Benford’s Law: Tables of Logarithms, Tax Cheats, and The Leading Digit Phenomenon

Michelle Manes (mmanes@math.hawaii.edu)

9 September, 2008
(1881) Simon Newcomb publishes “Note on the frequency of use of the different digits in natural numbers.” The world ignores it.
(1881) Simon Newcomb publishes “Note on the frequency of use of the different digits in natural numbers.” The world ignores it.

(1938) Frank Benford (unaware of Newcomb’s work, presumably) publishes “The law of anomalous numbers.”
Newcomb noticed that the early pages of the book of tables of logarithms were much dirtier than the later pages, so were presumably referenced more often.
Newcomb noticed that the early pages of the book of tables of logarithms were much dirtier than the later pages, so were presumably referenced more often.

He stated the rule this way:

$$\text{Prob(first significant digit } = d) = \log_{10} \left( 1 + \frac{1}{d} \right).$$
Benford Distribution

**Definition**

A sequence of positive numbers \( \{x_n\} \) is **Benford** if

\[
\text{Prob(first significant digit } = d) = \log_{10} \left( 1 + \frac{1}{d} \right).
\]
Benford Distribution

Definition

A sequence of positive numbers \( \{x_n\} \) is **Benford** if

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

A sequence of positive numbers \( \{x_n\} \) is **Benford base** \( b \) if

\[
\text{Prob}(\text{first significant digit} = d) = \log_b \left(1 + \frac{1}{d}\right).
\]
Benford’s Law

**Base 10 Predictions**

<table>
<thead>
<tr>
<th>digit</th>
<th>probability it occurs as a leading digit</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>30.1%</td>
</tr>
<tr>
<td>2</td>
<td>17.6%</td>
</tr>
<tr>
<td>3</td>
<td>12.5%</td>
</tr>
<tr>
<td>4</td>
<td>9.7%</td>
</tr>
<tr>
<td>5</td>
<td>7.9%</td>
</tr>
<tr>
<td>6</td>
<td>6.7%</td>
</tr>
<tr>
<td>7</td>
<td>5.8%</td>
</tr>
<tr>
<td>8</td>
<td>5.1%</td>
</tr>
<tr>
<td>9</td>
<td>4.6%</td>
</tr>
</tbody>
</table>
# Benford’s Data

## Table I

<table>
<thead>
<tr>
<th>Group</th>
<th>Title</th>
<th>First Digit</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>A</td>
<td>Rivers, Area</td>
<td>31.0</td>
<td>64.1</td>
</tr>
<tr>
<td>B</td>
<td>Population</td>
<td>33.9</td>
<td>20.4</td>
</tr>
<tr>
<td>C</td>
<td>Constants</td>
<td>41.3</td>
<td>14.4</td>
</tr>
<tr>
<td>D</td>
<td>Newspapers</td>
<td>30.0</td>
<td>18.0</td>
</tr>
<tr>
<td>E</td>
<td>Spec. Heat</td>
<td>24.0</td>
<td>18.4</td>
</tr>
<tr>
<td>F</td>
<td>Pressure</td>
<td>29.6</td>
<td>18.3</td>
</tr>
<tr>
<td>G</td>
<td>H.P. Lost</td>
<td>30.0</td>
<td>18.4</td>
</tr>
<tr>
<td>H</td>
<td>Mol. Wgt.</td>
<td>26.7</td>
<td>25.2</td>
</tr>
<tr>
<td>I</td>
<td>Drainage</td>
<td>27.1</td>
<td>23.9</td>
</tr>
<tr>
<td>J</td>
<td>Atomic Wgt.</td>
<td>47.2</td>
<td>18.7</td>
</tr>
<tr>
<td>K</td>
<td>( n^{-1}, \sqrt{n}, \cdot \cdot \cdot )</td>
<td>25.7</td>
<td>20.3</td>
</tr>
<tr>
<td>L</td>
<td>Design</td>
<td>26.8</td>
<td>14.8</td>
</tr>
<tr>
<td>M</td>
<td>Digest</td>
<td>33.4</td>
<td>18.5</td>
</tr>
<tr>
<td>N</td>
<td>Cost Data</td>
<td>32.4</td>
<td>18.8</td>
</tr>
<tr>
<td>O</td>
<td>X-Ray Volts</td>
<td>27.9</td>
<td>17.5</td>
</tr>
<tr>
<td>P</td>
<td>Am. League</td>
<td>32.7</td>
<td>17.6</td>
</tr>
<tr>
<td>Q</td>
<td>Black Body</td>
<td>31.0</td>
<td>17.3</td>
</tr>
<tr>
<td>R</td>
<td>Addresses</td>
<td>28.9</td>
<td>19.2</td>
</tr>
<tr>
<td>S</td>
<td>( n^1, n^2 \cdot \cdot \cdot n )</td>
<td>25.3</td>
<td>16.0</td>
</tr>
<tr>
<td>T</td>
<td>Death Rate</td>
<td>27.0</td>
<td>18.6</td>
</tr>
<tr>
<td></td>
<td><strong>Average</strong></td>
<td><strong>30.6</strong></td>
<td><strong>18.5</strong></td>
</tr>
<tr>
<td></td>
<td><strong>Probable Error</strong></td>
<td><strong>±0.8</strong></td>
<td>±0.4</td>
</tr>
</tbody>
</table>

---

**Note:** The table provides the percentage of times each digit (1 to 9) appears as the first digit in various real-world datasets, along with the count for each digit in the dataset, and the average and probable error for each digit.
More Data

Benford’s Law compared with: numbers from the front pages of newspapers, U.S. county populations, and the Dow Jones Industrial Average.
Dow Illustrates Benford’s Law

Suppose the Dow Jones average is about $1,000. If the average goes up at a rate of about 20% a year, it would take four years to get from 1 to 2 as a first digit. If we start with a first digit 5, it only requires a 20% increase to get from $5,000 to $6,000, and that is achieved in one year.

When the Dow reaches $9,000, it takes only an 11% increase and just seven months to reach the $10,000 mark. This again has first digit 1, so it will take another doubling (and four more years) to get back to first digit 2.
Dow Illustrates Benford’s Law

Suppose the Dow Jones average is about $1,000. If the average goes up at a rate of about 20% a year, it would take four years to get from 1 to 2 as a first digit.
Dow Illustrates Benford’s Law

Suppose the Dow Jones average is about $1,000. If the average goes up at a rate of about 20% a year, it would take four years to get from 1 to 2 as a first digit.

If we start with a first digit 5, it only requires a 20% increase to get from $5,000 to $6,000, and that is achieved in one year.
Dow Illustrates Benford’s Law

Suppose the Dow Jones average is about $1,000. If the average goes up at a rate of about 20% a year, it would take four years to get from 1 to 2 as a first digit.

If we start with a first digit 5, it only requires a 20% increase to get from $5,000 to $6,000, and that is achieved in one year.

When the Dow reaches $9,000, it takes only an 11% increase and just seven months to reach the $10,000 mark. This again has first digit 1, so it will take another doubling (and four more years) to get back to first digit 2.
Benford’s Law and Tax Fraud (Nigrini, 1992)

Most people can’t fake data convincingly. Many states (including California) and the IRS now use fraud-detection software based on Benford’s Law.
Benford’s Law and Tax Fraud (Nigrini, 1992)

Most people can’t fake data convincingly.
Most people can’t fake data convincingly.

Many states (including California) and the IRS now use fraud-detection software based on Benford’s Law.
True Life Tale

Manager from Arizona State Treasurer was embezzling funds.
True Life Tale

- Manager from Arizona State Treasurer was embezzling funds.
- Most amounts were below $100,000 (critical threshold for checks that would require more scrutiny).
Manager from Arizona State Treasurer was embezzling funds.

Most amounts were below $100,000 (critical threshold for checks that would require more scrutiny).

Over 90% of the checks had a first digit 7, 8, or 9. (Trying to get close to the threshold without going over — artificially changes the data and so breaks fit with Benford’s law.)
**True Life Tale**

**Exhibit 3: Check Fraud in Arizona**

The table lists the checks that a manager in the office of the Arizona State Treasurer wrote to divert funds for his own use. The vendors to whom the checks were issued were fictitious.

<table>
<thead>
<tr>
<th>Date of Check</th>
<th>Amount</th>
</tr>
</thead>
<tbody>
<tr>
<td>October 9, 1992</td>
<td>$ 1,927.48</td>
</tr>
<tr>
<td></td>
<td>27,902.31</td>
</tr>
<tr>
<td>October 14, 1992</td>
<td>86,241.90</td>
</tr>
<tr>
<td></td>
<td>72,117.46</td>
</tr>
<tr>
<td></td>
<td>81,321.75</td>
</tr>
<tr>
<td></td>
<td>97,473.96</td>
</tr>
<tr>
<td>October 19, 1992</td>
<td>93,249.11</td>
</tr>
<tr>
<td></td>
<td>89,658.17</td>
</tr>
<tr>
<td></td>
<td>87,776.89</td>
</tr>
<tr>
<td></td>
<td>92,105.83</td>
</tr>
<tr>
<td></td>
<td>79,949.16</td>
</tr>
<tr>
<td></td>
<td>87,602.93</td>
</tr>
<tr>
<td></td>
<td>96,879.27</td>
</tr>
<tr>
<td></td>
<td>91,806.47</td>
</tr>
<tr>
<td></td>
<td>84,991.67</td>
</tr>
<tr>
<td></td>
<td>90,831.83</td>
</tr>
<tr>
<td></td>
<td>93,766.67</td>
</tr>
<tr>
<td></td>
<td>88,338.72</td>
</tr>
<tr>
<td></td>
<td>94,639.49</td>
</tr>
<tr>
<td></td>
<td>83,709.28</td>
</tr>
<tr>
<td></td>
<td>96,412.21</td>
</tr>
<tr>
<td></td>
<td>88,432.86</td>
</tr>
<tr>
<td></td>
<td>71,552.16</td>
</tr>
<tr>
<td><strong>TOTAL</strong></td>
<td>$ 1,878,687.58</td>
</tr>
</tbody>
</table>
What types of sequences are Benford?

Real-world data can be a good fit or not, depending on the type of data. Data that is a good fit is “suitably random” — comes in many different scales, and is a large and randomly distributed data set, with no artificial or external limitations on the range of the numbers.
What types of sequences are Benford?

Real-world data can be a good fit or not, depending on the type of data. Data that is a good fit is “suitably random” — comes in many different scales, and is a large and randomly distributed data set, with no artificial or external limitations on the range of the numbers.

Some numerical sequences are clearly not Benford:
What types of sequences are Benford?

Real-world data can be a good fit or not, depending on the type of data. Data that is a good fit is “suitably random” — comes in many different scales, and is a large and randomly distributed data set, with no artificial or external limitations on the range of the numbers.

Some numerical sequences are clearly not Benford:
- $1, 2, 3, 4, 5, 6, 7, \ldots$ (uniform distribution)
What types of sequences are Benford?

Real-world data can be a good fit or not, depending on the type of data. Data that is a good fit is “suitably random” — comes in many different scales, and is a large and randomly distributed data set, with no artificial or external limitations on the range of the numbers.

Some numerical sequences are clearly *not* Benford:

- $1, 2, 3, 4, 5, 6, 7, \ldots$ (uniform distribution)

- $1, 10, 100, 1000, \ldots$ (first digit is always 1)
Powers of Two
Powers of Two

1, 2, 4, 8, 16, 32, 64, ...
Powers of Two

1, 2, 4, 8, 16, 32, 64, \ldots

\[ t_n = 2^n \]
Powers of Two

Plot of first digit frequencies versus Benford’s Law.
Fibonacci Numbers

\[
F_{n+1} = F_n + F_{n-1}
\]
Fibonacci Numbers

1, 1, 2, 3, 5, 8, 13, 21, …
Fibonacci Numbers

1, 1, 2, 3, 5, 8, 13, 21, \ldots

\[ F_{n+1} = F_n + F_{n-1} \]
Mantissa Function

**Definition**

Define the *mantissa function*

\[ M : \mathbb{R}^+ \rightarrow [1, 10) \]

\[ x \mapsto r \]

where \( r \) is the unique number in \([1, 10)\) such that \( x = r \times 10^n \) for some \( n \in \mathbb{Z} \).
Mantissa Function

**Definition**

Define the *mantissa function*

\[ M : \mathbb{R}^+ \rightarrow [1, 10) \]

\[ x \mapsto r \]

where \( r \) is the unique number in \([1, 10)\) such that \( x = r \times 10^n \) for some \( n \in \mathbb{Z} \).

In other words, write the number in scientific notation.
Define the *mantissa function*

\[ M : \mathbb{R}^+ \rightarrow [1, 10) \]

\[ x \mapsto r \]

where \( r \) is the unique number in \([1, 10)\) such that \( x = r \times 10^n \) for some \( n \in \mathbb{Z} \).

**Examples**

- \( M(9,000,001) = 9.0000001 \).
- \( M(0.01247) = 1.247 \).
Logarithms and Benford’s Law

**Fundamental Equivalence**

Data set \( \{x_i\} \) is Benford if \( \{y_i\} \) is equidistributed mod 1, where \( y_i = \log_{10} x_i \).
Logarithms and Benford’s Law

Fundamental Equivalence

Data set \( \{x_i\} \) is Benford if \( \{y_i\} \) is equidistributed mod 1, where \( y_i = \log_{10} x_i \).

Proof:

\[ x = M(x) \cdot 10^k \text{ for some } k \in \mathbb{Z}. \]
Logarithms and Benford’s Law

**Fundamental Equivalence**

Data set \( \{x_i\} \) is Benford if \( \{y_i\} \) is equidistributed mod 1, where \( y_i = \log_{10} x_i \).

**Proof:**

- \( x = M(x) \cdot 10^k \) for some \( k \in \mathbb{Z} \).
- First digit of \( x \) is \( d \) iff \( d \leq M(x) < d + 1 \).
Logarithms and Benford’s Law

Fundamental Equivalence

Data set \( \{x_i\} \) is Benford if \( \{y_i\} \) is equidistributed mod 1, where \( y_i = \log_{10} x_i \).

Proof:

- \( x = M(x) \cdot 10^k \) for some \( k \in \mathbb{Z} \).
- First digit of \( x \) is \( d \) iff \( d \leq M(x) < d + 1 \).
- \( \log_{10} d \leq y < \log_{10}(d + 1) \), where
  \[ y = \log_{10}(M(x)) = \log_{10} x \mod 1. \]
Logarithms and Benford’s Law

Fundamental Equivalence

Data set \( \{x_i\} \) is Benford if \( \{y_i\} \) is equidistributed mod 1, where \( y_i = \log_{10} x_i \).

Proof:

- \( x = M(x) \cdot 10^k \) for some \( k \in \mathbb{Z} \).
- First digit of \( x \) is \( d \) iff \( d \leq M(x) < d + 1 \).
- \( \log_{10} d \leq y < \log_{10}(d + 1) \), where \( y = \log_{10}(M(x)) = \log_{10} x \mod 1 \).
- If the distribution is uniform (mod 1), then the probability \( y \) is in this range is

\[
\log_{10}(d+1) - \log_{10}(d) = \log_{10} \left( \frac{d + 1}{d} \right) = \log_{10} \left( 1 + \frac{1}{d} \right).
\]
Logarithms and Benford’s Law

**Fundamental Equivalence**

Data set \( \{ x_i \} \) is Benford if \( \{ y_i \} \) is equidistributed mod 1, where \( y_i = \log_{10} x_i \).
Logarithms and Benford’s Law

**Fundamental Equivalence**

Data set \( \{x_i\} \) is Benford if \( \{y_i\} \) is equidistributed mod 1, where \( y_i = \log_{10} x_i \).
Logarithms and Benford’s Law

**Fundamental Equivalence**

Data set \( \{x_i\} \) is Benford if \( \{y_i\} \) is equidistributed mod 1, where \( y_i = \log_{10} x_i \).

**Kronecker-Weyl Theorem**

If \( \beta \notin \mathbb{Q} \) then \( n\beta \mod 1 \) is equidistributed. (Thus if \( \log_{10} \alpha \notin \mathbb{Q} \), then \( \alpha^n \) is Benford.)
The sequence \( \{2^n\} \) for \( n \geq 0 \) is Benford.
The sequence \( \{2^n\} \) for \( n \geq 0 \) is Benford.

Proof:

- Consider the sequence of logarithms \( \{ n \log_{10} 2 \} \).
- By the Kronecker-Weyl Theorem, this is uniform (mod 1) because \( \log_{10} 2 \notin \mathbb{Q} \).
- Since the sequence of logarithms is uniformly distributed (mod 1), the original sequence is Benford.
Fibonacci Numbers

Theorem

The sequence \( \{F_n\} \) of Fibonacci numbers is Benford.
Theorem

The sequence \( \{F_n\} \) of Fibonacci numbers is Benford.

Heuristic Argument:

Closed form for Fibonacci numbers:

\[
F_n = \frac{1}{\sqrt{5}} \left[ \left( \frac{1 + \sqrt{5}}{2} \right)^n - \left( \frac{1 - \sqrt{5}}{2} \right)^n \right].
\]
The sequence \( \{F_n\} \) of Fibonacci numbers is Benford.

**Heuristic Argument:**

- Closed form for Fibonacci numbers:

\[
F_n = \frac{1}{\sqrt{5}} \left[ \left( \frac{1 + \sqrt{5}}{2} \right)^n - \left( \frac{1 - \sqrt{5}}{2} \right)^n \right].
\]

- \( \left| \left( \frac{1-\sqrt{5}}{2} \right) \right| < 1 \), so the leading digits are completely determined by \( \frac{1}{\sqrt{5}} \left( \frac{1+\sqrt{5}}{2} \right)^n \).
Fibonacci Numbers

Theorem

The sequence \( \{F_n\} \) of Fibonacci numbers is Benford.

Heuristic Argument:

- Closed form for Fibonacci numbers:

\[
F_n = \frac{1}{\sqrt{5}} \left[ \left( \frac{1 + \sqrt{5}}{2} \right)^n - \left( \frac{1 - \sqrt{5}}{2} \right)^n \right].
\]

- \( \left| \left( \frac{1 - \sqrt{5}}{2} \right) \right| < 1 \), so the leading digits are completely determined by \( \frac{1}{\sqrt{5}} \left( \frac{1 + \sqrt{5}}{2} \right)^n \).

- This sequence is Benford because \( \log_{10} \left( \frac{1 + \sqrt{5}}{2} \right) \not\in \mathbb{Q} \)
Consider the sequence \( \{a_n\} \) given by some initial conditions \( a_0, a_1, \ldots, a_{k-1} \) and then a recurrence relation

\[
a_{n+k} = c_1 a_{n+k-1} + c_2 a_{n+k-2} + \cdots + c_k a_n,
\]

with \( c_1, c_2, \ldots, c_k \) fixed real numbers.
Consider the sequence \( \{a_n\} \) given by some initial conditions \( a_0, a_1, \ldots, a_{k-1} \) and then a recurrence relation

\[
a_{n+k} = c_1 a_{n+k-1} + c_2 a_{n+k-2} + \cdots + c_k a_n,
\]

with \( c_1, c_2, \ldots, c_k \) fixed real numbers.

Find the eigenvalues of the recurrence relation and order them so that \( |\lambda_1| \geq |\lambda_2| \geq \cdots \geq |\lambda_k| \).
Linear Recurrence Sequences

Consider the sequence \( \{ a_n \} \) given by some initial conditions \( a_0, a_1, \ldots, a_{k-1} \) and then a recurrence relation

\[
a_{n+k} = c_1 a_{n+k-1} + c_2 a_{n+k-2} + \cdots + c_k a_n,
\]

with \( c_1, c_2, \ldots, c_k \) fixed real numbers.

Find the eigenvalues of the recurrence relation and order them so that \( |\lambda_1| \geq |\lambda_2| \geq \cdots \geq |\lambda_k| \).

There exist numbers \( u_1, u_2, \ldots, u_k \) (which depend on the initial conditions) so that

\[
a_n = u_1 \lambda_1^n + u_2 \lambda_2^n + \cdots + u_k \lambda_k^n.
\]
Linear Recurrence Sequences

**Theorem**

*With a linear recurrence sequence as described, if \( \log_{10} |\lambda_1| \not\in \mathbb{Q} \) and the initial conditions are such that \( u_1 \neq 0 \), then the sequence \( \{a_n\} \) is Benford.*
Linear Recurrence Sequences

**Theorem**

*With a linear recurrence sequence as described, if \( \log_{10} |\lambda_1| \not\in \mathbb{Q} \) and the initial conditions are such that \( u_1 \neq 0 \), then the sequence \( \{a_n\} \) is Benford.*

**Sketch of Proof:**

- Rewrite the closed form as \( a_n = u_1 \lambda_1^n \left( 1 + \mathcal{O} \left( \frac{ku\lambda_2^n}{\lambda_1^n} \right) \right) \)

where \( u = \max_i |u_i| + 1. \)
### Linear Recurrence Sequences

**Theorem**

*With a linear recurrence sequence as described, if* \( \log_{10} |\lambda_1| \notin \mathbb{Q} \) *and the initial conditions are such that* \( u_1 \neq 0 \), *then the sequence* \( \{a_n\} \) *is Benford.*

**Sketch of Proof:**

- Rewrite the closed form as \( a_n = u_1 \lambda_1^n \left( 1 + \mathcal{O} \left( \frac{ku^n}{\lambda_1^n} \right) \right) \)
  where \( u = \max_i |u_i| + 1 \).

- Some clever algebra using our assumptions to find that \( y_n = \log_{10}(a_n) = n \log_{10} \lambda_1 + \log_{10} u_1 + \mathcal{O}(\beta^n) \) for some appropriate \( \beta \).
Linear Recurrence Sequences

**Theorem**

*With a linear recurrence sequence as described, if* \( \log_{10}|\lambda_1| \notin \mathbb{Q} \) *and the initial conditions are such that* \( u_1 \neq 0 \), *then the sequence* \( \{ a_n \} \) *is Benford.*

**Sketch of Proof:**

- Rewrite the closed form as \( a_n = u_1 \lambda_1^n \left( 1 + \mathcal{O} \left( \frac{ku\lambda_2^n}{\lambda_1^n} \right) \right) \)
  where \( u = \max_i |u_i| + 1 \).

- Some clever algebra using our assumptions to find that \( y_n = \log_{10}(a_n) = n \log_{10} \lambda_1 + \log_{10} u_1 + \mathcal{O}(\beta^n) \) for some appropriate \( \beta \).

- Show in the limit the error term affects a vanishingly small portion of the distribution.
Elliptic Divisibility Sequences

**Definition**

An *integral divisibility sequence* is a sequence of integers \(\{u_n\}\) satisfying

\[
u_n \mid u_m \quad \text{whenever} \quad n \mid m.
\]

An *elliptic divisibility sequence* is an integral divisibility sequence which satisfies the following recurrence relation for all \(m \geq n \geq 1\):

\[
u_{m+n}u_{m-n}u_1^2
\]

\[= u_{m+1}u_{m-1}u_n^2 - u_{n+1}u_{n-1}u_m^2.
\]
Boring Elliptic Divisibility Sequences

- The sequences of integers, where $u_n = n$. 
Boring Elliptic Divisibility Sequences

- The sequences of integers, where $u_n = n$.
- The sequence $0, 1, -1, 0, 1, -1, \ldots$. 

Boring Elliptic Divisibility Sequences

- The sequences of integers, where $u_n = n$.

- The sequence $0, 1, -1, 0, 1, -1, \ldots$.

- The sequence $1, 3, 8, 21, 55, 144, 377, 987, 2584, 6765, \ldots$ (this is every-other Fibonacci number).
Not-So-Boring Elliptic Divisibility Sequences

The sequences which begins
0, 1, 1, −1, 1, 2, −1, −3, −5, 7, −4, −28, 29, 59, 129, −314, −65, 1529, −3689, −8209, −16264, 833313, 113689, −620297, 2382785, 7869898, 7001471, −126742987, −398035821, 168705471, ... (This is sequence A006769 in the On-Line Encyclopedia of Integer Sequences.)
The sequences which begins
\[0, 1, 1, -1, 1, 2, -1, -3, -5, 7, -4, -28, 29, 59, 129, -314, -65, 1529, -3689, -8209, -16264, 833313, 113689, -620297, 2382785, 7869898, 7001471, -126742987, -398035821, 168705471, \ldots\]
(This is sequence A006769 in the On-Line Encyclopedia of Integer Sequences.)

The sequence which begins
\[1, 1, -3, 11, 38, 249, -2357, 8767, 496036, -3769372, -299154043, -12064147359, \ldots\]
Why We Like Elliptic Divisibility Sequences

- Special case of *Somos sequences*, which are interesting and an active area of research.
Why We Like Elliptic Divisibility Sequences

- Special case of *Somos sequences*, which are interesting and an active area of research.

- Connection to elliptic curves, also an active area of research.

\[
\begin{align*}
\text{Elliptic curve} & \quad \Rightarrow \quad \text{EDS: denominators of the sequence of points} \\
E \text{ over } \mathbb{Q} \quad & \quad \leftrightarrow \quad \{ P, 2P, 3P, \ldots \} \\
\text{with rational point} & \quad \text{of points} \\
P \text{ on } E &
\end{align*}
\]
Why We Like Elliptic Divisibility Sequences

- Special case of *Somos sequences*, which are interesting and an active area of research.

- Connection to elliptic curves, also an active area of research.

\[
\begin{align*}
\text{Elliptic curve } & \quad E \text{ over } \mathbb{Q} \\
\text{with rational point } & \quad P \text{ on } E \\
\end{align*}
\quad \leftrightarrow \quad
\begin{align*}
\text{EDS: denominators } & \quad \text{of the sequence} \\
\text{of points } & \quad \{P, 2P, 3P, \ldots\} \\
\end{align*}
\]

- Applications to elliptic curve cryptography.
Elliptic Divisibility Sequences are Benford?
Elliptic Divisibility Sequences are Benford?

Plot of first digit frequencies versus Benford's Law.
Elliptic Divisibility Sequences are Benford?

Plot of first digit frequencies versus Benford's Law.
It’s well-known that elliptic divisibility sequences satisfy a growth condition like $u_n \approx c^{n^2}$ where the constant $c$ depends on the arithmetic height of the point $P$ and on the curve $E$. 
Heuristic Argument

- It’s well-known that elliptic divisibility sequences satisfy a growth condition like \( u_n \approx c^{n^2} \) where the constant \( c \) depends on the arithmetic height of the point \( P \) and on the curve \( E \).

- Weyl’s theorem tells us that \( \{n^k \alpha\} \) is uniformly distributed \((\text{mod } 1)\) iff \( \alpha \not\in \mathbb{Q} \).
History

Applications

Benford and Integer Sequences

Formalism

Benford and Recurrence Relations

Heuristic Argument

- It’s well-known that elliptic divisibility sequences satisfy a growth condition like $u_n \approx c^{n^2}$ where the constant $c$ depends on the arithmetic height of the point $P$ and on the curve $E$.

- Weyl’s theorem tells us that $\{n^k \alpha\}$ is uniformly distributed (mod 1) iff $\alpha \notin \mathbb{Q}$.

- So we should at least be able to conclude that a given EDS is Benford base $b$ for almost every $b$. 
Heuristic Argument

- It’s well-known that elliptic divisibility sequences satisfy a growth condition like $u_n \approx c^{n^2}$ where the constant $c$ depends on the arithmetic height of the point $P$ and on the curve $E$.

- Weyl’s theorem tells us that $\{n^k \alpha\}$ is uniformly distributed (mod 1) iff $\alpha \notin \mathbb{Q}$.

- So we should at least be able to conclude that a given EDS is Benford base $b$ for almost every $b$.

- **But:** The argument with the big-$O$ error terms is delicate, and not enough is known in the case of EDS.