q}$: Intervals and Roughly Equal Probability

$\mathcal{I}_\ell = \{x_\ell, x_\ell + 1, \dots, x_\ell + N^\delta - 1\}$.

Let $x_0 = N/2$. It follows that $x_\ell = N/2 + \ell N^\delta$.

$$\left| \binom{N}{x_\ell} - \binom{N}{x_{\ell+1}} \right| \leq \binom{N}{x_\ell} N^{-\frac{1}{2}+\delta+\epsilon},$$

when $\delta < 1/2 - \epsilon$ and $\ell \leq N^{1/2-\delta+\epsilon}$.

$\frac{1-p}{p} = 10^y$, y $\notin \mathbb{Q}$: Equidistribution

Definition

$\{x_n\}_{n=1}^{\infty}$ is equidistributed modulo 1 if for any $[a,b] \subset [0,1]$, $P\left(x_n \bmod 1 \in [a,b]\right) \to b-a$:

$$\lim_{N\to\infty} \frac{\#\{n \le N : x_n \bmod 1 \in [a,b]\}}{N} = b-a.$$

Recall: Leading digits of stick lengths are Benford if their logarithms are equidistributed modulo 1.

$\frac{1-p}{p} = 10^y$, y $\notin \mathbb{Q}$: Equidistribution

Consider an interval I_ℓ where

$$I_\ell = \{x_\ell + i : i \in \{0, 1, \dots, N^\delta - 1\}\}$$

$$J_\ell \subset \{0, 1, \dots, N^\delta - 1\} = \{i : \log(x_\ell + i) \bmod 1 \in [a, b]\}.$$

$\frac{1-p}{p} = 10^y$, y $\notin \mathbb{Q}$: Equidistribution

Consider an interval I_ℓ where

$I_\ell = \{x_\ell + i : i \in \{0, 1, \ldots, N^\delta - 1\}\}$

$J_\ell \subset \{0, 1, \ldots, N^\delta - 1\} = \{i : \log(x_\ell + i) \bmod 1 \in [a, b]\}$.

If the irrationality exponent κ of y is finite,

$$|J_\ell| \quad = \quad (b-a)N^\delta + O(N^{\delta(1-\frac{1}{\kappa}+\epsilon')})$$

$\frac{1-p}{p} = 10^y$, $y \notin \mathbb{Q}$: Equidistribution

Using

- equidistribution within intervals
- roughly equal probability

we have

$$\sum_{\ell} \sum_{i \in J_\ell} f(x_\ell + i) \quad = \quad (b-a) + O(N^{\delta(-\frac{1}{\kappa}+\epsilon')} + N^{-\frac{1}{2}+\delta+\epsilon}).$$

where κ is the irrationality exponent of y.

$\frac{1-p}{p} = 10^y$, y $\notin \mathbb{Q}$: Equidistribution

Remark

- Rate of convergence depends on $\kappa < \infty$
- Still Benford for $\kappa = \infty$, but no quantified rate.

Additive Stick Decomposition Processes: Conjectures

Benford

- Stop at evens (proved)
- Stop at primes
- Cutting into *m* pieces

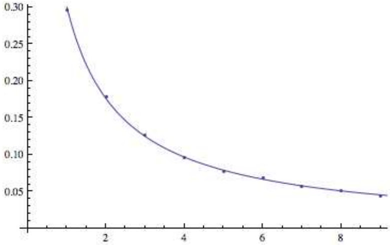
Non-Benford

- Stop at squares
- Stop at powers of two
- Stop at powers of three
- Stop at Fibonnaci numbers

Any Questions?

Additive Decomposition Process (Evens Model)

```
Digit Count After 23127 iterations: {6858, 4137, 2941, 2241, 1796, 1587, 1331, 1205, 1032}
Benford Prob for d = 1 is 0.30103 and observe 0.296524.
Benford Prob for d = 2 is 0.176091 and observe 0.178874.
Benford Prob for d = 3 is 0.124939 and observe 0.127162.
Benford Prob for d = 4 is 0.09691 and observe 0.0968955.
Benford Prob for d = 5 is 0.0791812 and observe 0.0776548.
Benford Prob for d = 6 is 0.0669468 and observe 0.0686181.
Benford Prob for d = 7 is 0.0579919 and observe 0.0575493.
Benford Prob for d = 8 is 0.0511525 and observe 0.0521013.
Benford Prob for d = 9 is 0.0457575 and observe 0.0446212.
There were 1 pieces of length one.
There were 23128 fragmented pieces.
The value of Chi Squared (goodness of fit for Benford) is 6.27561.
```



Irrationality Exponent

Let $x \in \mathbb{R}$. Denote by $\mathcal{A}$ the set of positive numbers n for which

$$0 \leq \left| x - \frac{p}{q} \right| \leq \frac{1}{q^n}$$

has at most finitely many solutions for $p, q \in \mathbb{Z}$.
The irrationality measure of x, denoted $n(x)$, is $\inf_{n \in \mathcal{A}} n$.
If $\mathcal{A}$ is empty, $n(x) = \infty$
For nonempty $\mathcal{A}$,

$$n(x) = \begin{cases} 1 \text{ if } x \text{ is rational} \\ 2 \text{ if } x \text{ is algebraic of degree } > 1 \\ \geq 2 \text{ if } x \text{ is transcendental} \end{cases}$$

Acknowledgements

We would like to thank our advisor, Steven J. Miller, our co-author Joseph R. Iafrate, as well as Frederick W. Strauch and Joy Jing.

We would also like to thank the Williams College SMALL summer research program.

This research was funded by NSF grant DMS0850577.

References

C Hongwei Chen (2010). *On the Summation of First Subseries in Closed Form.*International Journal of Mathematical Education in Science and Technology 41:4, 538-547.

JKKKM D. Jang, J. U. Kang, A. Kruckman, J. Kudo and S. J. Miller, *Chains of distributions, hierarchical Bayesian models and Benford's Law*, to appear in the Journal of Algebra, Number Theory: Advances and Applications.

KM Alex V. Kontorovich and Steven J. Miller. *Benford's Law, Values of L-functions and the* $3x + 1$ *Problem.* http://arxiv.org/pdf/math/0412003v2.pdf

MT Steven J. Miller and Ramin Takloo-Bighash. “Needed Gaussian Integral." *An Invitation to Modern Number Theory.* Princeton: Princeton UP, 2006. 222. Print.

BGMRS T. Becker, A. Greaves-Tunnell, S. Miller, R. Ronan, and F. Strauch, *Benford's Law and Continuous Dependent Random Variables*. http://arxiv.org/pdf/1111.0568v1.pdf

UNIF *Distribution of First Digit.* Digital image. Benford's Law. DataGenetics, n.d. Web. 15 July 2013. http://www.datagenetics.com/blog/march52012/.

BLD *Benford's Law Distribution Leading Digit.* Digital image. ISACA. N.p., n.d. Web. 15 June 2013. http://www.isaca.org/Journal/Past-Issues/2011/Volume-3/PublishingImages/11v3-understanding-and.jpg.

BLA Singleton, Tommie W., Ph.D. *Benford's Law Test/Comparison.* Digital image. ISACA. N.p., 2011. Web. 16 July 2013. http://www.isaca.org/Journal/Past-Issues/2011/Volume-3/Pages/Understanding-and-Applying-Benfords-Law.aspx.

JJ Jing, Joy. *Benford's Law and Stick Decomposition.* Undergraduate Mathematics Thesis, Williams College. Advisor: Steven Miller.

W Weisstein, Eric W. "Irrationality Measure." From MathWorld–A Wolfram Web Resource. http://mathworld.wolfram.com/IrrationalityMeasure.html