Linear Recurrence Relations with Non-constant Coefficients and Benford's Law

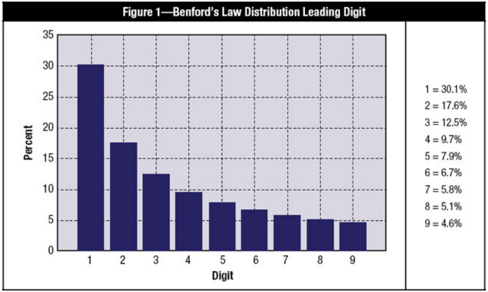
Mengxi Wang, Lily Shao
University of Michigan, Williams College

mengxiw@umich.edu
ls12@williams.edu

Young Mathematics Conference, Ohio State University, August 10, 2018

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?

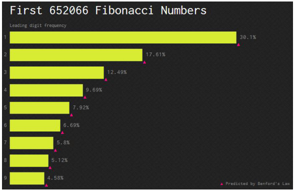
Motivating Question: For a nice data set, such as the Fibonacci numbers, stock prices, street addresses of college employees and students, ..., what percent of the leading digits are 1?



Answer: Benford's law!

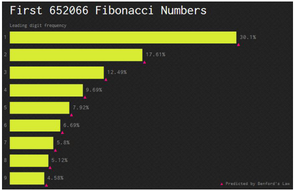
Examples with First Digit Bias

Fibonacci numbers

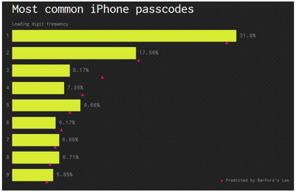


Examples with First Digit Bias

Fibonacci numbers

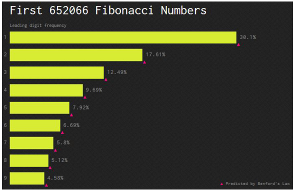


Most common iPhone passcodes

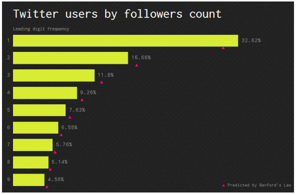


Examples with First Digit Bias

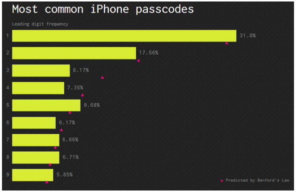
Fibonacci numbers



Twitter users by # followers

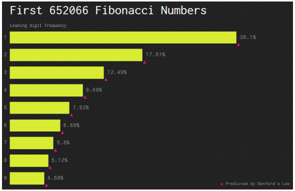


Most common iPhone passcodes

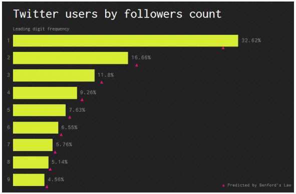


Examples with First Digit Bias

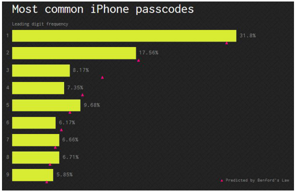
Fibonacci numbers



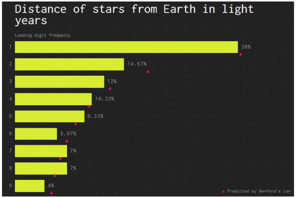
Twitter users by # followers



Most common iPhone passcodes



Distance of stars from Earth



Applications

- Analyzing round-off errors.
- Determining the optimal way to store numbers.
- Detecting tax and image fraud, and data integrity.

An Interesting Question

- From previous works: sequences generated by linear recurrence relations with constant coefficients obey Benford's Law.

An Interesting Question

- From previous works: sequences generated by linear recurrence relations with constant coefficients obey Benford's Law.
- Example: The Fibonacci Sequence 1, 2, 3, 5, 8, 13, 21, 34, ...

An Interesting Question

- From previous works: sequences generated by linear recurrence relations with constant coefficients obey Benford's Law.
- Example: The Fibonacci Sequence 1, 2, 3, 5, 8, 13, 21, 34, ...
- Question: Non-constant coefficients?

Outline

- Benford's Law.
- Linear Recurrence relations (degree 2 and higher degree).
- Multiplicative Recurrence relations.
- Open problems and references.

Equidistribution and Benford's Law

Benford's Law

Benford's Law

A set of numbers is Benford (base b) if the probability of observing a first digit of d is $\log_b\left(1+\frac{1}{d}\right)$.

Benford's Law

Benford's Law

A set of numbers is Benford (base b) if the probability of observing a first digit of d is $\log_b\left(1+\frac{1}{d}\right)$.

Strong Benford

A set of numbers is Strong Benford (base b) if the probability of observing a significand in $[1, s)$ is $log_b(s)$. Then the probability of observing a significand in $[s, s+1)$ is $\log_b\left(1+\frac{1}{s}\right)$.

Equidistribution and Benford's Law

Equidistribution

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

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

Theorem

$\beta \notin \mathbb{Q}$, $n\beta$ is equidistributed 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 $y_i = \log_B x_i$.

Logarithms and Benford's Law

Fundamental Equivalence

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

$$x = S_{10}(x) \cdot 10^k \text{ then}$$
$$\log_{10} x = \log_{10} S_{10}(x) + k = \log_{10} S_{10} x \bmod 1.$$

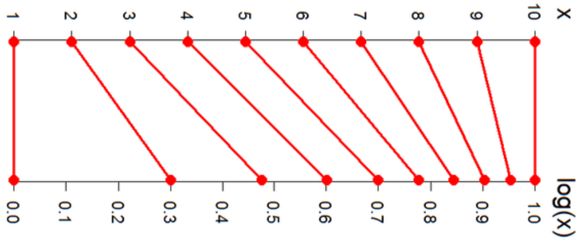
Logarithms and Benford's Law

Fundamental Equivalence

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

$$x = S_{10}(x) \cdot 10^k \text{ then}$$

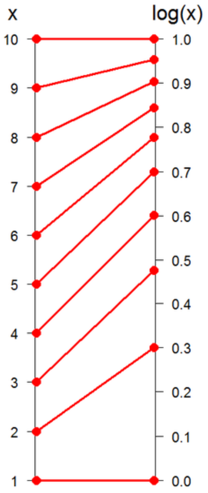
$$\log_{10} x = \log_{10} S_{10}(x) + k = \log_{10} S_{10} x \bmod 1.$$



Logarithms and Benford's Law

$$\begin{aligned}&\mathrm{Prob}(\text{leading digit } d)\\ &= \log_{10}(d+1) - \log_{10}(d)\\ &= \log_{10}\left(\frac{d+1}{d}\right)\\ &= \log_{10}\left(1+\frac{1}{d}\right).\end{aligned}$$

Have Benford's law $\leftrightarrow$ mantissa (fractional part) of logarithms of data are uniformly distributed



Examples

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

Examples

- Remember: $\beta \notin \mathbb{Q}$, $n\beta$ is equidistributed mod 1.
- 2^n is Benford base 10 as $\log_{10} 2 \notin \mathbb{Q}$.

Examples

- Remember: $\beta \notin \mathbb{Q}$, $n\beta$ is equidistributed mod 1.
- 2^n is Benford base 10 as $\log_{10} 2 \notin \mathbb{Q}$.
- Fibonacci numbers are Benford base 10.
  Binet: $a_n = \frac{1}{\sqrt{5}}\left(\frac{1+\sqrt{5}}{2}\right)^n - \frac{1}{\sqrt{5}}\left(\frac{1-\sqrt{5}}{2}\right)^n$.

Useful Theorem

It suffices to analyze the main term of the sequence:

Theorem

If a sequence $\{a_n\}$ *is Benford and* $\lim_{n\to\infty} b_n = a_n$ *then* $\{b_n\}$ *is Benford as well.*

Linear Recurrence Relations

Linear Recurrence Relations of Degree 2

- $a_{n+1} = f(n)a_n + g(n)a_{n-1}$ with non-constant coefficients $f(n)$ and $g(n)$.

Linear Recurrence Relations of Degree 2

- $a_{n+1} = f(n)a_n + g(n)a_{n-1}$ with non-constant coefficients $f(n)$ and $g(n)$.
- Explore conditions on f and g such that the sequence generated obeys Benford's Law for all initial values.

Linear Recurrence Relations of Degree 2

- $a_{n+1} = f(n)a_n + g(n)a_{n-1}$ with non-constant coefficients $f(n)$ and $g(n)$.
- Explore conditions on f and g such that the sequence generated obeys Benford's Law for all initial values.
- First solve the closed form of the sequence (a_n), then analyze its main term.

Linear Recurrence Relations of Degree 2

To solve for the closed form of the sequence:

- Main idea: reduce the degree of recurrence.

Linear Recurrence Relations of Degree 2

To solve for the closed form of the sequence:

- Main idea: reduce the degree of recurrence.
- Define an auxiliary sequence $\{b_n\}_{n=1}^{\infty}$ by $b_n = a_{n+1} - \lambda(n)a_n$ for $n \geq 1$.
  ((a_n) recurrent of degree 2, so (b_n) of degree 1).

  Set $a_{n+1} - \lambda(n)a_n = \mu(n)(a_n - \lambda(n-1)a_{n-1})$ for $n \geq 2$.

- $a_{n+1} = (\lambda(n) + \mu(n))a_n - \mu(n)\lambda(n-1)a_{n-1}$, and compare the coefficients:

$$\begin{aligned} f(n) &= \lambda(n) + \mu(n) \\ g(n) &= -\lambda(n-1)\mu(n). \end{aligned}$$

- $a_{n+1} = (\lambda(n) + \mu(n))a_n - \mu(n)\lambda(n-1)a_{n-1}$, and compare the coefficients:

$$f(n) = \lambda(n) + \mu(n)$$
$$g(n) = -\lambda(n-1)\mu(n).$$

- We show that for any given pair of f and g, such λ and μ always exist.

Linear Recurrence Relations of Degree 2

- Recurrence relations of degree 1:

$$\begin{aligned} a_{n+1} &= \lambda(n)a_n + b_n \\ b_n &= \mu(n)b_{n-1}. \end{aligned}$$

Linear Recurrence Relations of Degree 2

- Recurrence relations of degree 1:

$$\begin{aligned} a_{n+1} &= \lambda(n) a_n + b_n \\ b_n &= \mu(n) b_{n-1}. \end{aligned}$$

- $a_{n+1} = r(n)\left(1 + \sum_{k=3}^{n} \prod_{i=k}^{n} \frac{\lambda(i)}{\mu(i)} + \frac{a_2}{b_1} \prod_{i=2}^{n} \frac{\lambda(i)}{\mu(i)}\right)$, where $r(n) := b_1 \prod_{i=2}^{n} \mu(i)$.

Linear Recurrence Relations of Degree 2

- Asymptotic analysis: show the main term dominates for suitable choices of μ and λ.

Linear Recurrence Relations of Degree 2

- Asymptotic analysis: show the main term dominates for suitable choices of μ and λ.
- Let $\frac{\lambda(n)}{\mu(n)}$ be a non-increasing function and $\lim_{n\to\infty} \frac{\lambda(n)}{\mu(n)} = 0$, then $\lim_{n\to\infty}(a_{n+1} - r(n)) = 0$.

Linear Recurrence Relations of Degree 2

- Asymptotic analysis: show the main term dominates for suitable choices of μ and λ.
- Let $\frac{\lambda(n)}{\mu(n)}$ be a non-increasing function and $\lim_{n\to\infty} \frac{\lambda(n)}{\mu(n)} = 0$, then $\lim_{n\to\infty}(a_{n+1} - r(n)) = 0$.
- $\lim_{n\to\infty} \frac{\lambda(n)}{\mu(n)} = 0$ implies $\lim_{n\to\infty} \frac{g(n)}{f(n)^2} = 0$.

Benford-ness of the Main Term

- Main term of a_{n+1} is $r(n) = b_1 \prod_{i=2}^{n} \mu(i)$.

Benford-ness of the Main Term

- Main term of a_{n+1} is $r(n) = b_1 \prod_{i=2}^{n} \mu(i)$.
- Since (strong) Benford-ness is preserved under translation and dilation we let $r(n) = \prod_{i=1}^{n} \mu(i)$ for simplicity.

Examples when f and g are functions

- If $\mu(k) = k$, then $r(n) = n!$.

Examples when f and g are functions

- If $\mu(k) = k$, then $r(n) = n!$.
- If $\mu(k) = k^\alpha$ where $\alpha \in \mathbb{R}$, then $r(n) = (n!)^\alpha$.

Examples when f and g are functions

- If $\mu(k) = k$, then $r(n) = n!$.
- If $\mu(k) = k^\alpha$ where $\alpha \in \mathbb{R}$, then $r(n) = (n!)^\alpha$.
- If $\mu(k) = \exp(\alpha h(k))$ where α is irrational and $h(k)$ is a monic polynomial, then $\log r(n) = \alpha \sum_{k=1}^{n} h(k)$.

Lemma

The sequence $\{\alpha p(n)\}$ is equidistributed mod 1 if $\alpha \notin \mathbb{Q}$ and $p(n)$ a monic polynomial.

Examples when f and g are random variables

- Take $\mu(n) \sim h(n)U_n$ where the U_n's are independent uniform distributions on $[0, 1]$, and $h(n)$ is a deterministic function in n such that $\prod_{i=1}^{n} h(i)$ is Benford.

  Then $r(n) = \prod_{i=1}^{n} h(i) \prod_{i=1}^{n} U_i$ is Benford.

Examples when f and g are random variables

- Take $\mu(n) \sim h(n)U_n$ where the U_n's are independent uniform distributions on $[0, 1]$, and $h(n)$ is a deterministic function in n such that $\prod_{i=1}^{n} h(i)$ is Benford.

  Then $r(n) = \prod_{i=1}^{n} h(i) \prod_{i=1}^{n} U_i$ is Benford.

- Take $\mu(n) \sim \exp(U_n)$ where the U_n's are i.i.d. random variables. Then take logarithm and sum up $\log(\mu(n))$. Apply Central Limit Theorem and get a Gaussian distribution with increasing variance.

Linear Recurrences of Higher Degree

- Use recurrence relation of degree 3 as an example.
  Similar main idea: reduce the degree.

Linear Recurrences of Higher Degree

- Use recurrence relation of degree 3 as an example. Similar main idea: reduce the degree.
- Define the sequence $\{a_n\}_{n=1}^{\infty}$ by $a_{n+1} \ = \ f_1(n)a_n + f_2(n)a_{n-1} + f_3(n)a_{n-2}$.

Linear Recurrences of Higher Degree

- Use recurrence relation of degree 3 as an example. Similar main idea: reduce the degree.
- Define the sequence $\{a_n\}_{n=1}^{\infty}$ by $a_{n+1} \ = \ f_1(n)a_n + f_2(n)a_{n-1} + f_3(n)a_{n-2}$.
- Define an auxiliary sequence $(b_n)_{n=1}^{\infty}$ by $b_n \ = \ a_{n+1} - \lambda(n)a_n$. Then (b_n) is degree 2.

Appendix
Multiplicative Recurrence Relations

Generalization to Multiplicative Recurrence Relations

- Define sequence $(A_n)_{n=1}^{\infty}$ by the recurrence relation $A_{n+1} = A_n^{f(n)} A_{n-1}^{g(n)}$ with initial values A_1, A_2.

Generalization to Multiplicative Recurrence Relations

- Define sequence $(A_n)_{n=1}^{\infty}$ by the recurrence relation $A_{n+1} = A_n^{f(n)} A_{n-1}^{g(n)}$ with initial values A_1, A_2.

- Then the closed form of A_n is $A_n = A_2^{x_n} A_1^{y_n}$, where the exponents (x_n) and (y_n) satisfy the linear recurrence relations

$$\begin{aligned} x_{n+1} &= f(n)x_n + g(n)x_{n-1} \\ y_{n+1} &= f(n)y_n + g(n)y_{n-1} \end{aligned}$$

with initial values

$$\begin{aligned} x_1 &= 0, x_2 = 1, \\ y_1 &= 1, y_2 = 0. \end{aligned}$$

Generalization to Multiplicative Recurrence Relations

- Again, solve for (x_n) and (y_n) with auxiliary functions $\lambda(n)$ and $\mu(n)$.

Generalization to Multiplicative Recurrence Relations

- Again, solve for (x_n) and (y_n) with auxiliary functions $\lambda(n)$ and $\mu(n)$.

- Let $\lambda(n)$ and $\mu(n)$ satisfy $\lim_{n\to\infty} \frac{\lambda(n)}{\mu(n)} = 0$.

Generalization to Multiplicative Recurrence Relations

- Take the main terms:

$$x_{n+1} \rightarrow (x_2 - \lambda(1)x_1) \prod_{i=2}^{n} \mu(i) = \prod_{i=2}^{n} \mu(i),$$

$$y_{n+1} \rightarrow (y_2 - \lambda(1)y_1) \prod_{i=2}^{n} \mu(i) = -\lambda(1) \prod_{i=2}^{n} \mu(i).$$

Generalization to Multiplicative Recurrence Relations

- Take the main terms:

$$x_{n+1} \;\to\; (x_2 - \lambda(1)x_1)\prod_{i=2}^{n}\mu(i) \;=\; \prod_{i=2}^{n}\mu(i),$$
$$y_{n+1} \;\to\; (y_2 - \lambda(1)y_1)\prod_{i=2}^{n}\mu(i) \;=\; -\lambda(1)\prod_{i=2}^{n}\mu(i).$$

- By the closed form $A_n = A_2^{x_n}A_1^{y_n}$,

$$\begin{aligned}\log(A_{n+1}) &= x_n\log(A_2) + y_n\log(A_1)\\ &\to \Big(\prod_{i=2}^{n}\mu(i)\Big)(\log(A_2) - \lambda(1)\log(A_1))\end{aligned}$$

as $n \to \infty$.

Benford-ness of the main term

- (A_n) is a Benford sequence if the main term of $\log(A_n)$ is equidistributed mod 1.

Benford-ness of the main term

- (A_n) is a Benford sequence if the main term of $\log(A_n)$ is equidistributed mod 1.
- We choose μ and the initial values such that $(\log(A_2) - \lambda(1)\log(A_1))\prod_{i=2}^{n}\mu(i)$ is equidistributed mod 1.

Examples

Remeber: The sequence $\{\alpha p(n)\}$ is equidistributed mod 1 if $\alpha \notin \mathbb{Q}$ and $p(n)$ a monic polynomial.

Examples

Remeber: The sequence $\{\alpha p(n)\}$ is equidistributed mod 1 if $\alpha \notin \mathbb{Q}$ and $p(n)$ a monic polynomial.

Our construction:

- $\frac{\log(A_2) - \lambda(1)\log(A_1)}{\mu(1)} =: \alpha \notin \mathbb{Q}$,
- $\mu(i) = \frac{p(i)}{p(i-1)}$ where $p(n)$ is a non-vanishing monic polynomial.

Open Questions and References

Open Questions

1. We mainly consider the case when $\frac{\lambda(n)}{\mu(n)} \to 0$ as $n \to \infty$. What about when $\frac{\lambda(n)}{\mu(n)} \to \infty$? In this case, there is no simple dominating term.

Open Questions

1. We mainly consider the case when $\frac{\lambda(n)}{\mu(n)} \to 0$ as $n \to \infty$. What about when $\frac{\lambda(n)}{\mu(n)} \to \infty$? In this case, there is no simple dominating term.

2. Applications of recurrence relations when $f(n)$ and $g(n)$ are random variables?

Acknowledgement

The presenters are partially supported by NSF grants, the University of Michigan, and Williams College. We are grateful to our excellent advisor Professor Steven J. Miller for valuable guidance, and thanks the other contributors Madeleine Farris and Noah Luntzlara for help.

References

- P. Diaconis, *The Distribution of Leading Digits and Uniform Distribution Mod 1*, The Annals of Probability **5** (1977), No. 1, 72–81.
- D. Jang, J. U. Kang, A. Kruckman, J. Kudo, and S. J. Miller, *Chains of Distributions, Hierarchical Bayesian Models and Benford's Law*, preprint, 2008. https://arxiv.org/pdf/0805.4226.pdf.
- A. Kontorovich and S. J. Miller, *Benford's Law, Values of L-functions and the* $3x + 1$ *Problem*, Acta Arithmetica, **120** (2005), 269–297.
- P. K. Romano and H. Mclaughlin, *On Non-Linear Recursive Sequences and Benford's Law*, Fibonacci Quarterly, **49** (2011), 134–138.
- T. Tao, *254B, Notes 1: Equidistribution of polynomial sequences in tori* 254A, 2010, https://terrytao.wordpress.com/2010/03/28/