Benford’s law, or: Why the IRS cares about number theory!

Steven J. Miller, Williams College

sjm1@williams.edu
http://www.williams.edu/Mathematics/sjmillermiller/

Hampshire College, July 14, 2011
Interesting Question

For a nice data set, such as the Fibonacci numbers, what percent of the leading digits are 1?
Interesting Question

For a nice data set, such as the Fibonacci numbers, what percent of the leading digits are 1?

Plausible answers:
Interesting Question

For a nice data set, such as the Fibonacci numbers, what percent of the leading digits are 1?

Plausible answers: 10%
Interesting Question

For a nice data set, such as the Fibonacci numbers, what percent of the leading digits are 1?

Plausible answers: 10%, 11%
Interesting Question

For a nice data set, such as the Fibonacci numbers, what percent of the leading digits are 1?

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

- State Benford’s Law.
- Discuss examples and applications.
- Sketch proofs.
- Describe open problems.
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.
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.
  - 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.
  - 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
First digit of number of hits Google returns for first 80 distinct words and numbers from Wikipedia’s *Bastille Day* entry on July 10, 2011. Chi-square of 46.2, highly non-Benford (99% of time should be at most 20).
Riemann Zeta Function

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

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

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

Mantissa: \( x = M_{10}(x) \cdot 10^k, \ k \text{ integer.} \)

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

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

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

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 \( y_n \mod 1 \in [a, b] \) 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: \( \log_{10} 2 \), \( \log_{10} \left( \frac{1+\sqrt{5}}{2} \right) \notin \mathbb{Q} \).
Equidistribution and Benford’s Law

Equidistribution

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

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

- **Examples:** \( \log_{10} 2 \), \( \log_{10} \left( \frac{1+\sqrt{5}}{2} \right) \notin \mathbb{Q} \).
  
  - **Proof:** if rational: \( 2 = 10^{p/q} \).
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: \( \beta \not\in \mathbb{Q} \), \( n\beta \) is equidistributed mod 1.

- Examples: \( \log_{10} 2 \), \( \log_{10} \left( \frac{1+\sqrt{5}}{2} \right) \not\in \mathbb{Q} \).

  *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 \( y_i = \log_B x_i \).
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

- Fibonacci numbers are Benford base 10.
Examples

Fibonacci numbers are Benford base 10.

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

- Fibonacci numbers are Benford base 10.
  \[ a_{n+1} = a_n + a_{n-1}. \]
  Guess \[ a_n = r^n: \ r^{n+1} = r^n + r^{n-1} \text{ or } r^2 = r + 1. \]
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: \( a_n = c_1 r_1^n + c_2 r_2^n. \)
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: \( a_n = \frac{1}{\sqrt{5}} \left( \frac{1+\sqrt{5}}{2} \right)^n - \frac{1}{\sqrt{5}} \left( \frac{1-\sqrt{5}}{2} \right)^n. \)
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:**
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:

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

- Fibonacci numbers are Benford base 10.

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

Guess \[ a_n = r^n: \quad r^{n+1} = r^n + r^{n-1} \] 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:

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

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

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

<table>
<thead>
<tr>
<th>$2^n$</th>
<th>Digit</th>
<th>Observation</th>
<th>Obs Prob</th>
<th>Benf Prob</th>
</tr>
</thead>
<tbody>
<tr>
<td>1024</td>
<td>1</td>
<td>18</td>
<td>.300</td>
<td>.301</td>
</tr>
<tr>
<td>2048</td>
<td>2</td>
<td>12</td>
<td>.200</td>
<td>.176</td>
</tr>
<tr>
<td>4096</td>
<td>3</td>
<td>6</td>
<td>.100</td>
<td>.125</td>
</tr>
<tr>
<td>8192</td>
<td>4</td>
<td>6</td>
<td>.100</td>
<td>.097</td>
</tr>
<tr>
<td>16384</td>
<td>5</td>
<td>6</td>
<td>.100</td>
<td>.079</td>
</tr>
<tr>
<td>32768</td>
<td>6</td>
<td>4</td>
<td>.067</td>
<td>.067</td>
</tr>
<tr>
<td>65536</td>
<td>7</td>
<td>2</td>
<td>.033</td>
<td>.058</td>
</tr>
<tr>
<td>131072</td>
<td>8</td>
<td>5</td>
<td>.083</td>
<td>.051</td>
</tr>
<tr>
<td>524288</td>
<td>9</td>
<td>1</td>
<td>.017</td>
<td>.046</td>
</tr>
</tbody>
</table>
Logarithms and Benford’s Law

$\chi^2$ 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

\[ \log(\chi^2) \text{ vs } N \text{ for } \pi^n \text{ (red) and } e^n \text{ (blue)}, \]
\[ n \in \{1, \ldots, N\}. \text{ Note } \pi^{175} \approx 1.0028 \cdot 10^{87}, (5\%, \]
\[ \log(\chi^2) \approx 2.74). \]
Applications
Applications for the IRS: Detecting Fraud
Applications for the IRS: Detecting Fraud
Bank Fraud

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

Numerous protests/complaints 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 $3x + 1$ Problem and Benford’s Law
3x + 1 Problem

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

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

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

- 7
3x + 1 Problem

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

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

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

- $7 \rightarrow_1 11$
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, $T(x) = \frac{3x + 1}{2^k}$, $2^k \mid 3x + 1$.

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

- $7 \rightarrow_1 11 \rightarrow_1 17 \rightarrow_2 13 \rightarrow_3 5$
3x + 1 Problem

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

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

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

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

- $7 \rightarrow_1 11 \rightarrow_1 17 \rightarrow_2 13 \rightarrow_3 5 \rightarrow_4 1 \rightarrow_2 1$, 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 $\log(3/4) < 1$. 
3x + 1 and Benford

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.
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).

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

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>
Products of Random Variables
Preliminaries

- $X_1 \cdots X_n \Leftrightarrow Y_1 + \cdots + Y_n \text{ mod } 1, \ Y_i = \log_B X_i$

- Density $Y_i$ is $g_i$, density $Y_i + Y_j$ is

$$ (g_i \ast g_j)(y) = \int_0^1 g_i(t)g_j(y - t)dt. $$

- $h_n = g_1 \ast \cdots \ast g_n, \ \hat{g}(\xi) = \hat{g}_1(\xi) \cdots \hat{g}_n(\xi)$. 
Modulo 1 Central Limit Theorem

Theorem (M– and Nigrini 2007)

\{ Y_m \} independent continuous random variables on \([0, 1]\) (not necc. i.i.d.), densities \(\{g_m\}\). 
\(Y_1 + \cdots + Y_M \mod 1\) converges to the uniform distribution as \(M \to \infty\) in \(L^1([0, 1])\) if and only if for all \(n \neq 0\), 
\[ \lim_{M \to \infty} \widehat{g}_1(n) \cdots \widehat{g}_M(n) = 0. \]

◊ Gives info on rate of convergence.
Proof under stronger conditions

- Use standard CLT to show $Y_1 + \cdots + Y_M$ tends to a Gaussian.

- Use Poisson Summation to show the Gaussian tends to the uniform modulo 1.
Proof under stronger conditions

**Figure:** Plot of normal (mean 0, stdev 1).
Proof under stronger conditions

Figure: Plot of normal (mean 0, stdev .1) modulo 1.
Proof under stronger conditions

Figure: Plot of normal (mean 0, stdev .5) modulo 1.
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


