Benford's Law, Elliptic Divisibility Sequences, and Canonical Heights

Michelle Manes (mmanes@math.hawaii.edu)

Sage Days for Women
July, 2013

History

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

History

- (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.”

Statement of Benford's Law

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.

Statement of Benford's Law

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:

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

Benford's Law

Base 10 Predictions

digit	probability it occurs as a leading digit
1	30.1%
2	17.6%
3	12.5%
4	9.7%
5	7.9%
6	6.7%
7	5.8%
8	5.1%
9	4.6%

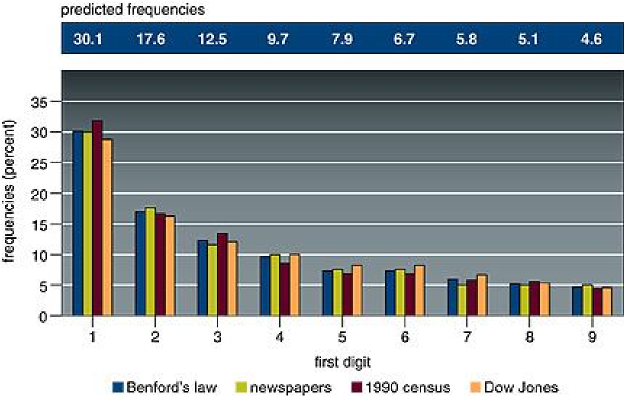
Benford's Data

TABLE I

PERCENTAGE OF TIMES THE NATURAL NUMBERS 1 TO 9 ARE USED AS FIRST DIGITS IN NUMBERS, AS DETERMINED BY 20,229 OBSERVATIONS

Group	Title	First Digit									Count
		1	2	3	4	5	6	7	8	9	
A	Rivers, Area	31.0	16.4	10.7	11.3	7.2	8.6	5.5	4.2	5.1	335
B	Population	33.9	20.4	14.2	8.1	7.2	6.2	4.1	3.7	2.2	3259
C	Constants	41.3	14.4	4.8	8.6	10.6	5.8	1.0	2.9	10.6	104
D	Newspapers	30.0	18.0	12.0	10.0	8.0	6.0	6.0	5.0	5.0	100
E	Spec. Heat	24.0	18.4	16.2	14.6	10.6	4.1	3.2	4.8	4.1	1389
F	Pressure	29.6	18.3	12.8	9.8	8.3	6.4	5.7	4.4	4.7	703
G	H.P. Lost	30.0	18.4	11.9	10.8	8.1	7.0	5.1	5.1	3.6	690
H	Mol. Wgt.	26.7	25.2	15.4	10.8	6.7	5.1	4.1	2.8	3.2	1800
I	Drainage	27.1	23.9	13.8	12.6	8.2	5.0	5.0	2.5	1.9	159
J	Atomic Wgt.	47.2	18.7	5.5	4.4	6.6	4.4	3.3	4.4	5.5	91
K	$n^{-1}, \sqrt{n}, \cdots$	25.7	20.3	9.7	6.8	6.6	6.8	7.2	8.0	8.9	5000
L	Design	26.8	14.8	14.3	7.5	8.3	8.4	7.0	7.3	5.6	560
M	*Digest*	33.4	18.5	12.4	7.5	7.1	6.5	5.5	4.9	4.2	308
N	Cost Data	32.4	18.8	10.1	10.1	9.8	5.5	4.7	5.5	3.1	741
O	X-Ray Volts	27.9	17.5	14.4	9.0	8.1	7.4	5.1	5.8	4.8	707
P	Am. League	32.7	17.6	12.6	9.8	7.4	6.4	4.9	5.6	3.0	1458
Q	Black Body	31.0	17.3	14.1	8.7	6.6	7.0	5.2	4.7	5.4	1165
R	Addresses	28.9	19.2	12.6	8.8	8.5	6.4	5.6	5.0	5.0	342
S	$n^1, n^2 \cdots n!$	25.3	16.0	12.0	10.0	8.5	8.8	6.8	7.1	5.5	900
T	Death Rate	27.0	18.6	15.7	9.4	6.7	6.5	7.2	4.8	4.1	418
Average		30.6	18.5	12.4	9.4	8.0	6.4	5.1	4.9	4.7	1011
Probable Error		±0.8	±0.4	±0.4	±0.3	±0.2	±0.2	±0.2	±0.2	±0.3	—

More Data



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

Example

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

Example

Suppose the Dow Jones average is about \$1K. If the average goes up at a rate of about 20% a year, it would take five 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 \$5K to \$6K, and that is achieved in one year.

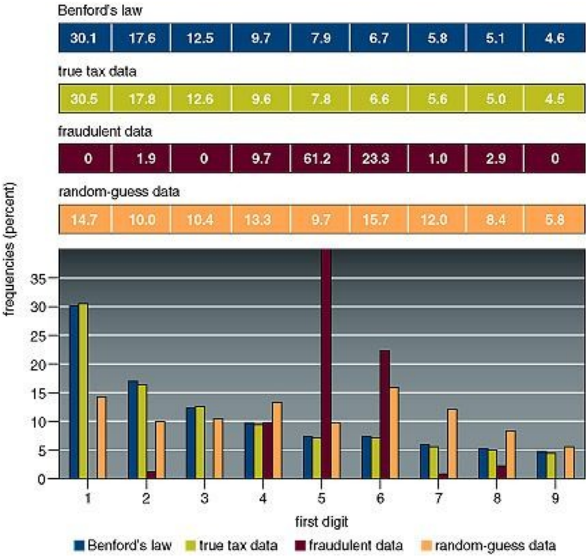
Example

Suppose the Dow Jones average is about \$1K. If the average goes up at a rate of about 20% a year, it would take five 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 \$5K to \$6K, and that is achieved in one year.

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

Benford's Law and Tax Fraud (Nigrini, 1992)



Benford's Law and Tax Fraud (Nigrini, 1992)

Most people can't fake data convincingly.

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.

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 $100K (critical threshold for checks that would require more scrutiny).

True Life Tale

- Manager from Arizona State Treasurer was embezzling funds.
- Most amounts were below \$100K (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.

Date of Check	Amount
October 9, 1992	$ 1,927.48
	27,902.31
October 14, 1992	86,241.90
	72,117.46
	81,321.75
	97,473.96
October 19, 1992	93,249.11
	89,658.17
	87,776.89
	92,105.83
	79,949.16
	87,602.93
	96,879.27
	91,806.47
	84,991.67
	90,831.83
	93,766.67
	88,338.72
	94,639.49
	83,709.28
	96,412.21
	88,432.86
	71,552.16
TOTAL	$ 1,878,687.58

Benford Base b

Definition

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

$$\text{Prob}(\text{first significant digit} = d) = \log_b\left(1 + \frac{1}{d}\right).$$

Problems with “Proofs” of Benford’s Law

- Discrete density and summability methods.

Problems with “Proofs” of Benford’s Law

- Discrete density and summability methods.
  $F_d = \{x \in \mathbb{N} \mid \text{first digit of } x \text{ is } d\}$. No natural density. That is,
  $$\lim_{n\to\infty} \frac{F_d \cap \{1, 2, \dots, n\}}{n}$$
  does not exist.

Problems with "Proofs" of Benford's Law

- Discrete density and summability methods.
- Continuous density and summability methods. (Same problem.)

Problems with “Proofs” of Benford’s Law

- Discrete density and summability methods.
- Continuous density and summability methods. (Same problem.)
- Scale invariance.

  If there is a reasonable first-digit law, it should be scale-invariant. That is, it shouldn’t matter if the measurements are in feet or meters, pounds or kilograms, etc.

Hill's Formulation (1988)

Definition

For each integer $b > 1$, define the *mantissa function*

$$\begin{aligned} M_b \colon \mathbb{R}^+ &\to [1, b) \\ x &\mapsto r \end{aligned}$$

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

Hill's Formulation (1988)

Definition

For each integer $b > 1$, define the *mantissa function*

$$M_b \colon \mathbb{R}^+ \to [1, b)$$
$$x \mapsto r$$

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

Examples

- $M_{10}(9) = 9 = M_{100}(9)$.
- $M_2(9) = 9/8 = 1.001$ (base 2).

Hill's Formulation (1988)

Definition

For $E \subset [1, b)$, let

$$\langle E \rangle_b = M_b^{-1}(E) = \bigcup_{n \in \mathbb{Z}} b^n E \subset \mathbb{R}^+.$$

Hill's Formulation (1988)

Definition

For $E \subset [1, b)$, let

$$\langle E \rangle_b = M_b^{-1}(E) = \bigcup_{n \in \mathbb{Z}} b^n E \subset \mathbb{R}^+.$$

Definition

$\mathcal{M}_b = \{\langle E \rangle_b \mid E \subset \mathbb{B}(1, b)\}$is the σ-algebra on $\mathbb{R}^+$ generated by M_b.

Hill's Formulation (1988)

Definition

Let P_b be the probability measure on $(\mathbb{R}^+, \mathcal{M}_b)$ defined by

$$P_b(\langle [1, \gamma) \rangle_b) = \log_b \gamma.$$

Hill's Formulation (1988)

Definition

Let P_b be the probability measure on $(\mathbb{R}^+, \mathcal{M}_b)$ defined by

$$P_b(\langle[1,\gamma)\rangle_b) = \log_b \gamma.$$

This probability measure:

Hill's Formulation (1988)

Definition

Let P_b be the probability measure on $(\mathbb{R}^+, \mathcal{M}_b)$ defined by

$$P_b(\langle[1,\gamma)\rangle_b) = \log_b \gamma.$$

This probability measure:

- Agrees with Benford's law.

Hill's Formulation (1988)

Definition

Let P_b be the probability measure on $(\mathbb{R}^+, \mathcal{M}_b)$ defined by

$$P_b(\langle[1,\gamma)\rangle_b) = \log_b \gamma.$$

This probability measure:

- Agrees with Benford's law.
- Is the unique scale-invariant probability measure on $(\mathbb{R}^+, \mathcal{M}_b)$.

Hill's Formulation (1988)

Definition

Let P_b be the probability measure on $(\mathbb{R}^+, \mathcal{M}_b)$ defined by

$$P_b(\langle [1, \gamma) \rangle_b) = \log_b \gamma.$$

This probability measure:

- Agrees with Benford's law.
- Is the unique scale-invariant probability measure on $(\mathbb{R}^+, \mathcal{M}_b)$.

Proof comes down to uniqueness of Haar measure.

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 distributed base-10:

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 distributed base-10:

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

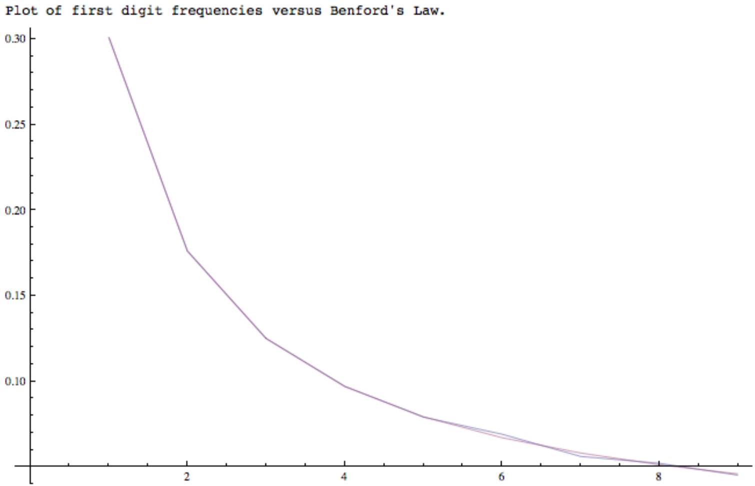
What types of sequences are Benford?

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 distributed base-10:

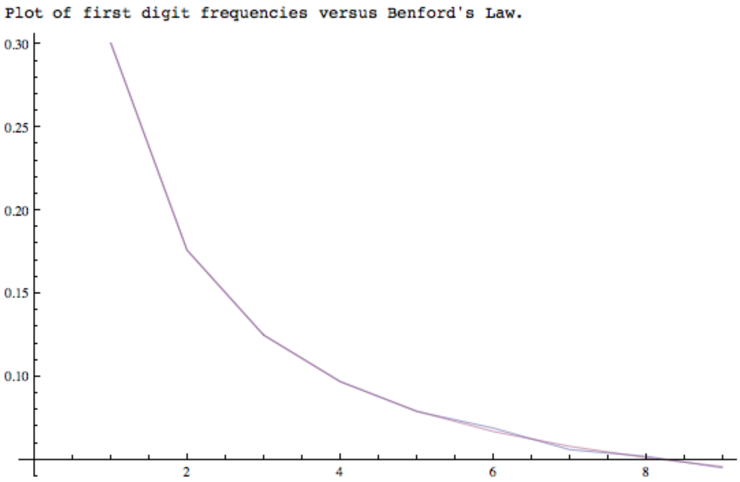
- $1, 2, 3, 4, 5, 6, 7, \ldots$ (uniform distribution)
- $1, 10, 100, 1000, \ldots$ (first digit is always 1)

Some numerical sequences seem to be a good fit



Powers of Two

Some numerical sequences seem to be a good fit



Fibonacci Numbers

Logarithms and Benford's Law

Fundamental Equivalence

Data set $\{x_i\}$ is Benford base b iff $\{y_i\}$ 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 iff $\{y_i\}$ is equidistributed mod 1, where $y_i = \log_b x_i$.

Proof:

- $x = M_b(x) \cdot b^k$ for some $k \in \mathbb{Z}$.

Logarithms and Benford's Law

Fundamental Equivalence

Data set $\{x_i\}$ is Benford base b iff $\{y_i\}$ is equidistributed mod 1, where $y_i = \log_b x_i$.

Proof:

- $x = M_b(x) \cdot b^k$ for some $k \in \mathbb{Z}$.
- First digit of x in base b is d iff $d \leq M_b(x) < d + 1$.

Logarithms and Benford's Law

Fundamental Equivalence

Data set $\{x_i\}$ is Benford base b iff $\{y_i\}$ is equidistributed mod 1, where $y_i = \log_b x_i$.

Proof:

- $x = M_b(x) \cdot b^k$ for some $k \in \mathbb{Z}$.
- First digit of x in base b is d iff $d \leq M_b(x) < d + 1$.
- $\log_b d \leq y < \log_b(d+1)$, where $y = \log_b(M_b(x)) = \log_b x \bmod 1$.

Logarithms and Benford's Law

Fundamental Equivalence

Data set $\{x_i\}$ is Benford base b iff $\{y_i\}$ is equidistributed mod 1, where $y_i = \log_b x_i$.

Proof:

- $x = M_b(x) \cdot b^k$ for some $k \in \mathbb{Z}$.
- First digit of x in base b is d iff $d \le M_b(x) < d + 1$.
- $\log_b d \le y < \log_b(d+1)$, where $y = \log_b(M_b(x)) = \log_b x \bmod 1$.
- If the distribution is uniform $\pmod 1$, then the probability y is in this range is

$$\log_b(d+1) - \log_b(d) = \log_b\left(\frac{d+1}{d}\right) = \log_b\left(1 + \frac{1}{d}\right).$$

Logarithms and Benford's Law

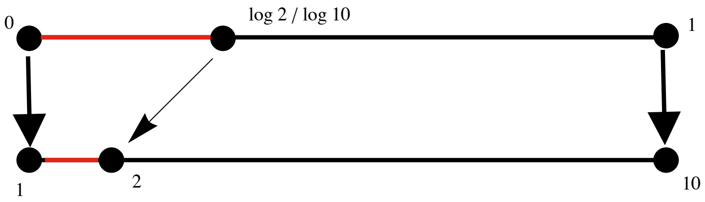
Fundamental Equivalence

Data set $\{x_i\}$ is Benford base b iff $\{y_i\}$ 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 iff $\{y_i\}$ 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 iff $\{y_i\}$ is equidistributed mod 1, where $y_i = \log_b x_i$.

Kronecker-Weyl Theorem

If $\beta \notin \mathbb{Q}$ then $n\beta$ mod 1 (resp. $n^2\beta$ mod 1) is equidistributed.

Thus if $\log_b \alpha \notin \mathbb{Q}$, then α^n (resp. α^{n^2}) is Benford.

Powers of 2

Theorem

The sequence $\{2^n\}$ *for* $n \geq 0$ *is Benford base* b *for any* b *that is not a rational power of* 2*.*

Powers of 2

Theorem

The sequence $\{2^n\}$ for $n \geq 0$ is Benford base b for any b that is not a rational power of 2.

Proof:

- Consider the sequence of logarithms $\{n(\log_b 2)\}$.
- By the Kronecker-Weyl Theorem, this is uniform (mod 1) as long as $\log_b 2 \notin \mathbb{Q}$.
- If b is not a rational power of 2, then the sequence of logarithms is uniformly distributed (mod 1), so the original sequence is Benford base b.

Fibonacci Numbers

Theorem

The sequence $\{F_n\}$ of Fibonacci numbers Benford base b for almost every b.

Fibonacci Numbers

Theorem

The sequence $\{F_n\}$ of Fibonacci numbers Benford base b for almost every b.

Heuristic Argument:

- Closed form for Fibonacci numbers:

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

Fibonacci Numbers

Theorem

The sequence $\{F_n\}$ of Fibonacci numbers Benford base b for almost every b.

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 Benford base b for almost every b.

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 will be Benford base-b for any b where $\log_b\left(\frac{1+\sqrt{5}}{2}\right) \notin \mathbb{Q}$.

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.

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 $|\lambda_1| \geq |\lambda_2| \geq \cdots \geq |\lambda_k|$.

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 number $u_1, u_2, \ldots, u_k$ (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 $\log_b |\lambda_1| \notin \mathbb{Q}$*and the initial conditions are such that* $u_1 \neq 0$*, then the sequence* $\{a_n\}$ *is Benford base b.*

Linear Recurrence Sequences

Theorem

With a linear recurrence sequence as described, if $\log_b |\lambda_1| \notin \mathbb{Q}$ and the initial conditions are such that $u_1 \neq 0$, then the sequence $\{a_n\}$ is Benford base b.

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 $\log_b |\lambda_1| \notin \mathbb{Q}$ and the initial conditions are such that $u_1 \neq 0$, then the sequence $\{a_n\}$ is Benford base b.

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 rewrite this as $a_n = u_1 \lambda_1^n (1 + \mathcal{O}(\beta^n))$.

Linear Recurrence Sequences

Theorem

With a linear recurrence sequence as described, if $\log_b|\lambda_1| \notin \mathbb{Q}$and the initial conditions are such that $u_1 \neq 0$, then the sequence $\{a_n\}$ is Benford base b.

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 rewrite this as $a_n = u_1\lambda_1^n(1 + \mathcal{O}(\beta^n))$.
- Then $y_n = \log_b(a_n) = n\log_b\lambda_1 + \log_b u_1 + \mathcal{O}(\beta^n)$.

Linear Recurrence Sequences

Theorem

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

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 rewrite this as $a_n = u_1 \lambda_1^n (1 + \mathcal{O}(\beta^n))$.
- Then $y_n = \log_b(a_n) = n \log_b \lambda_1 + \log_b u_1 + \mathcal{O}(\beta^n)$.
- 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

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

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

$$u_{m+n}u_{m-n}u_1^2 = u_{m+1}u_{m-1}u_n^2 - u_{n+1}u_{n-1}u_m^2. \qquad (*)$$

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

Not-So-Boring Elliptic Divisibility Sequences

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

- The sequence which begins
$1, 1, -3, 11, 38, 249, -2357, 8767, 496036, -3769372, -299154043, -12064147359, \ldots$.

That Recurrence Relation

$$u_{m+n}u_{m-n}u_1^2 = u_{m+1}u_{m-1}u_n^2 - u_{n+1}u_{n-1}u_m^2. \qquad (*)$$

If $u_1 = 1$, $u_2, u_3 \in \mathbb{Z} \setminus \{0\}$ and $u_4/u_2 \in \mathbb{Z} \setminus \{0\}$, then $u_n \in \mathbb{Z}$ for all n. Why?

That Recurrence Relation

$$u_{m+n}u_{m-n}u_1^2 = u_{m+1}u_{m-1}u_n^2 - u_{n+1}u_{n-1}u_m^2. \qquad (*)$$

If $u_1 = 1$, $u_2, u_3 \in \mathbb{Z} \setminus \{0\}$ and $u_4/u_2 \in \mathbb{Z} \setminus \{0\}$, then $u_n \in \mathbb{Z}$ for all n. Why?

- Induction.

That Recurrence Relation

$$u_{m+n}u_{m-n}u_1^2 = u_{m+1}u_{m-1}u_n^2 - u_{n+1}u_{n-1}u_m^2. \qquad (*)$$

If $u_1 = 1$, $u_2, u_3 \in \mathbb{Z} \setminus \{0\}$ and $u_4/u_2 \in \mathbb{Z} \setminus \{0\}$, then $u_n \in \mathbb{Z}$ for all n. Why?

- Induction.
- $|u_n|$ counts perfect matchings on certain graphs (Bousquet-Mélu–West, Speyer, others)

That Recurrence Relation

$$u_{m+n}u_{m-n}u_1^2 = u_{m+1}u_{m-1}u_n^2 - u_{n+1}u_{n-1}u_m^2. \qquad (*)$$

If $u_1 = 1$, $u_2, u_3 \in \mathbb{Z} \setminus \{0\}$ and $u_4/u_2 \in \mathbb{Z} \setminus \{0\}$, then $u_n \in \mathbb{Z}$ for all n. Why?

- Induction.
- $|u_n|$ counts perfect matchings on certain graphs (Bousquet-Mélu–West, Speyer, others)
- Laurentness of u_n in terms of u_1, u_2, u_3, u_4 (Fomin–Zelevinsky: cluster algebras)

That Recurrence Relation

$$u_{m+n}u_{m-n}u_1^2 = u_{m+1}u_{m-1}u_n^2 - u_{n+1}u_{n-1}u_m^2. \qquad (*)$$

If $u_1 = 1$, $u_2, u_3 \in \mathbb{Z} \setminus \{0\}$ and $u_4/u_2 \in \mathbb{Z} \setminus \{0\}$, then $u_n \in \mathbb{Z}$ for all n. Why?

- Induction.
- $|u_n|$ counts perfect matchings on certain graphs (Bousquet-Mélu–West, Speyer, others)
- Laurentness of u_n in terms of u_1, u_2, u_3, u_4 (Fomin–Zelevinsky: cluster algebras)
- u_n is the denominator of a point on an elliptic curve.

That Recurrence Relation

$$u_{m+n}u_{m-n}u_1^2 = u_{m+1}u_{m-1}u_n^2 - u_{n+1}u_{n-1}u_m^2. \qquad (*)$$

If $u_1 = 1$, $u_2, u_3 \in \mathbb{Z} \setminus \{0\}$ and $u_4/u_2 \in \mathbb{Z} \setminus \{0\}$, then $u_n \in \mathbb{Z}$ for all n. Why?

- Induction.
- $|u_n|$ counts perfect matchings on certain graphs (Bousquet-Mélu–West, Speyer, others)
- Laurentness of u_n in terms of u_1, u_2, u_3, u_4 (Fomin–Zelevinsky: cluster algebras)
- u_n is the denominator of a point on an elliptic curve.

Example: $y^2 + y = x^3 + x^2 - 2x$

$$u_1 = 1$$
$$u_2 = 1$$
$$u_3 = -3$$
$$u_4 = 11$$
$$u_5 = 38$$
$$u_6 = 249$$
$$u_7 = -2357$$

Example: $y^2 + y = x^3 + x^2 - 2x$

$$u_1 = 1 \qquad P = (0,0)$$

$$u_2 = 1$$

$$u_3 = -3$$

$$u_4 = 11$$

$$u_5 = 38$$

$$u_6 = 249$$

$$u_7 = -2357$$

Example: $y^2 + y = x^3 + x^2 - 2x$

$$\begin{aligned}
u_1 &= 1 & P &= (0,0)\\
u_2 &= 1 & [2]P &= (3,5)\\
u_3 &= -3 & [3]P &= \left(-\frac{11}{9}, \frac{28}{27}\right)\\
u_4 &= 11 & [4]P &= \left(\frac{114}{121}, -\frac{267}{1331}\right)\\
u_5 &= 38 & [5]P &= \left(-\frac{2739}{1444}, -\frac{77033}{54872}\right)\\
u_6 &= 249 & [6]P &= \left(\frac{89566}{62001}, -\frac{31944320}{15438249}\right)\\
u_7 &= -2357 & [7]P &= \left(-\frac{2182983}{5555449}, -\frac{20464084173}{13094193293}\right)
\end{aligned}$$

Example: $y^2 + y = x^3 + x^2 - 2x$

$$
\begin{aligned}
u_1 &= 1 & P &= (0,0) \\
u_2 &= 1 & [2]P &= (3,5) \\
u_3 &= -3 & [3]P &= \left(-\frac{11}{9}, \frac{28}{27}\right) \\
u_4 &= 11 & [4]P &= \left(\frac{114}{121}, -\frac{267}{1331}\right) \\
u_5 &= 38 & [5]P &= \left(-\frac{2739}{1444}, -\frac{77033}{54872}\right) \\
u_6 &= 249 & [6]P &= \left(\frac{89566}{62001}, -\frac{31944320}{15438249}\right) \\
u_7 &= -2357 & [7]P &= \left(-\frac{2182983}{5555449}, -\frac{20464084173}{13094193293}\right)
\end{aligned}
$$

Example: $y^2 + y = x^3 + x^2 - 2x$

$$
\begin{aligned}
u_1 &= 1 & P &= (0,0) \\
u_2 &= 1 & [2]P &= (3,5) \\
u_3 &= -3 & [3]P &= \left(-\frac{11}{3^2}, \frac{28}{3^3}\right) \\
u_4 &= 11 & [4]P &= \left(\frac{114}{11^2}, -\frac{267}{11^3}\right) \\
u_5 &= 38 & [5]P &= \left(-\frac{2739}{38^2}, -\frac{77033}{38^3}\right) \\
u_6 &= 249 & [6]P &= \left(\frac{89566}{249^2}, -\frac{31944320}{249^3}\right) \\
u_7 &= -2357 & [7]P &= \left(-\frac{2182983}{2357^2}, -\frac{20464084173}{2357^3}\right)
\end{aligned}
$$

Division Polynomials

One defines elliptic functions Ψ_n on $E : y^2 = x^3 + Ax + B$ with

$$\begin{cases} \text{zeroes at the } n\text{-torsion points of } E \\ \text{poles supported on } \mathbf{O} \end{cases}$$

Then for

$$P = (x, y) \in E, \qquad [n]P = \left(\frac{\phi_n(P)}{\Psi_n(P)^2}, \frac{\omega_n(P)}{\Psi_n(P)^3} \right).$$

If P is an integral point,

$$\Psi_1 = 1, \qquad \Psi_2 = 2y, \qquad \Psi_3 = 3x^4 + 6Ax^2 + 12Bx - A^2,$$
$$\Psi_4 = 4y(x^6 + 5Ax^4 + 20Bx^3 - 5A^2x^2 - 4ABx - 8B^2 - A^3), \ldots$$

Division Polynomials

One defines elliptic functions Ψ_n on $E : y^2 = x^3 + Ax + B$ with

$$\begin{cases} \text{zeroes at the } n\text{-torsion points of } E \\ \text{poles supported on } \mathbf{O} \end{cases}$$

Then for

$$P = (x, y) \in E, \qquad [n]P = \left(\frac{\phi_n(P)}{\Psi_n(P)^2}, \frac{\omega_n(P)}{\Psi_n(P)^3} \right).$$

If P is an integral point,

$$\Psi_1 = 1, \qquad \Psi_2 = 2y, \qquad \Psi_3 = 3x^4 + 6Ax^2 + 12Bx - A^2,$$

$$\Psi_4 = 4y(x^6 + 5Ax^4 + 20Bx^3 - 5A^2x^2 - 4ABx - 8B^2 - A^3), \ldots$$

Ψ_n satisfy $(*)$.

Division Polynomials

Note:

- $\gcd(\phi_n(P), \Psi_n(P)) = 1$ in $\mathbb{Z}[A, B, x, y]$.
- $\gcd(\phi_n(P), \Psi_n(P))$ is supported on $p \mid \Delta_E$ for $P \in E(\mathbb{Q})$.
- So $\Psi_n(P)$ is almost the denominator of $[n]P$.

Fundamental Correspondence

Theorem (Ward, 1948)

If $u_n : \mathbb{Z} \to \mathbb{Q}$ *satisfies* $(*)$, *and if* $u_1 = 1$, *then for some*

$$E : y^2 = x^3 + Ax + B, \qquad A, B \in \mathbb{Q} \qquad P \in E(\mathbb{Q}),$$

we have

$$u_n = \Psi_n(E, P).$$

Fundamental Correspondence

Theorem (Ward, 1948)

If $u_n : \mathbb{Z} \to \mathbb{Q}$ satisfies $(*)$, and if $u_1 = 1$, then for some

$$E : y^2 = x^3 + Ax + B, \qquad A, B \in \mathbb{Q} \qquad P \in E(\mathbb{Q}),$$

we have

$$u_n = \Psi_n(E, P).$$

Ward's Correspondence:

$$\left\{ \begin{array}{c} \text{curve-point pairs } (E, P) \\ E : y^2 = x^3 + Ax + B, \\ A, B \in \mathbb{Q}, \qquad P \in E(\mathbb{Q}) \\ P \notin E[2] \cup E[3] \end{array} \right\} \longleftrightarrow \left\{ \begin{array}{c} \text{elliptic divisibility} \\ \text{sequences} \\ u_n : \mathbb{Z} \to \mathbb{Q} \\ u_1 = 1, \quad u_2 u_3 \neq 0 \end{array} \right\}$$

Growth Rate

1,
1,
3,
11,
38,
249,
2357,
8767,
496035,
3769372,
299154043,
12064147359,
632926474117,
65604679199921,
6662962874355342,
720710377683595651,
285131375126739646739,
52061747034847247 19135,
36042157766246923788837209,
1414637218637532261361 0002376,
13926071420933252466435774939177,
189071401739889824822835298962 28001,
2356334609742356570409387470315 4629107,
52613843196106605131800510111110767937939,
1910424746438412543757555272420136901439312318,
201143562868610416717760281868105570520101027137,
509582199125499055223626535390012994946103658226 8645,
1619616042354576251961847118847539207230645302109465 2577,
390721759789017211388827166946590849427517620851066278956107,
5986280055034962587902117411856626799800260564768380372311618644,
108902005168517871203290899980149905032338645609229377887214046958803,
401059645553397223298394061792754188929061320344964142960722012585998 3231,
152506207465652277762531462142393791012856442441235840714430103762819736595413,
5286491728223134626400431117234262142530209508718504849234889569684083125892420201,
835397059891704991632636814121353141297683871830623235928141040342038068512341019315446,
10861789122218115292139551508417628820932571356531654998704845795890033629344542872385904645,
13351876087649817486050732736119541016235802111163925747732171131926421411306436158323451057508131,
2042977307842020707295863142858393936350596442010700266977612272386600979584155605002856821221263113151,
666758599738582427580962194986025574476589178060749335314959464037321543378395210027048006648288905711378993,
33316708658847856167209825975212203644033544158093267723708612909985155910861815688221530712645593855290823134401 6,
150866730291138374331025045659005244449458695650548930543174261374298387455590141700233602162964721944201442274446853073,
113760065777234882865006940654654895718896520042025048306493515052149363166271410666963494813413836495437803419621982027412929,
15925316996730732137567755513631456943452993717700763595310711172026756582128668133207380379874720393868837984396576246231406779343 07,
44416310167318880256461428190965193979854149844320579714027500283754273952989380044808517851663079825097686172334231751637837837673262107, ...

$$c^{n^2}$$

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.

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 $\{n^2\alpha\}$ is uniform distributed $\pmod 1$ iff $\alpha \notin \mathbb{Q}$.

Heuristic Argument

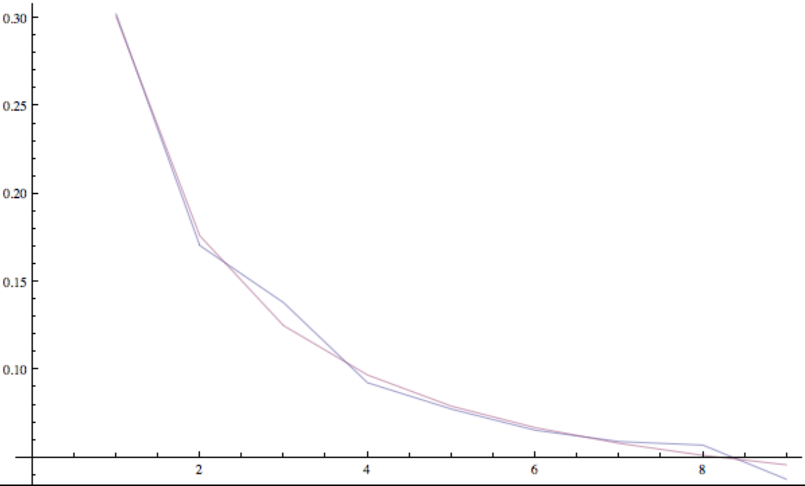
- It's well-known that elliptic divisibility sequences satisfy a growth condition like $u_n \approx c^{n^2}$ where the constant c depends on the arithmetic height of the point P and on the curve E.
- Weyl's theorem tells us that $\{n^2\alpha\}$ is uniform distributed $\pmod 1$ iff $\alpha \notin \mathbb{Q}$.
- 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 $\{n^2\alpha\}$ is uniform distributed $\pmod 1$ 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-$\mathcal{O}$ error terms is delicate, and we need to work out some details.

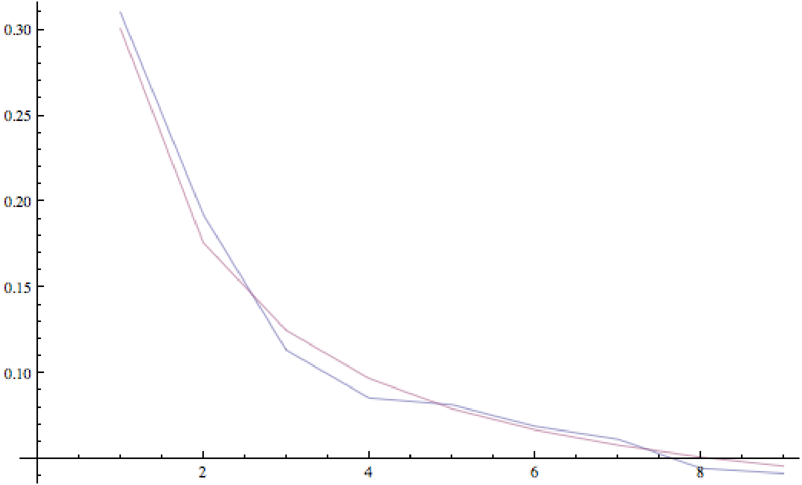
Elliptic Divisibility Sequences are Benford?

Plot of first digit frequencies versus Benford's Law.



Elliptic Divisibility Sequences are Benford?

Plot of first digit frequencies versus Benford's Law.



Elliptic Divisibility Sequences are Benford?

Plot of first digit frequencies versus Benford's Law.