Benford's Law: From Logarithms to Dynamical Systems

William Hao

University of Texas at Austin

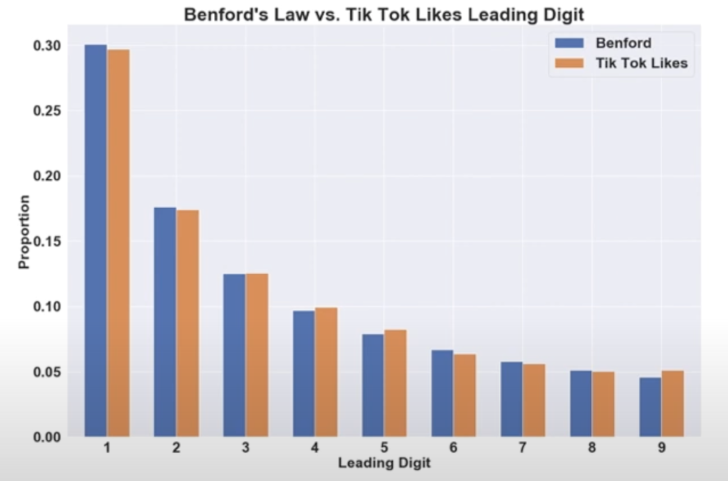
william.hao@utexas.edu

April 24, 2025

Motivation

- Given a collection of positive numbers from a set of random data
- We'd expect that the **leading digits** are distributed evenly...
- **Benford's Law**:
  - Leading digit 1 ~30% of the time
  - Leading digit 9 only ~4.6% of the time
- This talk will connect Benford's Law to logarithms, randomness, and dynamical systems

Motivation



Data from: https://www.youtube.com/watch?v=42fGFDNs-0A

Historical Origins: Early Observations

- **First Observed by Simon Newcomb (1881):**
  - Earlier pages of log tables were significantly more worn out than later pages.
  - He made a rule similar to Benford's Law
- **Rediscovered by Frank Benford (1938):**
  - Observed the same phenomenon
  - **Tested the law** extensively. Tests included data like:
    - Surface areas of 335 rivers
    - Sizes of 3,259 U.S. populations
    - 104 physical constants
    - 1,800 molecular weights
    - Street addresses from *American Men of Science*
  - He popularized the law.

What is Benford's Law?

- **Benford's Law:** In naturally occurring, or extensive numerical datasets, the probability $P(d)$ that the **leading digit** is d ($d \in \{1, 2, \dots, 9\}$) is:

$$\boxed{P(d) = \log_{10}\left(1 + \frac{1}{d}\right)}$$

- **Generalization to any base b:** The probability that the first digit is $d \in \{1, 2, \dots, b-1\}$ is:

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

- **Base 10 Examples:**
  - $P(1) = \log_{10}(2) \approx 0.301\ (30.1\%)$
  - $P(2) = \log_{10}(1.5) \approx 0.176\ (17.6\%)$
  - ...
  - $P(9) = \log_{10}(10/9) \approx 0.046\ (4.6\%)$

What digits appear in 2^n? (Israel Gelfand's Question)

- Seemingly unrelated application:
- Consider the sequence 2^n: $2, 4, 8, 16, 32, 64, 128, 256, 512, 1024, \ldots$
- What are the leading digits of each term?

- **Claim:** The distribution of these leading digits follows Benford's Law.
- To prove, we need a way to analyze the distribution of first digits

Proof Idea 1: Logarithms Determine First Digits

Relating First Digit to Logarithms

- Any power $2^n = d \times 10^k$; $1 \le d < 10$, $k = \lfloor \log_{10}(2^n) \rfloor$.
  - Ex. $2^{14} = 16384 = 1.6384 \times 10^4$
- First digit, D, is always equal to $\lfloor d \rfloor$ (floor function)
- Take $\log_{10}$:

$$\log_{10}(2^n) = \log_{10}(d \times 10^k)$$

$$n \log_{10}(2) = \log_{10}(d) + k$$

- Let $X = n \log_{10}(2)$. Note that $k = \lfloor X \rfloor$

$$X - k = \log_{10}(d)$$

$$X - \lfloor X \rfloor = \log_{10}(d)$$

$$\boxed{frac\{X\} = frac\{n \log_{10}(2)\} = \log_{10}(d)}$$

- Mathematical inequality for floor function:

$$\log_{10}(D) \leq \log_{10}(d) < \log_{10}(D+1)$$

$$\boxed{\log_{10}(D) \leq frac\{n\log_{10}(2)\} < \log_{10}(D+1)}$$

- **Key Insight:** Distribution of the first digit depends **entirely** on how $frac\{n\log_{10}(2)\}$ is distributed within $[0, 1)$

Proof Idea 2: The Circle Rotation Analogy

Connecting to Dynamical Systems: Irrational Circle Rotation

- Consider the sequence of fractional parts: $x_n = frac\{n\alpha\}$; $\alpha = \log_{10}(2)$.
- This sequence is generated by the **dynamical system** of rotation on a circle:
    - *Space:* The circle $\mathbb{T} = \mathbb{R}/\mathbb{Z}$ (or interval $[0, 1)$)
    - *Transformation:* $T(x) = \{x + \alpha\}$ (Rotation by α)

Proof Idea 3: Input from Dynamical Systems Theory

Applying Weyl's Theorem

- **Key Fact:** $\alpha = \log_{10}(2)$ is *irrational*.
- **Weyl's Equidistribution Theorem:** For irrational α, the orbit $\{n\alpha\}$ is **uniformly distributed** in $[0, 1)$.

- **Conclusion:** Proportion of n where $x_n = frac\{n\alpha\}$ falls into $[\log_{10}(D), \log_{10}(D+1))$ is just length.

$$P(D) = \text{Length} = \log_{10}(D+1) - \log_{10}(D) = \log_{10}\left(1 + \frac{1}{D}\right)$$

- This is exactly Benford's Law!

Scale Invariance

- Key property arising from Benford's Law: **scale invariance**.
- If dataset $\{x_i\}$ follows Benford's law, then scaled dataset $\{c \cdot x_i\}$ (for $c > 0$) should also follow.

- **Analogy:** Imagine converting units of a data set, like feet to meters, dollars to euros, miles to kilometers. The distribution of first digits shouldn't change. (And it doesn't!)

Scale Invariance

Mathematical expression:

- If X is a r.v., Benford's Law holds if $frac\{\log_{10} X\}$ is uniform on $[0, 1)$
- Scale by c: $\log_{10}(cX) = \log_{10} c + \log_{10} X$
- Take fractions: $frac\{\log_{10}(cX)\} = frac\{\log_{10} c\} + frac\{\log_{10} X\}$.

- **What this means:** Adding a constant $\log_{10} c$ and taking the fractional part corresponds to *rotation* on a circle.
- If original distribution $frac\{\log_{10} X\}$ was uniform, the rotated distribution remains uniform.

Connections to Other Systems

- Benford's Law appears naturally in many systems with growth, multiplicative, or chaotic behavior
  - 3^n
  - Fibonacci sequence (ratio of terms approaches ϕ)
  - Logistic growth for population modeling

- Textbook: Nillsen discusses equivalence/connection of a *Kronecker system* (irrational rotations on a circle) and a *Benford system* (related to multiplication by 10 modulo 1) (Section 3.18)

Randomness: Benford vs. Normal Numbers

- **Normal Numbers (Borel):** Numbers that exhibit maximum digit randomness.
  - Normal Number: All digits appear with equal asymptotic frequency ($1/b$) in a *single* number's base-b expansion.
  - *Borel's Theorem (1909):* Almost all real numbers (Lebesgue measure) are normal to every base $b \geq 2$
- **Distinction:**
  - Normality $\rightarrow$ Distribution of **all** digits in a **single** number.
  - Benford $\rightarrow$ Distribution of the **first** digit across a **set** of numbers.
- **Analogy: "Logarithmic" Normality**
  - Benford's Law: the fractional part of numbers' logarithms are uniformly distributed on $[0, 1)$
  - *Uniformity of logarithms* parallels the *uniformity of digits* in Borel's NNT

Applications of Benford's Law

- **Fraud Detection:** One of the major applications
  - Financial statements, tax returns, credit card transactions, election results
  - Human-fabricated data often doesn't follow Benford's Law
- **Validating Scientific Data:** Checking if experimental results or physical constants conform to Benford's Law
- **Natural Phenomena:** Population sizes, river lengths, stock market data, earthquake magnitudes
  - **Important**: Benford's Law applies more strongly to natural data sets when the data **spans several orders of magnitude**

Takeaways and Conclusion

- Benford's Law is an interesting, counter-intuitive statistical law that seems to pop up all the time. Governs what leading digit looks like in a variety of datasets
- Fundamentally connected to properties of logarithms and scale invariance
- Deeply connected to dynamical systems (Weyl's Theorem - irrational increment ”rotations” are uniformly spread on a circle)
- Broad connections to randomness, recurrence, and chaotic behavior in mathematical systems

References

- Rodney V. Nillsen (2010)
  Randomness and Recurrence in Dynamical Systems
  Section 3.14-3.19
- Minding the Data (2020)
  Benford's Law in Real Life | Finding Real World Examples
  YouTube Video
- Wikipedia
  Benford's law
  Wikipedia entry

Questions?