

Number Theory and Probability SMALL '13 Projects

Steven J. Miller (sjm1@williams.edu)

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

Williams College, June 19, 2013

Zeckendorf

Previous Results

Fibonacci Numbers: $F_{n+1} = F_n + F_{n-1}$;
 $F_1 = 1, F_2 = 2, F_3 = 3, F_4 = 5, \dots$

Previous Results

Fibonacci Numbers: $F_{n+1} = F_n + F_{n-1}$;
 $F_1 = 1, F_2 = 2, F_3 = 3, F_4 = 5, \dots$

Zeckendorf's Theorem

Every positive integer can be written uniquely as a sum of non-consecutive Fibonacci numbers.

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

Previous Results

Fibonacci Numbers: $F_{n+1} = F_n + F_{n-1}$;
 $F_1 = 1, F_2 = 2, F_3 = 3, F_4 = 5, \dots$

Zeckendorf's Theorem

Every positive integer can be written uniquely as a sum of non-consecutive Fibonacci numbers.

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

Lekkerkerker's Theorem (1952)

The average number of summands in the Zeckendorf decomposition for integers in $[F_n, F_{n+1})$ tends to $\frac{n}{\varphi^2 + 1} \approx .276n$, where $\varphi = \frac{1+\sqrt{5}}{2}$ is the golden mean.

Old Results

Central Limit Type Theorem

As $n \rightarrow \infty$, the distribution of the number of summands in the Zeckendorf decomposition for integers in $[F_n, F_{n+1})$ is Gaussian.

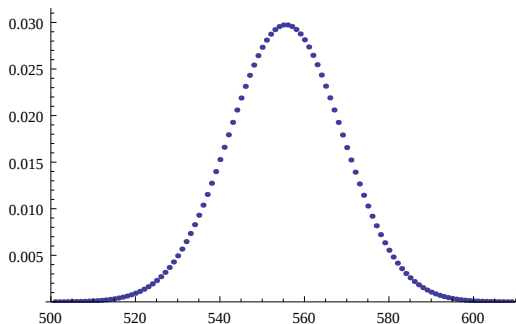


Figure: Number of summands in $[F_{2010}, F_{2011})$; $F_{2010} \approx 10^{420}$.

New Results: Bulk Gaps: $m \in [F_n, F_{n+1})$ and $\phi = \frac{1+\sqrt{5}}{2}$

$$m = \sum_{j=1}^{k(m)=n} F_{i_j}, \quad \nu_{m;n}(x) = \frac{1}{k(m)-1} \sum_{j=2}^{k(m)} \delta(x - (i_j - i_{j-1})).$$

Theorem (Zeckendorf Gap Distribution)

Gap measures $\nu_{m;n}$ converge to average gap measure where $P(k) = 1/\phi^k$ for $k \geq 2$.

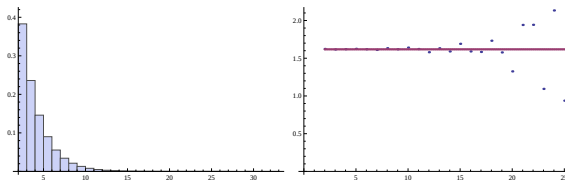


Figure: Distribution of gaps in $[F_{1000}, F_{1001})$; $F_{2010} \approx 10^{208}$.

New Results: Longest Gap

Fair coin: largest gap tightly concentrated around $\log n / \log 2$.

Theorem (Longest Gap)

As $n \rightarrow \infty$, the probability that $m \in [F_n, F_{n+1})$ has longest gap less than or equal to $f(n)$ converges to

$$\text{Prob}(L_n(m) \leq f(n)) \approx e^{-e^{\log n - f(n) \cdot \log \phi}}$$

- $\mu_n = \frac{\log\left(\frac{\phi^2}{\phi^2+1}n\right)}{\log \phi} + \frac{\gamma}{\log \phi} - \frac{1}{2} + \text{Small Error}.$
- If $f(n)$ grows **slower** (resp. **faster**) than $\log n / \log \phi$, then $\text{Prob}(L_n(m) \leq f(n))$ goes to **0** (resp. **1**).

The Cookie Problem: Reinterpretation

Reinterpreting the Cookie Problem

The number of solutions to $x_1 + \cdots + x_P = C$ with $x_i \geq 0$ is $\binom{C+P-1}{P-1}$.

The Cookie Problem: Reinterpretation

Reinterpreting the Cookie Problem

The number of solutions to $x_1 + \cdots + x_p = C$ with $x_i \geq 0$ is $\binom{C+p-1}{p-1}$.

Let $p_{n,k} = \# \{N \in [F_n, F_{n+1}) : \text{the Zeckendorf decomposition of } N \text{ has exactly } k \text{ summands}\}$.

For $N \in [F_n, F_{n+1})$, the **largest summand is F_n** .

$$N = F_{i_1} + F_{i_2} + \cdots + F_{i_{k-1}} + F_n,$$

The Cookie Problem: Reinterpretation

Reinterpreting the Cookie Problem

The number of solutions to $x_1 + \cdots + x_P = C$ with $x_i \geq 0$ is $\binom{C+P-1}{P-1}$.

Let $p_{n,k} = \# \{N \in [F_n, F_{n+1}): \text{the Zeckendorf decomposition of } N \text{ has exactly } k \text{ summands}\}$.

For $N \in [F_n, F_{n+1})$, the **largest summand is F_n** .

$$N = F_{i_1} + F_{i_2} + \cdots + F_{i_{k-1}} + F_n,$$

$$1 \leq i_1 < i_2 < \cdots < i_{k-1} < i_k = n, i_j - i_{j-1} \geq 2.$$

$$d_1 := i_1 - 1, d_j := i_j - i_{j-1} - 2 \ (j > 1).$$

$$d_1 + d_2 + \cdots + d_k = n - 2k + 1, d_j \geq 0.$$

The Cookie Problem: Reinterpretation

Reinterpreting the Cookie Problem

The number of solutions to $x_1 + \cdots + x_P = C$ with $x_i \geq 0$ is $\binom{C+P-1}{P-1}$.

Let $p_{n,k} = \# \{N \in [F_n, F_{n+1}) : \text{the Zeckendorf decomposition of } N \text{ has exactly } k \text{ summands}\}$.

For $N \in [F_n, F_{n+1})$, the **largest summand is F_n** .

$$N = F_{i_1} + F_{i_2} + \cdots + F_{i_{k-1}} + F_n,$$

$$1 \leq i_1 < i_2 < \cdots < i_{k-1} < i_k = n, i_j - i_{j-1} \geq 2.$$

$$d_1 := i_1 - 1, d_j := i_j - i_{j-1} - 2 \ (j > 1).$$

$$d_1 + d_2 + \cdots + d_k = n - 2k + 1, d_j \geq 0.$$

$$\text{Cookie counting} \Rightarrow p_{n,k} = \binom{n-2k+1+k-1}{k-1} = \binom{n-k}{k-1}.$$

(Sketch of the) Proof of Gaussianity

The probability density for the number of Fibonacci numbers that add up to an integer in $[F_n, F_{n+1})$ is $f_n(k) = \binom{n-1-k}{k} / F_{n-1}$. Consider the density for the $n+1$ case. Then we have, by Stirling

$$\begin{aligned} f_{n+1}(k) &= \binom{n-k}{k} \frac{1}{F_n} \\ &= \frac{(n-k)!}{(n-2k)!k!} \frac{1}{F_n} = \frac{1}{\sqrt{2\pi}} \frac{(n-k)^{n-k+\frac{1}{2}}}{k^{(k+\frac{1}{2})(n-2k+\frac{1}{2})}} \frac{1}{F_n} \end{aligned}$$

plus a lower order correction term.

Also we can write $F_n = \frac{1}{\sqrt{5}} \phi^{n+1} = \frac{\phi}{\sqrt{5}} \phi^n$ for large n , where ϕ is the golden ratio (we are using relabeled Fibonacci numbers where $1 = F_1$ occurs once to help dealing with uniqueness and $F_2 = 2$). We can now split the terms that exponentially depend on n .

$$f_{n+1}(k) = \left(\frac{1}{\sqrt{2\pi}} \sqrt{\frac{(n-k)}{k(n-2k)}} \frac{\sqrt{5}}{\phi} \right) \left(\phi^{-n} \frac{(n-k)^{n-k}}{k^k (n-2k)^{n-2k}} \right).$$

Define

$$N_n = \frac{1}{\sqrt{2\pi}} \sqrt{\frac{(n-k)}{k(n-2k)}} \frac{\sqrt{5}}{\phi}, \quad S_n = \phi^{-n} \frac{(n-k)^{n-k}}{k^k (n-2k)^{n-2k}}.$$

Thus, write the density function as

$$f_{n+1}(k) = N_n S_n$$

where N_n is the first term that is of order $n^{-1/2}$ and S_n is the second term with exponential dependence on n .

(Sketch of the) Proof of Gaussianity (cont)

Model the distribution as centered around the mean by the change of variable $k = \mu + x\sigma$ where μ and σ are the mean and the standard deviation, and depend on n . The discrete weights of $f_n(k)$ will become continuous. This requires us to use the change of variable formula to compensate for the change of scales:

$$f_n(k)dk = f_n(\mu + \sigma x)\sigma dx.$$

Using the change of variable, we can write N_n as

$$\begin{aligned} N_n &= \frac{1}{\sqrt{2\pi}} \sqrt{\frac{n-k}{k(n-2k)}} \frac{\phi}{\sqrt{5}} \\ &= \frac{1}{\sqrt{2\pi n}} \sqrt{\frac{1-k/n}{(k/n)(1-2k/n)}} \frac{\sqrt{5}}{\phi} \\ &= \frac{1}{\sqrt{2\pi n}} \sqrt{\frac{1-(\mu+\sigma x)/n}{((\mu+\sigma x)/n)(1-2(\mu+\sigma x)/n)}} \frac{\sqrt{5}}{\phi} \\ &= \frac{1}{\sqrt{2\pi n}} \sqrt{\frac{1-C-y}{(C+y)(1-2C-2y)}} \frac{\sqrt{5}}{\phi} \end{aligned}$$

where $C = \mu/n \approx 1/(\phi+2)$ (note that $\phi^2 = \phi+1$) and $y = \sigma x/n$. But for large n , the y term vanishes since $\sigma \sim \sqrt{n}$ and thus $y \sim n^{-1/2}$. Thus

$$N_n \approx \frac{1}{\sqrt{2\pi n}} \sqrt{\frac{1-C}{C(1-2C)}} \frac{\sqrt{5}}{\phi} = \frac{1}{\sqrt{2\pi n}} \sqrt{\frac{(\phi+1)(\phi+2)}{\phi}} \frac{\sqrt{5}}{\phi} = \frac{1}{\sqrt{2\pi n}} \sqrt{\frac{5(\phi+2)}{\phi}} = \frac{1}{\sqrt{2\pi\sigma^2}}$$

since $\sigma^2 = n \frac{\phi}{5(\phi+2)}$.

(Sketch of the) Proof of Gaussianity (cont)

For the second term S_n , take the logarithm and once again change variables by $k = \mu + x\sigma$,

$$\begin{aligned}
 \log(S_n) &= \log \left(\phi^{-n} \frac{(n-k)^{(n-k)}}{k^k (n-2k)^{(n-2k)}} \right) \\
 &= -n \log(\phi) + (n-k) \log(n-k) - (k) \log(k) \\
 &\quad - (n-2k) \log(n-2k) \\
 &= -n \log(\phi) + (n - (\mu + x\sigma)) \log(n - (\mu + x\sigma)) \\
 &\quad - (\mu + x\sigma) \log(\mu + x\sigma) \\
 &\quad - (n - 2(\mu + x\sigma)) \log(n - 2(\mu + x\sigma)) \\
 &= -n \log(\phi) \\
 &\quad + (n - (\mu + x\sigma)) \left(\log(n - \mu) + \log \left(1 - \frac{x\sigma}{n - \mu} \right) \right) \\
 &\quad - (\mu + x\sigma) \left(\log(\mu) + \log \left(1 + \frac{x\sigma}{\mu} \right) \right) \\
 &\quad - (n - 2(\mu + x\sigma)) \left(\log(n - 2\mu) + \log \left(1 - \frac{x\sigma}{n - 2\mu} \right) \right) \\
 &= -n \log(\phi) \\
 &\quad + (n - (\mu + x\sigma)) \left(\log \left(\frac{n}{\mu} - 1 \right) + \log \left(1 - \frac{x\sigma}{n - \mu} \right) \right) \\
 &\quad - (\mu + x\sigma) \log \left(1 + \frac{x\sigma}{\mu} \right) \\
 &\quad - (n - 2(\mu + x\sigma)) \left(\log \left(\frac{n}{\mu} - 2 \right) + \log \left(1 - \frac{x\sigma}{n - 2\mu} \right) \right).
 \end{aligned}$$

(Sketch of the) Proof of Gaussianity (cont)

Note that, since $n/\mu = \phi + 2$ for large n , the constant terms vanish. We have $\log(S_n)$

$$\begin{aligned}
 &= -n \log(\phi) + (n-k) \log\left(\frac{n}{\mu} - 1\right) - (n-2k) \log\left(\frac{n}{\mu} - 2\right) + (n - (\mu + x\sigma)) \log\left(1 - \frac{x\sigma}{n - \mu}\right) \\
 &\quad - (\mu + x\sigma) \log\left(1 + \frac{x\sigma}{\mu}\right) - (n - 2(\mu + x\sigma)) \log\left(1 - \frac{x\sigma}{n - 2\mu}\right) \\
 &= -n \log(\phi) + (n-k) \log(\phi + 1) - (n-2k) \log(\phi) + (n - (\mu + x\sigma)) \log\left(1 - \frac{x\sigma}{n - \mu}\right) \\
 &\quad - (\mu + x\sigma) \log\left(1 + \frac{x\sigma}{\mu}\right) - (n - 2(\mu + x\sigma)) \log\left(1 - \frac{x\sigma}{n - 2\mu}\right) \\
 &= n(-\log(\phi) + \log(\phi^2) - \log(\phi)) + k(\log(\phi^2) + 2\log(\phi)) + (n - (\mu + x\sigma)) \log\left(1 - \frac{x\sigma}{n - \mu}\right) \\
 &\quad - (\mu + x\sigma) \log\left(1 + \frac{x\sigma}{\mu}\right) - (n - 2(\mu + x\sigma)) \log\left(1 - \frac{x\sigma}{n - 2\mu}\right) \\
 &= (n - (\mu + x\sigma)) \log\left(1 - \frac{x\sigma}{n - \mu}\right) - (\mu + x\sigma) \log\left(1 + \frac{x\sigma}{\mu}\right) \\
 &\quad - (n - 2(\mu + x\sigma)) \log\left(1 - \frac{x\sigma}{n - 2\mu}\right).
 \end{aligned}$$

(Sketch of the) Proof of Gaussianity (cont)

Finally, we expand the logarithms and collect powers of $x\sigma/n$.

$$\begin{aligned}
 \log(S_n) &= (n - (\mu + x\sigma)) \left(-\frac{x\sigma}{n - \mu} - \frac{1}{2} \left(\frac{x\sigma}{n - \mu} \right)^2 + \dots \right) \\
 &\quad - (\mu + x\sigma) \left(\frac{x\sigma}{\mu} - \frac{1}{2} \left(\frac{x\sigma}{\mu} \right)^2 + \dots \right) \\
 &\quad - (n - 2(\mu + x\sigma)) \left(-2\frac{x\sigma}{n - 2\mu} - \frac{1}{2} \left(2\frac{x\sigma}{n - 2\mu} \right)^2 + \dots \right) \\
 &= (n - (\mu + x\sigma)) \left(-\frac{x\sigma}{n \frac{(\phi+1)}{(\phi+2)}} - \frac{1}{2} \left(\frac{x\sigma}{n \frac{(\phi+1)}{(\phi+2)}} \right)^2 + \dots \right) \\
 &\quad - (\mu + x\sigma) \left(\frac{x\sigma}{\frac{n}{\phi+2}} - \frac{1}{2} \left(\frac{x\sigma}{\frac{n}{\phi+2}} \right)^2 + \dots \right) \\
 &\quad - (n - 2(\mu + x\sigma)) \left(-\frac{2x\sigma}{n \frac{\phi}{\phi+2}} - \frac{1}{2} \left(\frac{2x\sigma}{n \frac{\phi}{\phi+2}} \right)^2 + \dots \right) \\
 &= \frac{x\sigma}{n} n \left(-\left(1 - \frac{1}{\phi+2} \right) \frac{(\phi+2)}{(\phi+1)} - 1 + 2 \left(1 - \frac{2}{\phi+2} \right) \frac{\phi+2}{\phi} \right) \\
 &\quad - \frac{1}{2} \left(\frac{x\sigma}{n} \right)^2 n \left(-2\frac{\phi+2}{\phi+1} + \frac{\phi+2}{\phi+1} + 2(\phi+2) - (\phi+2) + 4\frac{\phi+2}{\phi} \right) \\
 &\quad + O\left(n(x\sigma/n)^3\right)
 \end{aligned}$$

(Sketch of the) Proof of Gaussianity (cont)

$$\begin{aligned}
 \log(S_n) &= \frac{x\sigma}{n} n \left(-\frac{\phi+1}{\phi+2} \frac{\phi+2}{\phi+1} - 1 + 2 \frac{\phi}{\phi+2} \frac{\phi+2}{\phi} \right) \\
 &\quad - \frac{1}{2} \left(\frac{x\sigma}{n} \right)^2 n(\phi+2) \left(-\frac{1}{\phi+1} + 1 + \frac{4}{\phi} \right) \\
 &\quad + O \left(n \left(\frac{x\sigma}{n} \right)^3 \right) \\
 &= -\frac{1}{2} \frac{(x\sigma)^2}{n} (\phi+2) \left(\frac{3\phi+4}{\phi(\phi+1)} + 1 \right) + O \left(n \left(\frac{x\sigma}{n} \right)^3 \right) \\
 &= -\frac{1}{2} \frac{(x\sigma)^2}{n} (\phi+2) \left(\frac{3\phi+4+2\phi+1}{\phi(\phi+1)} \right) + O \left(n \left(\frac{x\sigma}{n} \right)^3 \right) \\
 &= -\frac{1}{2} x^2 \sigma^2 \left(\frac{5(\phi+2)}{\phi n} \right) + O \left(n (x\sigma/n)^3 \right).
 \end{aligned}$$

(Sketch of the) Proof of Gaussianity (cont)

But recall that

$$\sigma^2 = \frac{\phi n}{5(\phi + 2)}.$$

Also, since $\sigma \sim n^{-1/2}$, $n \left(\frac{x\sigma}{n} \right)^3 \sim n^{-1/2}$. So for large n , the $O \left(n \left(\frac{x\sigma}{n} \right)^3 \right)$ term vanishes. Thus we are left with

$$\begin{aligned} \log S_n &= -\frac{1}{2}x^2 \\ S_n &= e^{-\frac{1}{2}x^2}. \end{aligned}$$

Hence, as n gets large, the density converges to the normal distribution:

$$\begin{aligned} f_n(k)dk &= N_n S_n dk \\ &= \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}x^2} \sigma dx \\ &= \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} dx. \end{aligned}$$

□

Current Questions

- Far Difference Representations.
- Recurrences with Negative Coefficients.
- Two-Dimensional Sequences.

MSTD

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: $A - A = \{a_i - a_j : a_i, a_j \in A\}$.

Arise in Goldbach's Problem, Twin Primes, Fermat's Last Theorem,

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: $A - A = \{a_i - a_j : a_i, a_j \in A\}$.

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.

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?

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.

Representative Result: SMALL '11

Question: Can we find A with $|kA + kA| > |kA - kA|$?

Representative Result: SMALL '11

Question: Can we find A with $|kA + kA| > |kA - kA|$?

- One set gives infinitely many (generalize Miller-Orosz-Scheinerman), more work get positive percentage.

Representative Result: SMALL '11

Question: Can we find A with $|kA + kA| > |kA - kA|$?

- One set gives infinitely many (generalize Miller-Orosz-Scheinerman), more work get positive percentage.
- How do we find one set?

Representative Result: SMALL '11

Question: Can we find A with $|kA + kA| > |kA - kA|$?

- One set gives infinitely many (generalize Miller-Orosz-Scheinerman), more work get positive percentage.
- How do we find one set?
- Computer simulations? We couldn't find a set for $k = 2$; the probability of finding some of these sets is less than 10^{-16} .

Representative Result: SMALL '11

Question: Can we find A with $|kA + kA| > |kA - kA|$?

- One set gives infinitely many (generalize Miller-Orosz-Scheinerman), more work get positive percentage.
- How do we find one set?
- Computer simulations? We couldn't find a set for $k = 2$; the probability of finding some of these sets is less than 10^{-16} .

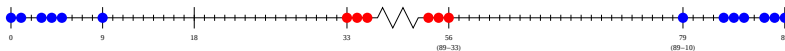
If A is symmetric ($A = c - A$ for some c) then

$$|A + A| = |A + (c - A)| = |A - A|.$$

Example: $|2A + 2A| > |2A - 2A|$

$$A = \{0, 1, 3, 4, 5, 9\} \cup [33, 56] \cup \{79, 83, 84, 85, 87, 88, 89\}$$

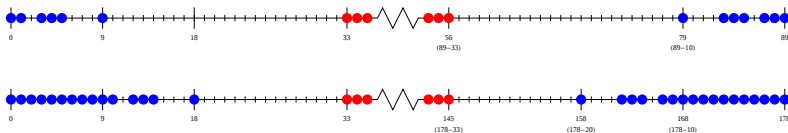
A



Example: $|2A + 2A| > |2A - 2A|$

$$A = \{0, 1, 3, 4, 5, 9\} \cup [33, 56] \cup \{79, 83, 84, 85, 87, 88, 89\}$$

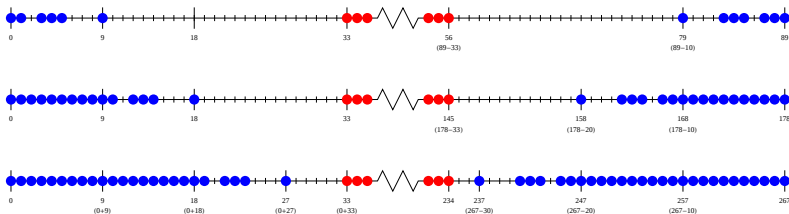
$A + A$



Example: $|2A + 2A| > |2A - 2A|$

$$A = \{0, 1, 3, 4, 5, 9\} \cup [33, 56] \cup \{79, 83, 84, 85, 87, 88, 89\}$$

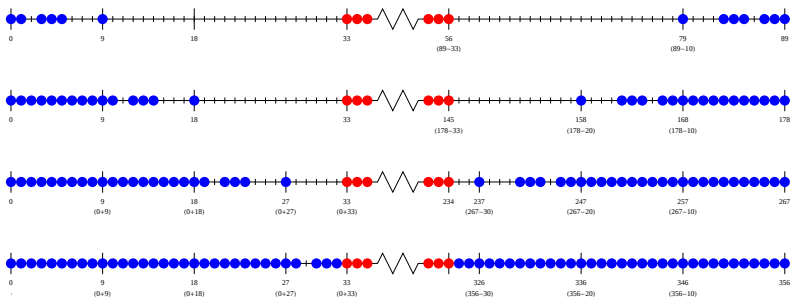
$A + A + A$



Example: $|2A + 2A| > |2A - 2A|$

$$A = \{0, 1, 3, 4, 5, 9\} \cup [33, 56] \cup \{79, 83, 84, 85, 87, 88, 89\}$$

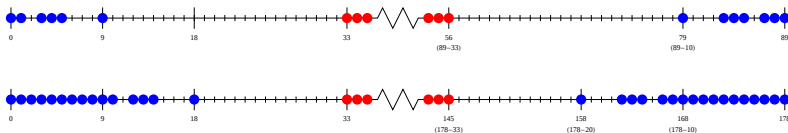
$$A + A + A + A$$



Example: $|2A + 2A| > |2A - 2A|$

$$A = \{0, 1, 3, 4, 5, 9\} \cup [33, 56] \cup \{79, 83, 84, 85, 87, 88, 89\}$$

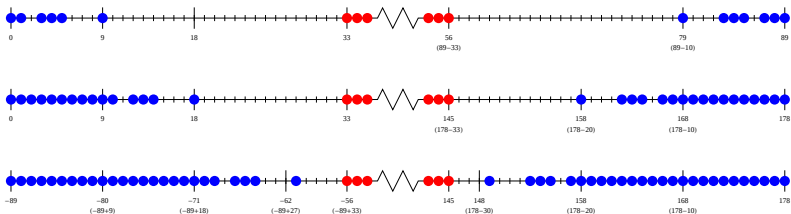
$A + A$



Example: $|2A + 2A| > |2A - 2A|$

$$A = \{0, 1, 3, 4, 5, 9\} \cup [33, 56] \cup \{79, 83, 84, 85, 87, 88, 89\}$$

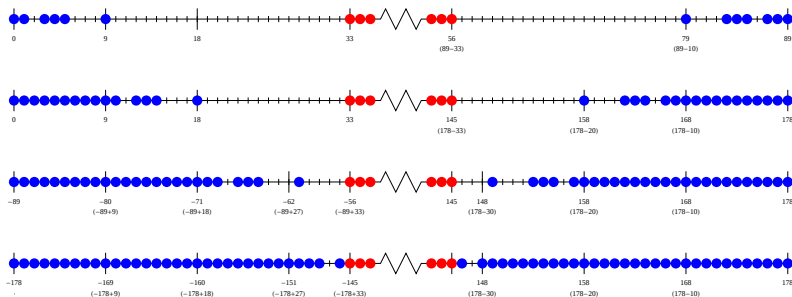
$$A + A - A$$



Example: $|2A + 2A| > |2A - 2A|$

$$A = \{0, 1, 3, 4, 5, 9\} \cup [33, 56] \cup \{79, 83, 84, 85, 87, 88, 89\}$$

$$A + A - A - A$$



Current Questions

- Study pairs A, B and allow dependence ($B = A^c$).
- Two-Dimensional Analogues (how does geometry matter).
- Phase Transitions.

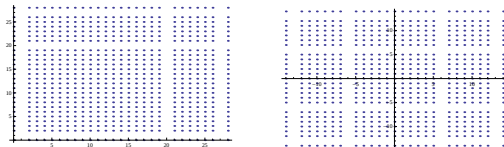


Figure: Can you find the $A \subset \mathbb{Z}^2$? $|A + A| = 676$, $|A - A| = 625$.

$$A = \{2, 4, 5\} \text{ and } A^c = \{0, 1, 3, 6, 7\}.$$

Benford's Law

Introduction

Interesting Question

For a nice data set, such as the Fibonacci numbers, stock prices, street addresses of audience, ..., what percent of the leading digits are 1?

Plausible answers:

Introduction

Interesting Question

For a nice data set, such as the Fibonacci numbers, stock prices, street addresses of audience, ..., what percent of the leading digits are 1?

Plausible answers: 10%

Introduction

Interesting Question

For a nice data set, such as the Fibonacci numbers, stock prices, street addresses of audience, ..., what percent of the leading digits are 1?

Plausible answers: 10%, 11%

Introduction

Interesting Question

For a nice data set, such as the Fibonacci numbers, stock prices, street addresses of audience, ..., what percent of the leading digits are 1?

Plausible answers: 10%, 11%, about 30%.

Benford's Law: Newcomb (1881), Benford (1938)

Statement

For many data sets, probability of observing a first digit of d base B is $\log_B \left(\frac{d+1}{d} \right)$; base 10 about 30% are 1s.

Benford's Law: Newcomb (1881), Benford (1938)

Statement

For many data sets, probability of observing a first digit of d base B is $\log_B \left(\frac{d+1}{d} \right)$; base 10 about 30% are 1s.

- Not all data sets satisfy Benford's Law.

Benford's Law: Newcomb (1881), Benford (1938)

Statement

For many data sets, probability of observing a first digit of d base B is $\log_B \left(\frac{d+1}{d} \right)$; base 10 about 30% are 1s.

- Not all data sets satisfy Benford's Law.
 - ◇ Long street $[1, L]$: $L = 199$ versus $L = 999$.

Benford's Law: Newcomb (1881), Benford (1938)

Statement

For many data sets, probability of observing a first digit of d base B is $\log_B \left(\frac{d+1}{d} \right)$; base 10 about 30% are 1s.

- Not all data sets satisfy Benford's Law.
 - ◇ Long street $[1, L]$: $L = 199$ versus $L = 999$.
 - ◇ Oscillates between $1/9$ and $5/9$ with first digit 1.

Benford's Law: Newcomb (1881), Benford (1938)

Statement

For many data sets, probability of observing a first digit of d base B is $\log_B \left(\frac{d+1}{d} \right)$; base 10 about 30% are 1s.

- Not all data sets satisfy Benford's Law.
 - ◇ Long street $[1, L]$: $L = 199$ versus $L = 999$.
 - ◇ Oscillates between $1/9$ and $5/9$ with first digit 1.
 - ◇ Many streets of different sizes: close to Benford.

Examples

- recurrence relations
- special functions (such as $n!$)
- iterates of power, exponential, rational maps
- products of random variables
- L -functions, characteristic polynomials
- iterates of the $3x + 1$ map
- differences of order statistics
- hydrology and financial data
- many hierarchical Bayesian models

Applications

- analyzing round-off errors
- determining the optimal way to store numbers
- detecting tax and image fraud, and data integrity

Significands

Significand: $x = S_{10}(x) \cdot 10^k$, k integer.

$S_{10}(x) = S_{10}(\tilde{x})$ if and only if x and $\tilde{x}$ have the same leading digits.

Key observation: $\log_{10}(x) = \log_{10}(\tilde{x}) \bmod 1$ if and only if x and $\tilde{x}$ have the same leading digits.

Significands

Significand: $x = S_{10}(x) \cdot 10^k$, k integer.

$S_{10}(x) = S_{10}(\tilde{x})$ if and only if x and $\tilde{x}$ have the same leading digits.

Key observation: $\log_{10}(x) = \log_{10}(\tilde{x}) \bmod 1$ if and only if x and $\tilde{x}$ have the same leading digits.

- ◇ Thus often study $y = \log_{10} x$.
- ◇ Benefit: $e^{2\pi i(y \bmod 1)} = e^{2\pi i y}$.

Equidistribution and Benford's Law

Equidistribution

$\{y_n\}_{n=1}^{\infty}$ is equidistributed modulo 1 if probability $y_n \bmod 1 \in [a, b]$ tends to $b - a$:

$$\frac{\#\{n \leq N : y_n \bmod 1 \in [a, b]\}}{N} \rightarrow b - a.$$

Equidistribution and Benford's Law

Equidistribution

$\{y_n\}_{n=1}^{\infty}$ is equidistributed modulo 1 if probability $y_n \bmod 1 \in [a, b]$ tends to $b - a$:

$$\frac{\#\{n \leq N : y_n \bmod 1 \in [a, b]\}}{N} \rightarrow b - a.$$

- Thm: $\beta \notin \mathbb{Q}$, $n\beta$ is equidistributed mod 1.

Equidistribution and Benford's Law

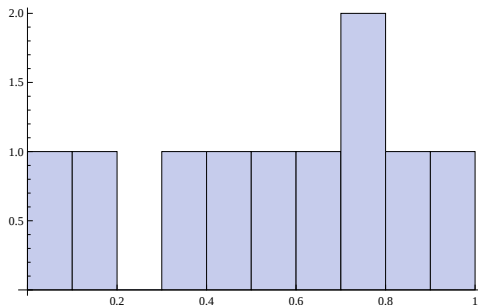
Equidistribution

$\{y_n\}_{n=1}^{\infty}$ is equidistributed modulo 1 if probability $y_n \bmod 1 \in [a, b]$ tends to $b - a$:

$$\frac{\#\{n \leq N : y_n \bmod 1 \in [a, b]\}}{N} \rightarrow b - a.$$

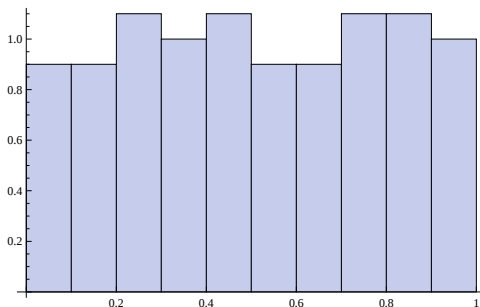
- Thm: $\beta \notin \mathbb{Q}$, $n\beta$ is equidistributed mod 1.
- Examples: $\log_{10} 2, \log_{10} \left(\frac{1+\sqrt{5}}{2} \right) \notin \mathbb{Q}$.

Example of Equidistribution: $n\sqrt{\pi} \bmod 1$



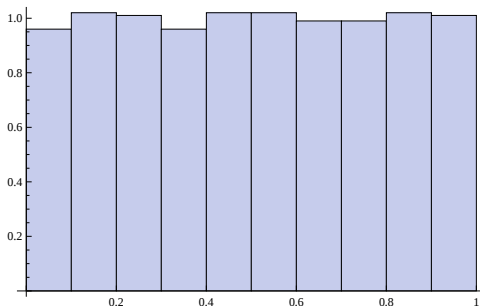
$n\sqrt{\pi} \bmod 1$ for $n \leq 10$

Example of Equidistribution: $n\sqrt{\pi} \bmod 1$



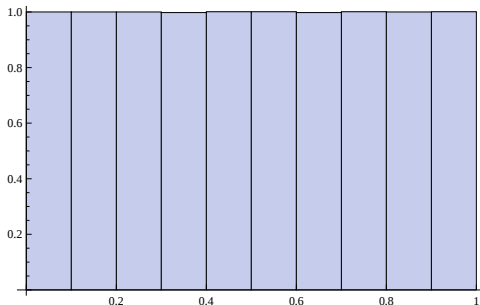
$n\sqrt{\pi} \bmod 1$ for $n \leq 100$

Example of Equidistribution: $n\sqrt{\pi} \bmod 1$



$n\sqrt{\pi} \bmod 1$ for $n \leq 1000$

Example of Equidistribution: $n\sqrt{\pi} \bmod 1$



$n\sqrt{\pi} \bmod 1$ for $n \leq 10,000$

Logarithms and Benford's Law

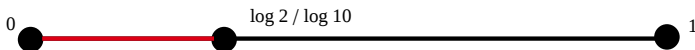
Fundamental Equivalence

Data set $\{x_i\}$ is Benford base B if $\{y_i\}$ is equidistributed mod 1, where $y_i = \log_B x_i$.

Logarithms and Benford's Law

Fundamental Equivalence

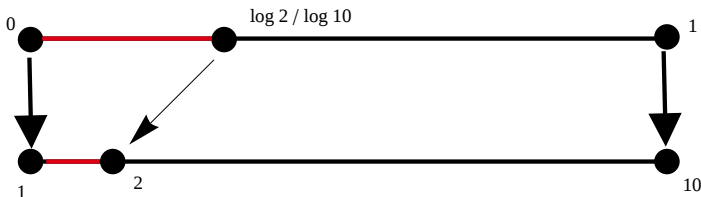
Data set $\{x_i\}$ is Benford base B if $\{y_i\}$ is equidistributed mod 1, where $y_i = \log_B x_i$.



Logarithms and Benford's Law

Fundamental Equivalence

Data set $\{x_i\}$ is Benford base B if $\{y_i\}$ is equidistributed mod 1, where $y_i = \log_B x_i$.



Examples

- 2^n is Benford base 10 as $\log_{10} 2 \notin \mathbb{Q}$.

Examples

- 2^n is Benford base 10 as $\log_{10} 2 \notin \mathbb{Q}$.
- Fibonacci numbers are Benford base 10.

$$a_{n+1} = a_n + a_{n-1}.$$

$$\text{Binet: } a_n = \frac{1}{\sqrt{5}} \left(\frac{1+\sqrt{5}}{2} \right)^n - \frac{1}{\sqrt{5}} \left(\frac{1-\sqrt{5}}{2} \right)^n.$$

Examples

- 2^n is Benford base 10 as $\log_{10} 2 \notin \mathbb{Q}$.
- Fibonacci numbers are Benford base 10.

$$a_{n+1} = a_n + a_{n-1}.$$

$$\text{Binet: } a_n = \frac{1}{\sqrt{5}} \left(\frac{1+\sqrt{5}}{2} \right)^n - \frac{1}{\sqrt{5}} \left(\frac{1-\sqrt{5}}{2} \right)^n.$$

- Most linear recurrence relations Benford.

$$a_{n+1} = 2a_n$$

$$a_{n+1} = 2a_n - a_{n-1}$$

Take $a_0 = a_1 = 1$ or $a_0 = 0, a_1 = 1$.

Digits of 2^n

First 60 values of 2^n (only displaying 30)

			digit	#	Obs Prob	Benf Prob
1	1024	1048576				
2	2048	2097152	1	18	.300	.301
4	4096	4194304	2	12	.200	.176
8	8192	8388608	3	6	.100	.125
16	16384	16777216	4	6	.100	.097
32	32768	33554432	5	6	.100	.079
64	65536	67108864	6	4	.067	.067
128	131072	134217728	7	2	.033	.058
256	262144	268435456	8	5	.083	.051
512	524288	536870912	9	1	.017	.046

Digits of 2^n

First 60 values of 2^n (only displaying 30)

			digit	#	Obs Prob	Benf Prob
1	1024	1048576				
2	2048	2097152	1	18	.300	.301
4	4096	4194304	2	12	.200	.176
8	8192	8388608	3	6	.100	.125
16	16384	16777216	4	6	.100	.097
32	32768	33554432	5	6	.100	.079
64	65536	67108864	6	4	.067	.067
128	131072	134217728	7	2	.033	.058
256	262144	268435456	8	5	.083	.051
512	524288	536870912	9	1	.017	.046

Digits of 2^n

First 60 values of 2^n (only displaying 30): $2^{10} = 1024 \approx 10^3$.

			digit	#	Obs Prob	Benf Prob
1	1024	1048576				
2	2048	2097152	1	18	.300	.301
4	4096	4194304	2	12	.200	.176
8	8192	8388608	3	6	.100	.125
16	16384	16777216	4	6	.100	.097
32	32768	33554432	5	6	.100	.079
64	65536	67108864	6	4	.067	.067
128	131072	134217728	7	2	.033	.058
256	262144	268435456	8	5	.083	.051
512	524288	536870912	9	1	.017	.046

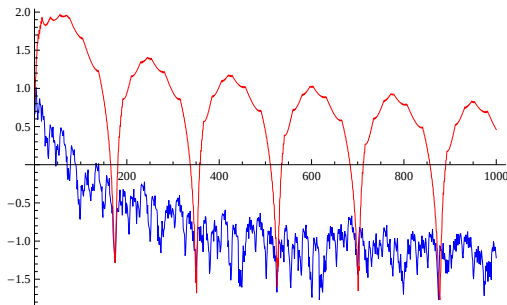
Logarithms and Benford's Law

χ^2 values for α^n , $1 \leq n \leq N$ (5% 15.5).

N	$\chi^2(\gamma)$	$\chi^2(e)$	$\chi^2(\pi)$
100	0.72	0.30	46.65
200	0.24	0.30	8.58
400	0.14	0.10	10.55
500	0.08	0.07	2.69
700	0.19	0.04	0.05
800	0.04	0.03	6.19
900	0.09	0.09	1.71
1000	0.02	0.06	2.90

Logarithms and Benford's Law: Base 10

$\log_{10}(\chi^2)$ vs N for π^n (red) and e^n (blue),
 $n \in \{1, \dots, N\}$. Note $\pi^{175} \approx 1.0028 \cdot 10^{87}$, (5%
and 8 d.f., $\log_{10}(\chi^2) \approx .44$).



[illegible]

Detecting Fraud

Bank Fraud

- Audit of a bank revealed huge spike of numbers

Detecting Fraud

Bank Fraud

- Audit of a bank revealed huge spike of numbers starting with 4

Detecting Fraud

Bank Fraud

- Audit of a bank revealed huge spike of numbers starting with 48 and 49

Detecting Fraud

Bank Fraud

- Audit of a bank revealed huge spike of numbers starting with 48 and 49, most due to one person.

Detecting Fraud

Bank Fraud

- Audit of a bank revealed huge spike of numbers starting with 48 and 49, most due to one person.
- Write-off limit of \$5,000. Officer had friends applying for credit cards, ran up balances just under \$5,000 then he would write the debts off.

$3x + 1$ Problem

- Kakutani (conspiracy), Erdős (not ready).
- x odd, $T(x) = \frac{3x+1}{2^k}$, $2^k || 3x + 1$.

$3x + 1$ Problem

- Kakutani (conspiracy), Erdős (not ready).
- x odd, $T(x) = \frac{3x+1}{2^k}$, $2^k \parallel 3x + 1$.
- Conjecture: for some $n = n(x)$, $T^n(x) = 1$.

3x + 1 Problem

- Kakutani (conspiracy), Erdős (not ready).
- x odd, $T(x) = \frac{3x+1}{2^k}$, $2^k || 3x + 1$.
- Conjecture: for some $n = n(x)$, $T^n(x) = 1$.
- 7

$3x + 1$ Problem

- Kakutani (conspiracy), Erdős (not ready).
- x odd, $T(x) = \frac{3x+1}{2^k}$, $2^k \parallel 3x + 1$.
- Conjecture: for some $n = n(x)$, $T^n(x) = 1$.
- $7 \rightarrow_1 11$

$3x + 1$ Problem

- Kakutani (conspiracy), Erdős (not ready).
- x odd, $T(x) = \frac{3x+1}{2^k}$, $2^k \parallel 3x + 1$.
- Conjecture: for some $n = n(x)$, $T^n(x) = 1$.
- $7 \rightarrow_1 11 \rightarrow_1 17$

$3x + 1$ Problem

- Kakutani (conspiracy), Erdős (not ready).
- x odd, $T(x) = \frac{3x+1}{2^k}$, $2^k \parallel 3x + 1$.
- Conjecture: for some $n = n(x)$, $T^n(x) = 1$.
- $7 \rightarrow_1 11 \rightarrow_1 17 \rightarrow_2 13$

$3x + 1$ Problem

- Kakutani (conspiracy), Erdős (not ready).
- x odd, $T(x) = \frac{3x+1}{2^k}$, $2^k \parallel 3x + 1$.
- Conjecture: for some $n = n(x)$, $T^n(x) = 1$.
- $7 \rightarrow_1 11 \rightarrow_1 17 \rightarrow_2 13 \rightarrow_3 5$

$3x + 1$ Problem

- Kakutani (conspiracy), Erdős (not ready).
- x odd, $T(x) = \frac{3x+1}{2^k}$, $2^k \parallel 3x + 1$.
- Conjecture: for some $n = n(x)$, $T^n(x) = 1$.
- $7 \rightarrow_1 11 \rightarrow_1 17 \rightarrow_2 13 \rightarrow_3 5 \rightarrow_4 1$

$3x + 1$ Problem

- Kakutani (conspiracy), Erdős (not ready).
- x odd, $T(x) = \frac{3x+1}{2^k}$, $2^k \parallel 3x + 1$.
- Conjecture: for some $n = n(x)$, $T^n(x) = 1$.
- $7 \rightarrow_1 11 \rightarrow_1 17 \rightarrow_2 13 \rightarrow_3 5 \rightarrow_4 1 \rightarrow_2 1$,

$3x + 1$ Problem

- Kakutani (conspiracy), Erdős (not ready).
- x odd, $T(x) = \frac{3x+1}{2^k}$, $2^k \parallel 3x + 1$.
- Conjecture: for some $n = n(x)$, $T^n(x) = 1$.
- $7 \rightarrow_1 11 \rightarrow_1 17 \rightarrow_2 13 \rightarrow_3 5 \rightarrow_4 1 \rightarrow_2 1$,
2-path $(1, 1)$, 5-path $(1, 1, 2, 3, 4)$.
 m -path: $(k_1, \dots, k_m)$.

Structure Theorem: Sinai, Kontorovich-Sinai

$$\mathbb{P}(A) = \lim_{N \rightarrow \infty} \frac{\#\{n \leq N : n \equiv 1, 5 \pmod{6}, n \in A\}}{\#\{n \leq N : n \equiv 1, 5 \pmod{6}\}}.$$

Structure Theorem: Sinai, Kontorovich-Sinai

$$\mathbb{P}(A) = \lim_{N \rightarrow \infty} \frac{\#\{n \leq N : n \equiv 1, 5 \pmod{6}, n \in A\}}{\#\{n \leq N : n \equiv 1, 5 \pmod{6}\}}.$$

$(k_1, \dots, k_m)$: two full arithm progressions:
 $6 \cdot 2^{k_1 + \dots + k_m} p + q.$

Structure Theorem: Sinai, Kontorovich-Sinai

$$\mathbb{P}(A) = \lim_{N \rightarrow \infty} \frac{\#\{n \leq N : n \equiv 1, 5 \pmod{6}, n \in A\}}{\#\{n \leq N : n \equiv 1, 5 \pmod{6}\}}.$$

$(k_1, \dots, k_m)$: two full arithm progressions:

$$6 \cdot 2^{k_1 + \dots + k_m} p + q.$$

Theorem (Sinai, Kontorovich-Sinai)

k_i -values are i.i.d.r.v. (geometric, $1/2$):

$$\mathbb{P} \left(\frac{\log_2 \left[\frac{x_m}{\left(\frac{3}{4}\right)^m x_0} \right]}{\sqrt{2m}} \leq a \right) = \mathbb{P} \left(\frac{S_m - 2m}{\sqrt{2m}} \leq a \right)$$

Structure Theorem: Sinai, Kontorovich-Sinai

$$\mathbb{P}(A) = \lim_{N \rightarrow \infty} \frac{\#\{n \leq N : n \equiv 1, 5 \pmod{6}, n \in A\}}{\#\{n \leq N : n \equiv 1, 5 \pmod{6}\}}.$$

$(k_1, \dots, k_m)$: two full arithm progressions:

$$6 \cdot 2^{k_1 + \dots + k_m} p + q.$$

Theorem (Sinai, Kontorovich-Sinai)

k_i -values are i.i.d.r.v. (geometric, $1/2$):

$$\mathbb{P} \left(\frac{\log_2 \left[\frac{x_m}{\left(\frac{3}{4}\right)^m x_0} \right]}{(\log_2 B) \sqrt{2m}} \leq a \right) = \mathbb{P} \left(\frac{S_m - 2m}{(\log_2 B) \sqrt{2m}} \leq a \right)$$

Structure Theorem: Sinai, Kontorovich-Sinai

$$\mathbb{P}(A) = \lim_{N \rightarrow \infty} \frac{\#\{n \leq N : n \equiv 1, 5 \pmod{6}, n \in A\}}{\#\{n \leq N : n \equiv 1, 5 \pmod{6}\}}.$$

$(k_1, \dots, k_m)$: two full arithm progressions:

$$6 \cdot 2^{k_1 + \dots + k_m} p + q.$$

Theorem (Sinai, Kontorovich-Sinai)

k_i -values are i.i.d.r.v. (geometric, $1/2$):

$$\mathbb{P} \left(\frac{\log_B \left[\frac{x_m}{\left(\frac{3}{4}\right)^m x_0} \right]}{\sqrt{2m}} \leq a \right) = \mathbb{P} \left(\frac{\frac{(S_m - 2m)}{\log_2 B}}{\sqrt{2m}} \leq a \right)$$

$3x + 1$ and Benford

Theorem (Kontorovich and M–, 2005)

As $m \rightarrow \infty$, $x_m / (3/4)^m x_0$ is Benford.

Theorem (Lagarias-Soundararajan 2006)

$X \geq 2^N$, for all but at most $c(B)N^{-1/36}X$ initial seeds the distribution of the first N iterates of the $3x + 1$ map are within $2N^{-1/36}$ of the Benford probabilities.

Prereq: Irrationality Type

Irrationality type

α has irrationality type κ if κ is the supremum of all γ with

$$\lim_{q \rightarrow \infty} q^{\gamma+1} \min_p \left| \alpha - \frac{p}{q} \right| = 0.$$

- Algebraic irrationals: type 1 (Roth's Thm).
- Theory of Linear Forms: $\log_B 2$ of finite type.

Prereq: Linear Forms

Theorem (Baker)

$\alpha_1, \dots, \alpha_n$ algebraic numbers height $A_j \geq 4$,
 $\beta_1, \dots, \beta_n \in \mathbb{Q}$ with height at most $B \geq 4$,

$$\Lambda = \beta_1 \log \alpha_1 + \dots + \beta_n \log \alpha_n.$$

If $\Lambda \neq 0$ then $|\Lambda| > B^{-C \Omega \log \Omega'}$, with
 $d = [\mathbb{Q}(\alpha_i, \beta_j) : \mathbb{Q}]$, $C = (16nd)^{200n}$,
 $\Omega = \prod_j \log A_j$, $\Omega' = \Omega / \log A_n$.

Gives $\log_{10} 2$ of finite type, with $\kappa < 1.2 \cdot 10^{602}$:

$$|\log_{10} 2 - p/q| = |q \log 2 - p \log 10| / q \log 10.$$

Prereq: Quantified Equidistribution

Theorem (Erdős-Turan)

$$D_N = \frac{\sup_{[a,b]} |N(b-a) - \#\{n \leq N : x_n \in [a,b]\}|}{N}$$

There is a C such that for all m :

$$D_N \leq C \cdot \left(\frac{1}{m} + \sum_{h=1}^m \frac{1}{h} \left| \frac{1}{N} \sum_{n=1}^N e^{2\pi i h x_n} \right| \right)$$

Prereq: Ideas for the proof of Erdős-Turan

Consider special case $x_n = n\alpha$, $\alpha \notin \mathbb{Q}$.

- Exponential sum $\leq \frac{1}{|\sin(\pi h\alpha)|} \leq \frac{1}{2||h\alpha||}$.
- Must control $\sum_{h=1}^m \frac{1}{h||h\alpha||}$, see irrationality type enter.
- type κ , $\sum_{h=1}^m \frac{1}{h||h\alpha||} = O(m^{\kappa-1+\epsilon})$, take $m = \lfloor N^{1/\kappa} \rfloor$.

Sketch of the proof: Notation

- $[a, b]$ is an arbitrary sub-interval of $[0, 1]$.
- $c \in (0, \frac{1}{2})$ and set $M = m^c$.
- $I_\ell := \{\ell M, \ell M + 1, \dots, (\ell + 1)M - 1\}$.
- $C := \log_B 2$ an irrational number of type κ .
- $\eta(x)$ the density of the standard normal.
- S_m sum of m independent $\text{Geom}(1/2)$ r.v.

Sketch of the proof

Step 1: Central Limit Theorem:

CLT: for any $k \in \mathbb{Z}$ have

$$\begin{aligned}\text{Prob}(C \cdot \bar{S}_m = C \cdot k) &= \text{Prob}\left(\frac{\bar{S}_m}{\sqrt{m}} = \frac{k}{\sqrt{m}}\right) \\ &= \frac{1}{\sqrt{m}} \eta\left(\frac{k}{\sqrt{m}}\right) + o\left(\frac{1}{\sqrt{m}}\right)\end{aligned}$$

Write $o\left(\frac{1}{\sqrt{m}}\right)$ as $O\left(\frac{1}{\sqrt{m}g(m)}\right)$ for some monotone increasing $g(m)$ which tends to infinity.

Sketch of the proof

Step 2: Variation in I_ℓ :

Let $k_1, k_2 \in I_\ell$. Then

$$\left| \frac{1}{\sqrt{m}} \eta \left(\frac{k_1}{\sqrt{m}} \right) - \frac{1}{\sqrt{m}} \eta \left(\frac{k_2}{\sqrt{m}} \right) \right| \\ \leq \frac{1}{\sqrt{m}} e^{-\ell^2 M^2 / 2m} \cdot \left(1 - \exp \left(-\frac{2\ell M^2 + M^2}{2m} \right) \right).$$

Means that for the ℓ we must study, there is negligible variation in the Gaussian for $k \in I_\ell$.

Sketch of the proof

Step 3: Poisson Summation:

By Poisson Summation:

$$\frac{1}{\sigma} \sum_{n=-\infty}^{\infty} e^{-n^2 \pi / \sigma^2} = \sum_{n=-\infty}^{\infty} e^{-n^2 \pi \sigma^2}, \quad \sigma > 0.$$

Take $\sigma^2 = \frac{2m}{\pi M^2}$, and use this to calculate the main term.

Sketch of the proof

Step 4: Quantified Equidistribution:

For any $\epsilon > 0$, letting $\delta = 1 + \epsilon - \frac{1}{\kappa} < 1$ we have

$$\#\{k \in I_\ell : kC \bmod 1 \in [a, b]\} = M(b-a) + O(M^\delta).$$

The quantification of the equidistribution of $kC \bmod 1$ is the key ingredient in proving Benford behavior base B . Follows from the Erdős-Turan Theorem.

Sketch of the proof

Step 5: Combining Pieces:

Must show as $m \rightarrow \infty$, for any $[a, b] \subset [0, 1]$,

$$P_m(a, b) := \text{Prob}(C\bar{S}_m \bmod 1 \in [a, b]) \longrightarrow b-a.$$

We have

$$\begin{aligned} P_m(a, b) &= \sum_{|\ell| \leq \frac{\sqrt{mh(m)}}{M}} \text{Prob}(\bar{S}_m = k \in I_\ell : kC \bmod 1 \in [a, b]) \\ &\quad + \sum_{|\ell| > \frac{\sqrt{mh(m)}}{M}} \text{Prob}(\bar{S}_m = k \in I_\ell : kC \bmod 1 \in [a, b]) \\ &= \sum_{|\ell| \leq \frac{\sqrt{mh(m)}}{M}} \text{Prob}(\bar{S}_m = k \in I_\ell : kC \bmod 1 \in [a, b]) + o(1). \end{aligned}$$

$3x + 1$ Data: random 10,000 digit number, $2^k \parallel 3x + 1$

80,514 iterations ($(4/3)^n = a_0$ predicts 80,319);
 $\chi^2 = 13.5$ (5% 15.5).

Digit	Number	Observed	Benford
1	24251	0.301	0.301
2	14156	0.176	0.176
3	10227	0.127	0.125
4	7931	0.099	0.097
5	6359	0.079	0.079
6	5372	0.067	0.067
7	4476	0.056	0.058
8	4092	0.051	0.051
9	3650	0.045	0.046

$3x + 1$ Data: random 10,000 digit number, $2|3x + 1$

241,344 iterations, $\chi^2 = 11.4$ (5% 15.5).

Digit	Number	Observed	Benford
1	72924	0.302	0.301
2	42357	0.176	0.176
3	30201	0.125	0.125
4	23507	0.097	0.097
5	18928	0.078	0.079
6	16296	0.068	0.067
7	13702	0.057	0.058
8	12356	0.051	0.051
9	11073	0.046	0.046

$5x + 1$ Data: random 10,000 digit number, $2^k \parallel 5x + 1$

27,004 iterations, $\chi^2 = 1.8$ (5% 15.5).

Digit	Number	Observed	Benford
1	8154	0.302	0.301
2	4770	0.177	0.176
3	3405	0.126	0.125
4	2634	0.098	0.097
5	2105	0.078	0.079
6	1787	0.066	0.067
7	1568	0.058	0.058
8	1357	0.050	0.051
9	1224	0.045	0.046

$5x + 1$ Data: random 10,000 digit number, $2|5x + 1$

241,344 iterations, $\chi^2 = 3 \cdot 10^{-4}$ (5% 15.5).

Digit	Number	Observed	Benford
1	72652	0.301	0.301
2	42499	0.176	0.176
3	30153	0.125	0.125
4	23388	0.097	0.097
5	19110	0.079	0.079
6	16159	0.067	0.067
7	13995	0.058	0.058
8	12345	0.051	0.051
9	11043	0.046	0.046

Current Questions

- Dependent Processes: Interval Decomposition.