Why the IRS cares about the Riemann Zeta Function and Number Theory
(and why you should too!)

Steven J. Miller
sjml@williams.edu,
Steven.Miller.MC.96@aya.yale.edu

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

Carnegie Mellon University, March 25, 2/12/15
Winona State University, March 30, 2/12/15
Interesting Question

Motivating Question: For a nice data set, such as the Fibonacci numbers, stock prices, street addresses of college employees and students, ..., what percent of the leading digits are 1?
Interesting Question

Motivating Question: For a nice data set, such as the Fibonacci numbers, stock prices, street addresses of college employees and students, ..., what percent of the leading digits are 1?

Natural guess: 10% (but immediately correct to 11%!).
Motivating Question: For a nice data set, such as the Fibonacci numbers, stock prices, street addresses of college employees and students, ..., what percent of the leading digits are 1?

Answer: Benford’s law!
Examples with First Digit Bias

Fibonacci numbers
Examples with First Digit Bias

Fibonacci numbers

Most common iPhone passcodes
Examples with First Digit Bias

- **Fibonacci numbers**
  
  ![Fibonacci Numbers Graph]

- **Twitter users by # followers**
  
  ![Twitter Followers Graph]

- **Most common iPhone passcodes**
  
  ![iPhone Passcodes Graph]
Examples with First Digit Bias

Fibonacci numbers

Twitter users by # followers

Most common iPhone passcodes

Distance of stars from Earth
Summary

- Explain Benford’s Law.
- Discuss examples and applications.
- Sketch proofs.
- Describe open problems.
Caveats!

- A math test indicating fraud is not proof of fraud: unlikely events, alternate reasons.
Caveats!

- A math test indicating fraud is *not* proof of fraud: unlikely events, alternate reasons.
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.
General Theory
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 (probabilities)
Background Material

- Modulo: $a = b \mod c$ if $a - b$ is an integer times $c$; thus $17 = 5 \mod 12$, and $4.5 = .5 \mod 1$. 
Background Material

- Modulo: \( a = b \mod c \) if \( a - b \) is an integer times \( c \); thus \( 17 = 5 \mod 12 \), and \( 4.5 = .5 \mod 1 \).

- Significand: \( x = S_{10}(x) \cdot 10^k \), \( k \) integer, \( 1 \leq S_{10}(x) < 10 \).
Background Material

- Modulo: $a = b \mod c$ if $a - b$ is an integer times $c$; thus $17 = 5 \mod 12$, and $4.5 = .5 \mod 1$.

- Significand: $x = S_{10}(x) \cdot 10^k$, $k$ integer, $1 \leq S_{10}(x) < 10$.

- $S_{10}(x) = S_{10}(\tilde{x})$ if and only if $x$ and $\tilde{x}$ have the same leading digits. Note $\log_{10} x = \log_{10} S_{10}(x) + k$. 
Background Material

- Modulo: \( a = b \mod c \) if \( a - b \) is an integer times \( c \); thus \( 17 = 5 \mod 12 \), and \( 4.5 = .5 \mod 1 \).

- Significand: \( x = S_{10}(x) \cdot 10^k \), \( k \) integer, \( 1 \leq S_{10}(x) < 10 \).

- \( S_{10}(x) = S_{10}(\tilde{x}) \) if and only if \( x \) and \( \tilde{x} \) have the same leading digits. Note \( \log_{10} x = \log_{10} S_{10}(x) + k \).

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

Thus often study \( y = \log_{10} x \mod 1 \).

**Advanced:** \( e^{2\pi i u} = e^{2\pi i (u \mod 1)} \).
Equidistribution and Benford’s Law

Equidistribution

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

\[
\frac{\# \{ n \leq N : y_n \mod 1 \in [a, b] \}}{N} \to b - a.
\]
Equidistribution and Benford’s Law

Equidistribution

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

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

Thm: \( \beta \notin \mathbb{Q} \), \( n\beta \) is equidistributed mod 1.
Equidistribution and Benford’s Law

**Equidistribution**

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

\[
\frac{\#\{n \leq N : y_n \mod 1 \in [a, b]\}}{N} \to b - a.
\]

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

- Examples: \( \log_{10} 2, \log_{10} \left(\frac{1+\sqrt{5}}{2}\right) \notin \mathbb{Q}. \)
Equidistribution and Benford's Law

**Equidistribution**

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

\[
\frac{\#\{n \leq N : y_n \mod 1 \in [a, b]\}}{N} \to b - a.
\]

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

- **Examples:** \( \log_{10} 2 \), \( \log_{10} \left( \frac{1+\sqrt{5}}{2} \right) \notin \mathbb{Q} \).

**Proof:** if rational: \( 2 = 10^{p/q} \).

\[
\text{Proof: if rational: } 2 = 10^{p/q}.
\]
Equidistribution and Benford’s Law

**Equidistribution**

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

\[
\frac{\#\{n \leq N : y_n \text{ mod 1 } \in [a, b]\}}{N} \to b - a.
\]

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

- **Examples:** $\log_{10} 2$, $\log_{10} \left(\frac{1+\sqrt{5}}{2}\right) \notin \mathbb{Q}$.

  **Proof:** if rational: $2 = 10^{p/q}$.

  Thus $2^q = 10^p$ or $2^{q-p} = 5^p$, impossible.
Example of Equidistribution: $n\sqrt{\pi} \mod 1$

$n\sqrt{\pi} \mod 1$ for $n \leq 10$
Example of Equidistribution: $n \sqrt{\pi} \mod 1$

$n \sqrt{\pi} \mod 1$ for $n \leq 100$
Example of Equidistribution: $n\sqrt{\pi} \mod 1$

$n\sqrt{\pi} \mod 1$ for $n \leq 1000$
Example of Equidistribution: $n\sqrt{\pi} \mod 1$

$n\sqrt{\pi} \mod 1$ 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 \( y_i = \log_B x_i \).
**Fundamental Equivalence**

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


\[
x = S_{10}(x) \cdot 10^k \text{ then } \log_{10} x = \log_{10} S_{10}(x) + k = \log_{10} S_{10}x \mod 1.
\]
Logarithms and Benford’s Law

Fundamental Equivalence

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

\[
x = S_{10}(x) \cdot 10^k \text{ then } \log_{10} x = \log_{10} S_{10}(x) + k = \log_{10} S_{10}x \mod 1.
\]
Logarithms and Benford’s Law

\[ \text{Prob(leading digit } d) = \log_{10}(d+1) - \log_{10}(d) = \log_{10} \left( \frac{d+1}{d} \right) = \log_{10} \left( 1 + \frac{1}{d} \right). \]

Have Benford’s law \(\leftrightarrow\) mantissa of logarithms of data are uniformly distributed
Examples

- $2^n$ is Benford base 10 as $\log_{10} 2 \not\in \mathbb{Q}$. 
Examples

- Fibonacci numbers are Benford base 10.
Examples

Fibonacci numbers are Benford base 10.

\[ a_{n+1} = a_n + a_{n-1}. \]
• Fibonacci numbers are Benford base 10.

\[ a_{n+1} = a_n + a_{n-1}. \]

Guess \( a_n = r^n: \)
\[ r^{n+1} = r^n + r^{n-1} \text{ or } r^2 = r + 1. \]
Fibonacci numbers are Benford base 10.

\[ a_{n+1} = a_n + a_{n-1}. \]

Guess \( a_n = r^n: \ r^{n+1} = r^n + r^{n-1} \) or \( r^2 = r + 1 \).

Roots \( r = (1 \pm \sqrt{5})/2. \)
Fibonacci numbers are Benford base 10.

\[ a_{n+1} = a_n + a_{n-1}. \]

Guess \( a_n = r^n: \ r^{n+1} = r^n + r^{n-1} \) or \( r^2 = r + 1. \)

Roots \( r = (1 \pm \sqrt{5})/2. \)

General solution: \( a_n = c_1 r^n_1 + c_2 r^n_2. \)
Examples

Fibonacci numbers are Benford base 10.

\[ a_{n+1} = a_n + a_{n-1}. \]

Guess \( a_n = r^n: r^{n+1} = r^n + r^{n-1} \) or \( r^2 = r + 1 \).

Roots \( r = \left(1 \pm \sqrt{5}\right)/2. \)

General solution: \( a_n = c_1 r_1^n + c_2 r_2^n. \)

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

- **Fibonacci numbers are Benford base 10.**
  
  \[ a_{n+1} = a_n + a_{n-1}. \]
  
  Guess \( a_n = r^n: \) \( r^{n+1} = r^n + r^{n-1} \) or \( r^2 = r + 1. \)
  
  Roots \( r = (1 \pm \sqrt{5})/2. \)
  
  General solution: \( a_n = c_1 r_1^n + c_2 r_2^n. \)
  
  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:**
Fibonacci numbers are Benford base 10.

\[ a_{n+1} = a_n + a_{n-1}. \]

Guess \( a_n = r^n \): \( r^{n+1} = r^n + r^{n-1} \) or \( r^2 = r + 1 \).

Roots \( r = (1 \pm \sqrt{5})/2 \).

General solution: \( a_n = c_1 r_1^n + c_2 r_2^n \).

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:

\[ a_{n+1} = 2a_n \]
Examples

- Fibonacci numbers are Benford base 10.
  \[ a_{n+1} = a_n + a_{n-1}. \]
  Guess \( a_n = r^n: \) \( r^{n+1} = r^n + r^{n-1} \) or \( r^2 = r + 1. \)
  Roots \( r = \left(1 \pm \sqrt{5}\right)/2. \)
  General solution: \( a_n = c_1 r_1^n + c_2 r_2^n. \)
  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:
  \( \diamond \) \( a_{n+1} = 2a_n - a_{n-1} \)
Examples

Fibonacci numbers are Benford base 10.

\[ a_{n+1} = a_n + a_{n-1}. \]

Guess \( a_n = r^n: \ r^{n+1} = r^n + r^{n-1} \) or \( r^2 = r + 1. \)

Roots \( r = (1 \pm \sqrt{5})/2. \)

General solution: \( a_n = c_1 r_1^n + c_2 r_2^n. \)

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:

\[ \Diamond \ a_{n+1} = 2a_n - a_{n-1} \]
\[ \Diamond \ \text{take } a_0 = a_1 = 1 \text{ or } a_0 = 0, \ a_1 = 1. \]
## Digits of $2^n$

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

<table>
<thead>
<tr>
<th></th>
<th>1024</th>
<th>1048576</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>2048</td>
<td>2097152</td>
</tr>
<tr>
<td>4</td>
<td>4096</td>
<td>4194304</td>
</tr>
<tr>
<td>8</td>
<td>8192</td>
<td>8388608</td>
</tr>
<tr>
<td>16</td>
<td>16384</td>
<td>16777216</td>
</tr>
<tr>
<td>32</td>
<td>32768</td>
<td>33554432</td>
</tr>
<tr>
<td>64</td>
<td>65536</td>
<td>67108864</td>
</tr>
<tr>
<td>128</td>
<td>131072</td>
<td>134217728</td>
</tr>
<tr>
<td>256</td>
<td>262144</td>
<td>268435456</td>
</tr>
<tr>
<td>512</td>
<td>524288</td>
<td>536870912</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>digit</th>
<th>#</th>
<th>Obs Prob</th>
<th>Benf Prob</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>18</td>
<td>.300</td>
<td>.301</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>12</td>
<td>.200</td>
<td>.176</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>6</td>
<td>.100</td>
<td>.125</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>6</td>
<td>.100</td>
<td>.097</td>
</tr>
<tr>
<td>5</td>
<td>5</td>
<td>6</td>
<td>.100</td>
<td>.079</td>
</tr>
<tr>
<td>6</td>
<td>6</td>
<td>4</td>
<td>.067</td>
<td>.067</td>
</tr>
<tr>
<td>7</td>
<td>7</td>
<td>2</td>
<td>.033</td>
<td>.058</td>
</tr>
<tr>
<td>8</td>
<td>8</td>
<td>5</td>
<td>.083</td>
<td>.051</td>
</tr>
<tr>
<td>9</td>
<td>9</td>
<td>1</td>
<td>.017</td>
<td>.046</td>
</tr>
</tbody>
</table>
## Digits of $2^n$

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

<table>
<thead>
<tr>
<th>n</th>
<th>Value of $2^n$</th>
<th>First 5 digits</th>
<th>Digit</th>
<th>Frequency</th>
<th>Obs Prob</th>
<th>Benf Prob</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1024</td>
<td>1048576</td>
<td>1</td>
<td>18</td>
<td>.300</td>
<td>.301</td>
</tr>
<tr>
<td>2</td>
<td>2048</td>
<td>2097152</td>
<td>2</td>
<td>12</td>
<td>.200</td>
<td>.176</td>
</tr>
<tr>
<td>4</td>
<td>4096</td>
<td>4194304</td>
<td>3</td>
<td>6</td>
<td>.100</td>
<td>.125</td>
</tr>
<tr>
<td>8</td>
<td>8192</td>
<td>8388608</td>
<td>4</td>
<td>6</td>
<td>.100</td>
<td>.097</td>
</tr>
<tr>
<td>16</td>
<td>16384</td>
<td>16777216</td>
<td>5</td>
<td>6</td>
<td>.100</td>
<td>.079</td>
</tr>
<tr>
<td>32</td>
<td>32768</td>
<td>33554432</td>
<td>6</td>
<td>4</td>
<td>.067</td>
<td>.067</td>
</tr>
<tr>
<td>64</td>
<td>65536</td>
<td>67108864</td>
<td>7</td>
<td>2</td>
<td>.033</td>
<td>.058</td>
</tr>
<tr>
<td>128</td>
<td>131072</td>
<td>134217728</td>
<td>8</td>
<td>5</td>
<td>.083</td>
<td>.051</td>
</tr>
<tr>
<td>256</td>
<td>262144</td>
<td>268435456</td>
<td>9</td>
<td>1</td>
<td>.017</td>
<td>.046</td>
</tr>
</tbody>
</table>
First 60 values of $2^n$ (only displaying 30): $2^{10} = 1024 \approx 10^3$.

<table>
<thead>
<tr>
<th>$2^n$</th>
<th>value</th>
<th>Value</th>
<th>digit</th>
<th>#</th>
<th>Obs Prob</th>
<th>Benf Prob</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1024</td>
<td>1048576</td>
<td>1</td>
<td>18</td>
<td>.300</td>
<td>.301</td>
</tr>
<tr>
<td>2</td>
<td>2048</td>
<td>2097152</td>
<td>2</td>
<td>12</td>
<td>.200</td>
<td>.176</td>
</tr>
<tr>
<td>4</td>
<td>4096</td>
<td>4194304</td>
<td>3</td>
<td>6</td>
<td>.100</td>
<td>.125</td>
</tr>
<tr>
<td>8</td>
<td>8192</td>
<td>8388608</td>
<td>4</td>
<td>6</td>
<td>.100</td>
<td>.097</td>
</tr>
<tr>
<td>16</td>
<td>16384</td>
<td>16777216</td>
<td>5</td>
<td>6</td>
<td>.100</td>
<td>.079</td>
</tr>
<tr>
<td>32</td>
<td>32768</td>
<td>33554432</td>
<td>6</td>
<td>4</td>
<td>.067</td>
<td>.067</td>
</tr>
<tr>
<td>64</td>
<td>65536</td>
<td>67108864</td>
<td>7</td>
<td>2</td>
<td>.033</td>
<td>.058</td>
</tr>
<tr>
<td>128</td>
<td>131072</td>
<td>134217728</td>
<td>8</td>
<td>5</td>
<td>.083</td>
<td>.051</td>
</tr>
<tr>
<td>256</td>
<td>262144</td>
<td>268435456</td>
<td>9</td>
<td>1</td>
<td>.017</td>
<td>.046</td>
</tr>
<tr>
<td>512</td>
<td>524288</td>
<td>536870912</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Logarithms and Benford’s Law

\[ \chi^2 \text{ values for } \alpha^n, \ 1 \leq n \leq N \ (5\% \ 15.5). \]

<table>
<thead>
<tr>
<th>( N )</th>
<th>( \chi^2(\gamma) )</th>
<th>( \chi^2(e) )</th>
<th>( \chi^2(\pi) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>0.72</td>
<td>0.30</td>
<td>46.65</td>
</tr>
<tr>
<td>200</td>
<td>0.24</td>
<td>0.30</td>
<td>8.58</td>
</tr>
<tr>
<td>400</td>
<td>0.14</td>
<td>0.10</td>
<td>10.55</td>
</tr>
<tr>
<td>500</td>
<td>0.08</td>
<td>0.07</td>
<td>2.69</td>
</tr>
<tr>
<td>700</td>
<td>0.19</td>
<td>0.04</td>
<td>0.05</td>
</tr>
<tr>
<td>800</td>
<td>0.04</td>
<td>0.03</td>
<td>6.19</td>
</tr>
<tr>
<td>900</td>
<td>0.09</td>
<td>0.09</td>
<td>1.71</td>
</tr>
<tr>
<td>1000</td>
<td>0.02</td>
<td>0.06</td>
<td>2.90</td>
</tr>
</tbody>
</table>
Logarithms and Benford’s Law: Base 10 (5%: \( \log(\chi^2) \approx 2.74 \))

\[
\log(\chi^2) \text{ vs } N \text{ for } \pi^n \text{ (red) and } e^n \text{ (blue), } \\
n \in \{1, \ldots, N\}.
\]
Logarithms and Benford’s Law: Base 10 (5%: $\log(\chi^2) \approx 2.74$)

$log(\chi^2) \text{ vs } N \text{ for } \pi^n \text{ (red) and } e^n \text{ (blue)}, \ n \in \{1, \ldots, N\}. \ \text{Note } \pi^{175} \approx 1.0028 \cdot 10^{87}$.
Why Benford’s Law?
Not all data sets satisfy Benford’s Law.

- Long street \([1, L]\): \(L = 199\) versus \(L = 999\).
- Oscillates b/w 1/9 and 5/9 with first digit 1.
Not all data sets satisfy Benford’s Law.

- Long street \([1, L]\): \(L = 199\) versus \(L = 999\).
- Oscillates b/w \(1/9\) and \(5/9\) with first digit \(1\).

Probability first digit 1 versus street length \(L\).
Not all data sets satisfy Benford’s Law.

- Long street \([1, L]\): \(L = 199\) versus \(L = 999\).
- Oscillates b/w 1/9 and 5/9 with first digit 1.

Probability first digit 1 versus \(\log(\text{street length } L)\).
Streets

Not all data sets satisfy Benford’s Law.

- Long street \([1, L]\): \(L = 199\) versus \(L = 999\).
- Oscillates b/w 1/9 and 5/9 with first digit 1.

Probability first digit 1 versus \(\log(\text{street length } L)\).

What if we have many streets of different lengths?
Amalgamating Streets

All houses: 1000 Streets, each from 1 to 10000.

First digit and first two digits vs Benford.
Amalgamating Streets

All houses: 1000 Streets, each from 1 to rand(10000).

First digit and first two digits vs Benford.
Amalgamating Streets

All houses: 1000 Streets, each 1 to rand(rand(10000)).

First digit and first two digits vs Benford.
Conclusion: More processes, closer to Benford.
Amalgamating Streets

All houses: 1000 Streets, each 1 to \( \text{rand}(\text{rand}(\text{rand}(10000))) \).

First digit and first two digits vs Benford.

Conclusion: More processes, closer to Benford.
Let $X$ be random variable with density $p(x)$:

- $p(x) \geq 0$; $\int_{-\infty}^{\infty} p(x)dx = 1$;
- $\text{Prob}(a \leq X \leq b) = \int_{a}^{b} p(x)dx$. 
Probability Review

Let $X$ be random variable with density $p(x)$:

- $p(x) \geq 0$; $\int_{-\infty}^{\infty} p(x)dx = 1$;
- $\text{Prob} \left( a \leq X \leq b \right) = \int_{a}^{b} p(x)dx$.

Mean $\mu = \int_{-\infty}^{\infty} xp(x)dx$. 
Probability Review

- Let $X$ be random variable with density $p(x)$:
  - $p(x) \geq 0$; $\int_{-\infty}^{\infty} p(x)dx = 1$;
  - $\text{Prob}(a \leq X \leq b) = \int_{a}^{b} p(x)dx$.
- Mean $\mu = \int_{-\infty}^{\infty} xp(x)dx$.
- Variance $\sigma^2 = \int_{-\infty}^{\infty} (x - \mu)^2 p(x)dx$. 

\begin{figure}
\centering
\includegraphics[width=0.8\textwidth]{probability_graph.png}
\end{figure}
Let $X$ be random variable with density $p(x)$:

- $p(x) \geq 0; \int_{-\infty}^{\infty} p(x)dx = 1$;
- $\text{Prob}(a \leq X \leq b) = \int_{a}^{b} p(x)dx$.

- **Mean** $\mu = \int_{-\infty}^{\infty} xp(x)dx$.
- **Variance** $\sigma^2 = \int_{-\infty}^{\infty} (x - \mu)^2 p(x)dx$.
- **Independence**: knowledge of one random variable gives no knowledge of the other.
Central Limit Theorem

Normal $N(\mu, \sigma^2)$: $p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$.

**Theorem**

If $X_1, X_2, \ldots$ independent, identically distributed random variables (mean $\mu$, variance $\sigma^2$, finite moments) then

$$S_N := \frac{X_1 + \cdots + X_N - N\mu}{\sigma\sqrt{N}}$$

converges to $N(0, 1)$. 
Central Limit Theorem: Sums of Uniform Random Variables

\( X_i \sim \text{Unif}(-1/2, 1/2) \) (adjusted to mean 0, variance 1)

\[ Y_1 = \frac{X_1}{\sigma_{X_1}} \text{ vs } \mathcal{N}(0, 1). \]
Central Limit Theorem: Sums of Uniform Random Variables

$X_i \sim \text{Unif}(-1/2, 1/2)$ (adjusted to mean 0, variance 1)

\[
Y_2 = \frac{(X_1 + X_2)}{\sigma_{X_1+X_2}} \text{ vs } N(0, 1).
\]
Central Limit Theorem: Sums of Uniform Random Variables

\( X_i \sim \text{Unif}(-1/2, 1/2) \) (adjusted to mean 0, variance 1)

\[
Y_4 = \frac{(X_1 + X_2 + X_3 + X_4)}{\sigma_{X_1+X_2+X_3+X_4}} \text{ vs } N(0, 1).
\]
Central Limit Theorem: Sums of Uniform Random Variables

\( X_i \sim \text{Unif}(-1/2, 1/2) \) (adjusted to mean 0, variance 1)

\[
Y_8 = \frac{(X_1 + \cdots + X_8)}{\sigma_{X_1+\cdots+X_8}} \text{ vs } N(0, 1).
\]
Central Limit Theorem: Sums of Uniform Random Variables

\( X_i \sim \text{Unif}(-1/2, 1/2) \) (adjusted to mean 0, variance 1)

Density of \( Y_4 = (X_1 + \cdots + X_4)/\sigma_{X_1+\cdots+X_4} \)

\[
\begin{align*}
\frac{1}{27} \left( 18 + 9 \sqrt{3} \, y - \sqrt{3} \, y^3 \right) & \quad y = 0 \\
\frac{1}{18} \left( 12 - 6 \, y^2 - \sqrt{3} \, y^3 \right) & \quad -\sqrt{3} < y < 0 \\
\frac{1}{54} \left( 72 - 36 \sqrt{3} \, y + 18 \, y^2 - \sqrt{3} \, y^3 \right) & \quad \sqrt{3} < y < 2 \sqrt{3} \\
\frac{1}{54} \left( 18 \sqrt{3} \, y - 18 \, y^2 + \sqrt{3} \, y^3 \right) & \quad y = \sqrt{3} \\
\frac{1}{18} \left( 12 - 6 \, y^2 + \sqrt{3} \, y^3 \right) & \quad 0 < y < \sqrt{3} \\
\frac{1}{54} \left( 72 + 36 \sqrt{3} \, y + 18 \, y^2 + \sqrt{3} \, y^3 \right) & \quad -2 \sqrt{3} < y \leq -\sqrt{3} \\
0 & \quad \text{True} \\
\sqrt{3} & \quad \text{True}
\end{align*}
\]

(Don’t even think of asking to see \( Y_8 \)’s!)
As $\sigma \to \infty$, $N(0, \sigma^2) \mod 1 \to \text{Unif}(0, 1)$. 

Variance is .01.
Normal Distributions Mod 1

As $\sigma \to \infty$, $N(0, \sigma^2)$ mod 1 $\to$ Unif(0, 1).

Variance is .1.
As $\sigma \to \infty$, $N(0, \sigma^2)$ mod 1 $\to$ Unif$(0, 1)$.

Variance is .5.
Products and Benford’s Law

**Pavlovian Response:** See a product, take a logarithm.
Pavlovian Response: See a product, take a logarithm.

\[ X_1, X_2, \ldots \text{ nice, } W_N = X_1 \cdot X_2 \cdots X_N. \]
Products and Benford’s Law

Pavlovian Response: See a product, take a logarithm.

\[ X_1, X_2, \ldots \text{ nice, } W_N = X_1 \cdot X_2 \cdots X_N. \]

\[ Y_i = \log_{10} X_i, \quad V_N := \log_{10} W_N. \]
Products and Benford’s Law

Pavlovian Response: See a product, take a logarithm.

\[ X_1, X_2, \ldots \text{ nice, } W_N = X_1 \cdot X_2 \cdot \ldots X_N. \]

\[ Y_i = \log_{10} X_i, \quad V_N := \log_{10} W_N. \]

\[ V_N = \log_{10}(X_1 \cdot X_2 \cdot \ldots X_N) \]
Products and Benford’s Law

Pavlovian Response: See a product, take a logarithm.

\[ X_1, X_2, \ldots \text{ nice, } W_N = X_1 \cdot X_2 \cdots X_N. \]

\[ Y_i = \log_{10} X_i, \ V_N := \log_{10} W_N. \]

\[
V_N = \log_{10}(X_1 \cdot X_2 \cdots X_N) \\
= \log_{10} X_1 + \log_{10} X_2 + \cdots + \log_{10} X_N
\]
Products and Benford's Law

Pavlovian Response: See a product, take a logarithm.

\[ X_1, X_2, \ldots \text{ nice, } W_N = X_1 \cdot X_2 \cdots X_N. \]

\[ Y_i = \log_{10} X_i, \ V_N := \log_{10} W_N. \]

\[ V_N = \log_{10}(X_1 \cdot X_2 \cdots X_N) \]
\[ = \log_{10} X_1 + \log_{10} X_2 + \cdots + \log_{10} X_N \]
\[ = Y_1 + Y_2 + \cdots + Y_N. \]

Need distribution of \( V_N \) mod 1, which by CLT becomes uniform, implying Benfordness!
Applications
Applications for the IRS: Detecting Fraud

A Tale of Two Steve Millers....
### Applications for the IRS: Detecting Fraud

![Image of a U.S. Individual Income Tax Return 1989](image)

#### Introduction
- **Why Benford?**

#### Applications
- **ζ(s) = 3x + 1**

#### Stick Decomposition

#### Conclusions
- Refs

---

**Note:** The image shows a U.S. Individual Income Tax Return from 1989. This type of document is used for filing tax returns and includes various sections for reporting income, deductions, and credits. The example shown reflects a hypothetical scenario for detecting fraud through the use of Benford's Law.
Applications for the IRS: Detecting Fraud
Bank Fraud

- Audit of a bank revealed huge spike of numbers starting with 82
Bank Fraud

- Audit of a bank revealed huge spike of numbers starting with 4
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.
Data Integrity: Stream Flow Statistics: 130 years, 457,440 records
Election Fraud: Iran 2009

Numerous questions over Iran’s 2009 elections.

Lot of analysis; data moderately suspicious:

- First and second leading digits;
- Last two digits (should almost be uniform);
- Last two digits differing by at least 2.

Warning: enough tests, even if nothing wrong will find a suspicious result (but when all tests are on the boundary...).
The Riemann Zeta Function $\zeta(s)$ and Benford’s Law
Riemann Zeta Function (for real part of $s$ greater than 1)

$$\zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s}$$
Riemann Zeta Function (for real part of $s$ greater than 1)

$$\zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s} = \prod_{\text{prime } p} \left(1 - \frac{1}{p^s}\right)^{-1}, \quad \Re(s) > 1.$$
Riemann Zeta Function (for real part of \( s \) greater than 1)

\[
\zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s} = \prod_{p \text{ prime}} \left(1 - \frac{1}{p^s}\right)^{-1}, \quad \text{Re}(s) > 1.
\]

Geometric Series Formula: 
\[
(1 - x)^{-1} = 1 + x + x^2 + \cdots.
\]

Unique Factorization: 
\[
n = p_1^{r_1} \cdots p_m^{r_m}.
\]
Riemann Zeta Function (for real part of $s$ greater than 1)

\[
\zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s} = \prod_{p \text{ prime}} \left(1 - \frac{1}{p^s}\right)^{-1}, \quad \text{Re}(s) > 1.
\]

Geometric Series Formula: \((1 - x)^{-1} = 1 + x + x^2 + \cdots\).
Unique Factorization: \(n = p_1^{r_1} \cdots p_m^{r_m}\).

\[
\prod_{p} \left(1 - \frac{1}{p^s}\right)^{-1} = \left[1 + \frac{1}{2^s} + \left(\frac{1}{2^s}\right)^2 + \cdots\right] \left[1 + \frac{1}{3^s} + \left(\frac{1}{3^s}\right)^2 + \cdots\right] \cdots
\]

\[
= \sum_n \frac{1}{n^s}.
\]
Riemann Zeta Function (cont)

\[ \zeta(s) = \sum_{n} \frac{1}{n^s} = \prod_{p} \left(1 - \frac{1}{p^s}\right)^{-1}, \quad \text{Re}(s) > 1 \]

\[ \pi(x) = \#\{p : p \text{ is prime}, p \leq x\} \]

Properties of \( \zeta(s) \) and Primes:
Riemann Zeta Function (cont)

\[
\zeta(s) = \sum_{n} \frac{1}{n^s} = \prod_{p} \left(1 - \frac{1}{p^s}\right)^{-1}, \quad \text{Re}(s) > 1
\]

\[
\pi(x) = \#\{p : p \text{ is prime, } p \leq x\}
\]

Properties of \(\zeta(s)\) and Primes:

- \(\lim_{s \to 1^+} \zeta(s) = \infty, \pi(x) \to \infty\).
Riemann Zeta Function (cont)

\[ \zeta(s) = \sum_{n} \frac{1}{n^s} = \prod_{p} \left(1 - \frac{1}{p^s}\right)^{-1}, \quad \text{Re}(s) > 1 \]

\[ \pi(x) = \#\{p : p \text{ is prime, } p \leq x\} \]

Properties of \( \zeta(s) \) and Primes:

- \( \lim_{s \to 1^+} \zeta(s) = \infty, \pi(x) \to \infty. \)
- \( \zeta(2) = \frac{\pi^2}{6}, \pi(x) \to \infty. \)
The Riemann Zeta Function and Benford’s Law

$$|\zeta \left( \frac{1}{2} + i \frac{k}{4} \right)|, \; k \in \{0, 1, \ldots, 65535\}.$$
The Riemann Zeta Function and Benford’s Law

\[ |\zeta \left( \frac{1}{2} + ik \frac{4}{4} \right) |, \ k \in \{0, 1, \ldots, 65535\}. \]

First digits of \( |\zeta \left( \frac{1}{2} + ik \frac{4}{4} \right) | \) versus Benford’s law.
Proof Sketch: ‘Good’ $L$-Functions

We say an $L$-function is **good** if:

- **Euler product:**

\[
L(s, f) = \sum_{n=1}^{\infty} \frac{a_f(n)}{n^s} = \prod_p \prod_{j=1}^{d} \left(1 - \alpha_{f,j}(p) p^{-s}\right)^{-1}.
\]

- $L(s, f)$ has a meromorphic continuation to $\mathbb{C}$, is of finite order, and has at most finitely many poles (all on the line $\text{Re}(s) = 1$).

- **Functional equation:**

\[
e^{i\omega} G(s)L(s, f) = e^{-i\omega} G(1 - \overline{s})L(1 - \overline{s}),
\]

where $\omega \in \mathbb{R}$ and

\[
G(s) = Q^s \prod_{i=1}^{h} \Gamma(\lambda_i s + \mu_i)
\]

with $Q, \lambda_i > 0$ and $\text{Re}(\mu_i) \geq 0$. 


Proof Sketch: ‘Good’ $L$-Functions (cont)

- For some $\kappa > 0$, $c \in \mathbb{C}$, $x \geq 2$ we have
  \[ \sum_{p \leq x} \frac{|a_f(p)|^2}{p} = \kappa \log \log x + c + O\left(\frac{1}{\log x}\right). \]

- The $\alpha_{f,j}(p)$ are (Ramanujan-Petersson) tempered: $|\alpha_{f,j}(p)| \leq 1$.

- If $N(\sigma, T)$ is the number of zeros $\rho$ of $L(s)$ with $\text{Re}(\rho) \geq \sigma$ and $\text{Im}(\rho) \in [0, T]$, then for some $\beta > 0$ we have
  \[ N(\sigma, T) = O\left(T^{1-\beta\left(\sigma - \frac{1}{2}\right)} \log T\right). \]

Known in some cases, such as $\zeta(s)$ and Hecke cuspidal forms of full level and even weight $k > 0$. 
Log-Normal Law

\[
\mu \left( \left\{ t \in [T, 2T] : \log |L(\sigma + it, f)| \in [a, b] \right\} \right) = \frac{1}{T} \int_a^b \frac{e^{-\pi u^2 / \psi(\sigma, T)}}{\sqrt{\psi(\sigma, T)}} du + \text{Error}
\]

\[
\psi(\sigma, T) = \aleph \log \left[ \min \left( \log T, \frac{1}{\sigma - \frac{1}{2}} \right) \right] + O(1)
\]

\[
\frac{1}{2} \leq \sigma \leq \frac{1}{2} + \frac{1}{\log^\delta T}, \quad \delta \in (0, 1).
\]
Result: Values of $L$-functions and Benford’s Law

**Theorem** (Kontorovich and M–, 2005)

$L(s, f)$ a good $L$-function, as $T \to \infty$, $L(\sigma_T + it, f)$ is Benford.

**Ingredients**

- Approximate $\log L(\sigma_T + it, f)$ with $\sum_{n \leq x} \frac{c(n)\Lambda(n)}{\log n} \frac{1}{n^{\sigma_T+it}}$.
- Study moments $\int_T^{2T} |\cdot|, \ k \leq \log^{1-\delta} T$.
- Montgomery-Vaughan: $\int_T^{2T} \sum a_n n^{-it} \sum b_m m^{-it} \, dt = H \sum a_n \overline{b_n} + O(1) \sqrt{\sum n|a_n|^2 \sum n|b_n|^2}$. 
Results: Explicit $L$-Function Statement

**Theorem (Kontorovich-Miller ’05)**

Let $L(s, f)$ be a good $L$-function. Fix a $\delta \in (0, 1)$. For each $T$, let $\sigma_T = \frac{1}{2} + \frac{1}{\log \delta T}$. Then as $T \to \infty$

\[
\mu \left\{ t \in [T, 2T] : \frac{M_B(|L(\sigma_T + it, f)|)}{T} \leq \tau \right\} \to \log_B \tau
\]

Thus the values of the $L$-function satisfy Benford’s Law in the limit for any base $B$. 
The $3x + 1$ Problem
and
Benford's Law
3x + 1 Problem

- Kakutani (conspiracy), Erdös (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), Erdös (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
3x + 1 Problem

- Kakutani (conspiracy), Erdös (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), Erdös (not ready).

- x odd, \( T(x) = \frac{3x+1}{2^k}, 2^k \mid |3x + 1| \).

- Conjecture: for some \( n = n(x), T^n(x) = 1 \).

- \( 7 \rightarrow_1 11 \rightarrow_1 17 \)
3x + 1 Problem

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

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

$x$ odd, $T(x) = \frac{3x+1}{2^k}, 2^k \mid 3x + 1$.

Conjecture: for some $n = n(x)$, $T^n(x) = 1$.

$7 \rightarrow_1 11 \rightarrow_1 17 \rightarrow_2 13 \rightarrow_3 5 \rightarrow_4 1 \rightarrow_2 1$, 

$$
\begin{align*}
\zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s} = \frac{1}{1-2^s} \zeta(s) \prod_{p \text{ odd}} \frac{1}{1-p^{-s}}
\end{align*}
$$
**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: \((k_1, \ldots, k_m)\).
Heuristic Proof of $3x + 1$ Conjecture

\[
a_{n+1} = T(a_n)
\]

\[
\mathbb{E}[\log a_{n+1}] \approx \sum_{k=1}^{\infty} \frac{1}{2^k} \log \left( \frac{3a_n}{2^k} \right)
\]

\[
= \log a_n + \log 3 - \log 2 \sum_{k=1}^{\infty} \frac{k}{2^k}
\]

\[
= \log a_n + \log \left( \frac{3}{4} \right).
\]

Geometric Brownian Motion, drift $\log(3/4) < 1$. 
Theorem (Kontorovich and M–, 2005)

As \( m \to \infty \), \( x_m/(3/4)^m x_0 \) 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.
Structure Theorem: Sinai, Kontorovich-Sinai

\[ P(A) = \lim_{N \to \infty} \frac{\# \{ n \leq N : n \equiv 1,5 \text{ mod } 6, n \in A \}}{\# \{ n \leq N : n \equiv 1,5 \text{ mod } 6 \}}. \]

\((k_1, \ldots, k_m)\): two full arithm progressions:

\[ 6 \cdot 2^{k_1 + \cdots + k_m} p + q. \]

**Theorem (Sinai, Kontorovich-Sinai)**

\(k_i\)-values are i.i.d. r.v. (geometric, 1/2):
Structure Theorem: Sinai, Kontorovich-Sinai

\[ P(A) = \lim_{N \to \infty} \frac{\#\{n \leq N: n \equiv 1,5 \mod 6, n \in A\}}{\#\{n \leq N: n \equiv 1,5 \mod 6\}} \cdot \]

\((k_1, \ldots, k_m)\): two full arithm progressions:
\[6 \cdot 2^{k_1 + \cdots + k_m} p + q.\]

Theorem (Sinai, Kontorovich-Sinai)

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

\[ P \left( \log_2 \left[ \frac{x_m}{\left( \frac{3}{4} \right)^m x_0} \right] \leq a \right) = P \left( \frac{S_m - 2m}{\sqrt{2m}} \leq a \right) \]
Structure Theorem: Sinai, Kontorovich-Sinai

\[ P(A) = \lim_{N \to \infty} \frac{\#\{n \leq N: n \equiv 1, 5 \mod 6, n \in A\}}{\#\{n \leq N: n \equiv 1, 5 \mod 6\}}. \]

\( (k_1, \ldots, k_m) \): two full arithm progressions:

\[ 6 \cdot 2^{k_1 + \cdots + k_m} p + q. \]

Theorem (Sinai, Kontorovich-Sinai)

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

\[
P \left( \frac{\log_2 \left[ \frac{x_m}{\left( \frac{3}{4} \right)^m x_0} \right]}{(\log_2 B)\sqrt{2m}} \leq a \right) = P \left( \frac{S_m - 2m}{(\log_2 B)\sqrt{2m}} \leq a \right)
\]
Structure Theorem: Sinai, Kontorovich-Sinai

\[ P(A) = \lim_{N \to \infty} \frac{\#\{n \leq N: n \equiv 1,5 \mod 6, n \in A\}}{\#\{n \leq N: n \equiv 1,5 \mod 6\}} \cdot \]

\( (k_1, \ldots, k_m) \): two full arithm progressions:

\[ 6 \cdot 2^{k_1+\ldots+k_m}p + q. \]

Theorem (Sinai, Kontorovich-Sinai)

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

\[ P \left( \log_B \left[ \frac{x_m}{\left(\frac{3}{4}\right)^m x_0} \right] \leq a \right) = P \left( \frac{(S_m-2m)}{\log_2 B} \leq a \right) \]
Sketch of the proof of Benfordness

- Failed Proof: lattices, bad errors.

- CLT: \((S_m - 2m)/\sqrt{2m} \to N(0, 1)\):

\[
P(S_m - 2m = k) = \frac{\eta(k/\sqrt{m})}{\sqrt{m}} + O\left(\frac{1}{g(m)\sqrt{m}}\right).
\]

- Quantified Equidistribution:

\(I_\ell = \{\ell M, \ldots, (\ell + 1)M - 1\}, \ M = m^c, \ c < 1/2\)

\(k_1, k_2 \in I_\ell: \left|\eta\left(\frac{k_1}{\sqrt{m}}\right) - \eta\left(\frac{k_2}{\sqrt{m}}\right)\right|\) small

\(C = \log_B 2\) of irrationality type \(\kappa < \infty\):

\(#\{k \in I_\ell : kC \in [a, b]\} = M(b - a) + O(M^{1+\epsilon-1/\kappa}).\)
Irrationality Type

\[ \alpha \text{ has irrationality type } \kappa \text{ if } \kappa \text{ is the supremum of all } \gamma \text{ with } \]
\[
\lim_{q \to \infty} q^{\gamma+1} \min_{p} \left| \frac{\alpha}{q} - \frac{p}{q} \right| = 0.
\]

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

\( \alpha_1, \ldots, \alpha_n \) algebraic numbers height \( A_j \geq 4 \),
\( \beta_1, \ldots, \beta_n \in \mathbb{Q} \) with height at most \( B \geq 4 \),
\[
\Lambda = \beta_1 \log \alpha_1 + \cdots + \beta_n \log \alpha_n.
\]

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

Gives \( \log_{10} 2 \) of finite type, with \( \kappa < 1.2 \cdot 10^{602} \):
\[
|\log_{10} 2 - p/q| = |q \log 2 - p \log 10| / q \log 10.
\]
Quantified Equidistribution

**Theorem (Erdős-Turan)**

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

There is a \( C \) such that for all \( m \):

\[
D_N \leq C \cdot \left( \frac{1}{m} + \sum_{h=1}^{m} \frac{1}{h} \right) \left( \frac{1}{N} \sum_{n=1}^{N} e^{2\pi ihx_n} \right)
\]
Proof of Erdös-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 $\kappa$, $\sum_{h=1}^{m} \frac{1}{h||h\alpha||} = O\left(m^{\kappa-1+\epsilon}\right)$, take $m = \left\lceil N^{1/\kappa} \right\rceil$. 
80,514 iterations \(((4/3)^n = a_0\) predicts 80,319); \(\chi^2 = 13.5\) (5\% 15.5).

<table>
<thead>
<tr>
<th>Digit</th>
<th>Number</th>
<th>Observed</th>
<th>Benford</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>24251</td>
<td>0.301</td>
<td>0.301</td>
</tr>
<tr>
<td>2</td>
<td>14156</td>
<td>0.176</td>
<td>0.176</td>
</tr>
<tr>
<td>3</td>
<td>10227</td>
<td>0.127</td>
<td>0.125</td>
</tr>
<tr>
<td>4</td>
<td>7931</td>
<td>0.099</td>
<td>0.097</td>
</tr>
<tr>
<td>5</td>
<td>6359</td>
<td>0.079</td>
<td>0.079</td>
</tr>
<tr>
<td>6</td>
<td>5372</td>
<td>0.067</td>
<td>0.067</td>
</tr>
<tr>
<td>7</td>
<td>4476</td>
<td>0.056</td>
<td>0.058</td>
</tr>
<tr>
<td>8</td>
<td>4092</td>
<td>0.051</td>
<td>0.051</td>
</tr>
<tr>
<td>9</td>
<td>3650</td>
<td>0.045</td>
<td>0.046</td>
</tr>
</tbody>
</table>
241,344 iterations, $\chi^2 = 11.4$ (5% 15.5).

<table>
<thead>
<tr>
<th>Digit</th>
<th>Number</th>
<th>Observed</th>
<th>Benford</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>72924</td>
<td>0.302</td>
<td>0.301</td>
</tr>
<tr>
<td>2</td>
<td>42357</td>
<td>0.176</td>
<td>0.176</td>
</tr>
<tr>
<td>3</td>
<td>30201</td>
<td>0.125</td>
<td>0.125</td>
</tr>
<tr>
<td>4</td>
<td>23507</td>
<td>0.097</td>
<td>0.097</td>
</tr>
<tr>
<td>5</td>
<td>18928</td>
<td>0.078</td>
<td>0.079</td>
</tr>
<tr>
<td>6</td>
<td>16296</td>
<td>0.068</td>
<td>0.067</td>
</tr>
<tr>
<td>7</td>
<td>13702</td>
<td>0.057</td>
<td>0.058</td>
</tr>
<tr>
<td>8</td>
<td>12356</td>
<td>0.051</td>
<td>0.051</td>
</tr>
<tr>
<td>9</td>
<td>11073</td>
<td>0.046</td>
<td>0.046</td>
</tr>
</tbody>
</table>
### 5x + 1 Data: random 10,000 digit number, $2^k || 5x + 1$

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

<table>
<thead>
<tr>
<th>Digit</th>
<th>Number</th>
<th>Observed</th>
<th>Benford</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>8154</td>
<td>0.302</td>
<td>0.301</td>
</tr>
<tr>
<td>2</td>
<td>4770</td>
<td>0.177</td>
<td>0.176</td>
</tr>
<tr>
<td>3</td>
<td>3405</td>
<td>0.126</td>
<td>0.125</td>
</tr>
<tr>
<td>4</td>
<td>2634</td>
<td>0.098</td>
<td>0.097</td>
</tr>
<tr>
<td>5</td>
<td>2105</td>
<td>0.078</td>
<td>0.079</td>
</tr>
<tr>
<td>6</td>
<td>1787</td>
<td>0.066</td>
<td>0.067</td>
</tr>
<tr>
<td>7</td>
<td>1568</td>
<td>0.058</td>
<td>0.058</td>
</tr>
<tr>
<td>8</td>
<td>1357</td>
<td>0.050</td>
<td>0.051</td>
</tr>
<tr>
<td>9</td>
<td>1224</td>
<td>0.045</td>
<td>0.046</td>
</tr>
</tbody>
</table>
$5x + 1$ Data: random 10,000 digit number, $2 | 5x + 1$

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

<table>
<thead>
<tr>
<th>Digit</th>
<th>Number</th>
<th>Observed</th>
<th>Benford</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>72652</td>
<td>0.301</td>
<td>0.301</td>
</tr>
<tr>
<td>2</td>
<td>42499</td>
<td>0.176</td>
<td>0.176</td>
</tr>
<tr>
<td>3</td>
<td>30153</td>
<td>0.125</td>
<td>0.125</td>
</tr>
<tr>
<td>4</td>
<td>23388</td>
<td>0.097</td>
<td>0.097</td>
</tr>
<tr>
<td>5</td>
<td>19110</td>
<td>0.079</td>
<td>0.079</td>
</tr>
<tr>
<td>6</td>
<td>16159</td>
<td>0.067</td>
<td>0.067</td>
</tr>
<tr>
<td>7</td>
<td>13995</td>
<td>0.058</td>
<td>0.058</td>
</tr>
<tr>
<td>8</td>
<td>12345</td>
<td>0.051</td>
<td>0.051</td>
</tr>
<tr>
<td>9</td>
<td>11043</td>
<td>0.046</td>
<td>0.046</td>
</tr>
</tbody>
</table>
Stick Decomposition
Fixed Proportion Decomposition Process

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 ’13)

Conjecture: The above decomposition process is Benford as $N \to \infty$ for any $p \in (0, 1)$, $p \neq \frac{1}{2}$.

(B) $p = 0.51$ and $N = 10000$. Benford distribution overlaid.
Conjecture: The above decomposition process is Benford as $N \to \infty$ for any $p \in (0, 1)$, $p \neq \frac{1}{2}$.

Counterexample (SMALL ’13): $p = \frac{1}{11}$, $1 - p = \frac{10}{11}$. 
Benford Analysis

At $N^{th}$ level,

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

$$p^N \left( \frac{1 - p}{p} \right)^j, \quad j \in \{0, \ldots, N\}, \quad \text{have } \binom{N}{j} \text{ times.}$$
Benford Analysis

At $N^{th}$ level,

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

$$p^N \left(\frac{1 - p}{p}\right)^j, \quad j \in \{0, \ldots, N\}, \quad \text{have } \binom{N}{j} \text{ times.}$$

(Weighted) Geometric with ratio $\frac{1-p}{p} = 10^y$; behavior depends on irrationality of $y$!
Benford Analysis

At $N^{th}$ level,

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

$$\rho^N \left( \frac{1 - \rho}{\rho} \right)^j, \quad j \in \{0, \ldots, N\}, \text{ have } \binom{N}{j} \text{ times.}$$

(Weighted) Geometric with ratio $\frac{1 - \rho}{\rho} = 10^y$; behavior depends on irrationality of $y$!

Theorem: Benford if and only if $y$ irrational.
Examples

\[ p = \frac{3}{11}, \text{ 1000 levels; } y = \log_{10}(\frac{8}{3}) \notin \mathbb{Q} \]  
(irrational)
Examples

\[ p = \frac{1}{11}, \text{ 1000 levels;} \ y = 1 \in \mathbb{Q} \]

(rational)
$\rho = \frac{1}{1 + 10^{33/10}}$, 1000 levels; $y = \frac{33}{10} \in \mathbb{Q}$ (rational)
Conclusions
Current / Future Investigations

- Develop more sophisticated tests for fraud.

- Study digits of other systems.
  - Break rod of fixed length a variable number of times.
  - Break rods of variable length a variable number of times.
  - Break rods of variable length, each piece then breaks with given probability.
  - Break rods of variable integer length, each piece breaks until is a prime, or a square, ....
Conclusions and Future Investigations

- See many different systems exhibit Benford behavior.

- Ingredients of proofs (logarithms, equidistribution).

- Applications to fraud detection / data integrity.
References


http://www.jstor.org/stable/2684926?seq=1#
page_scan_tab_contents.


