Benford's Law

Discrete Mathematics Seminar 2023

Jesse Campbell

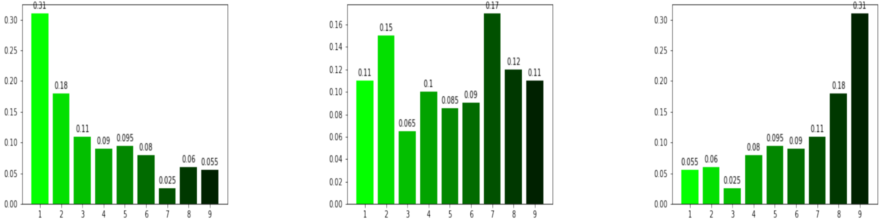
Duke Kunshan University

May 12, 2023

Outline

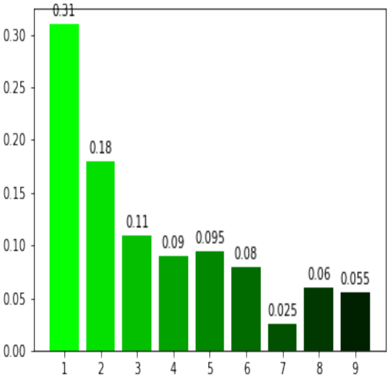
Open questions will be in orange!

Country Populations



Frequency of leading digits of population by country (2020).

Country Populations



Frequency of leading digits of population by country (2020).

Fibonacci Numbers

The Fibonacci sequence is given by,

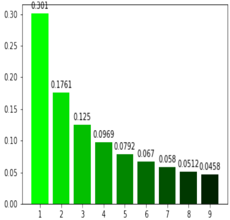
$$F(n) = F(n-1) + F(n-2) \text{ where } F(0) = F(1) = 1$$

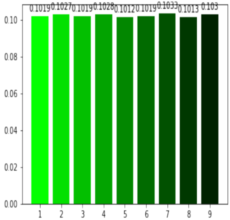
Fibonacci Numbers

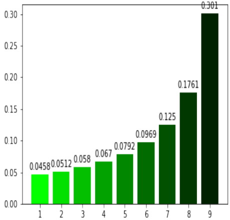
The Fibonacci sequence is given by,

$$F(n) = F(n-1) + F(n-2) \text{ where } F(0) = F(1) = 1$$

Which graph is correct?

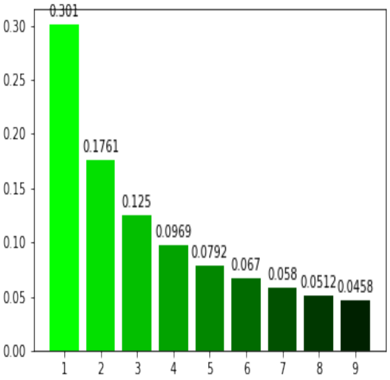






Frequency of leading digits of first 50,000 Fibonacci numbers.

Fibonacci Numbers



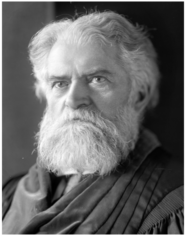
Frequency of leading digits of first 50,000 Fibonacci numbers.

Benford's Law Visualization

Digit	Probability $P(d)$	Relative Size of $P(d)$
1	0.301	
2	0.176	
3	0.125	
4	0.097	
5	0.079	
6	0.067	
7	0.058	
8	0.051	
9	0.046	

History of Benford's Law

That the ten digits do not occur with equal frequency must be evident to any one making much use of logarithmic tables, and noticing how much faster the first pages wear out than the last ones. —*Simon Newcomb (1881)*



Simon Newcomb, 1905

History of Benford's Law

- Benford's Law was rediscovered by physicist Frank Benford in 1938.
- Compiled over 20 tables containing over 20,000 data points supporting the law

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	—

Frank Benford's original data supporting Benford's Law (1938)

Notation

Let $D_n(x)$ be the n^{th} significant decimal digit

Notation

Let $D_n(x)$ be the n^{th} significant decimal digit

- $D_1(\pi) = 3$, $D_2(\pi) = 1$, $D_3(\pi) = 4$
- $D_n(300) = D_n(3) = D_n(0.003)$

What is Benford's Law

Benford's Law (1st digit)

$$Prob(D_1 = d) = \log_{10}(1 + \frac{1}{d})$$

What is Benford's Law

Benford's Law (1st digit)

$$Prob(D_1 = d) = \log_{10}(1 + \frac{1}{d})$$

And to make it a bit more general...

What is Benford's Law

Benford's Law (1st digit)

$$Prob(D_1 = d) = \log_{10}(1 + \tfrac{1}{d})$$

And to make it a bit more general...

Benford's Law

$$Prob(D_1 = d_1, D_2 = d_2, ..., D_m = d_m) = \log_{10}(1 + (\textstyle\sum_{j=1}^{m} 10^{m-j} d_j)^{-1})$$

Example

Pick any number from a distribution that follows Benford's Law.

What's the probability that the first five digits are the same as π?

Example

Pick any number from a distribution that follows Benford's Law.

What's the probability that the first five digits are the same as π?

$\text{Prob}(\text{D}_1 = 3, D_2 = 1, D_3 = 4, D_4 = 1, D_5 = 5) = \log_{10}(1 + \frac{1}{31415})$

$= \log_{10}(\frac{31416}{31415}) \approx 0.0000138$

Suprising Result?

$$\text{Prob}(D_2 = 1) = \sum_{j=1}^{9} \log_{10}(1 + \frac{1}{10j+1}) = \log_{10}(\frac{6029312}{4638501}) \approx 0.1138$$

Suprising Result?

$$\text{Prob}(D_2 = 1) = \sum_{j=1}^{9} \log_{10}(1 + \frac{1}{10j+1}) = \log_{10}(\frac{6029312}{4638501}) \approx 0.1138$$

But...

Suprising Result?

Prob$(\mathsf{D}_2 = 1) = \sum_{j=1}^{9} \log_{10}(1 + \frac{1}{10j+1}) = \log_{10}(\frac{6029312}{4638501}) \approx 0.1138$

But...

Prob$(\mathsf{D}_2 = 1 | D_1 = 1) = \frac{\log_{10}(1+\frac{1}{11})}{\log_{10}(2)} = \frac{log_{10}(12)-log_{10}(11)}{log_{10}(2)} = 0.1255$

Suprising Result?

$$\text{Prob}(\text{D}_2 = 1) = \sum_{j=1}^{9} \log_{10}(1 + \frac{1}{10j+1}) = \log_{10}(\frac{6029312}{4638501}) \approx 0.1138$$

But...

$$\text{Prob}(\text{D}_2 = 1|D_1 = 1) = \frac{\log_{10}(1+\frac{1}{11})}{\log_{10}(2)} = \frac{log_{10}(12)-log_{10}(11)}{log_{10}(2)} = 0.1255$$

Conclusion: Significant digits are **dependent**.

Significand

Another useful concept when taking about Benford's Law is the **significand**, also called the *mantissa*.

The significand of a number, call it $S(x)$, is its coeffcient when expressed in "scientific" (floating-point) notation.

Significand

Another useful concept when taking about Benford's Law is the **significand**, also called the *mantissa*.

The significand of a number, call it $S(x)$, is its coeffcient when expressed in "scientific" (floating-point) notation.

$P = 6.626 \times 10^{-34}$ (Plank Constant)
$S(P) = 6.626$

Significand Function

Explicity, the base-10 significand function $S : \mathbb{R} \to [1, 10)$ is given by,

Significand Function

$$S(x) = 10^{\log|x| - \lfloor \log|x| \rfloor}, \quad S(0) := 0$$

Significand Function

Explicity, the base-10 significand function $S : \mathbb{R} \to [1, 10)$ is given by,

Significand Function

$$S(x) = 10^{\log|x| - \lfloor \log|x| \rfloor}, \quad S(0) := 0$$

- $S(\sqrt{2}) = S(-\sqrt{2}) = S(10\sqrt{2}) = \sqrt{2} = 1.414...$
- $S(\pi^{-1}) = S(10\pi^{-1}) = 10\pi^{-1} = 3.183$

Significand Function

Explicity, the base-10 significand function $S : \mathbb{R} \to [1, 10)$ is given by,

Significand Function

$$S(x) = 10^{\log|x| - \lfloor \log|x| \rfloor}, \quad S(0) := 0$$

- $S(\sqrt{2}) = S(-\sqrt{2}) = S(10\sqrt{2}) = \sqrt{2} = 1.414...$
- $S(\pi^{-1}) = S(10\pi^{-1}) = 10\pi^{-1} = 3.183$

Using the significand we can state Benford's Law in a new (and super concise) way:

Benford's Law

$$Prob(S \leq t) = \log(t),\ t \in [1, 10)$$

σ-Algebras

Informal statements about Benford's Law all involve *probabilities*, therefore for mathematical precision it is necessary to reformulate these statements in the setting of rigorous probability theory.

σ-Algebras

Informal statements about Benford's Law all involve *probabilities*, therefore for mathematical precision it is necessary to reformulate these statements in the setting of rigorous probability theory.

A probability space $(\Omega, \mathcal{F}, \mathbb{P})$ consists of three objects: an outcome space Ω, a σ-algebra $\mathcal{F}$, and a probability measure $\mathbb{P}$.

σ-Algebras

Informal statements about Benford's Law all involve *probabilities*, therefore for mathematical precision it is necessary to reformulate these statements in the setting of rigorous probability theory.

A probability space $(\Omega, \mathcal{F}, \mathbb{P})$ consists of three objects: an outcome space Ω, a σ-algebra $\mathcal{F}$, and a probability measure $\mathbb{P}$.

A σ-algebra $\mathcal{F}$ on Ω is simply a subset of Ω such that $\emptyset \in \mathcal{F}$ and $\mathcal{F}$ is closed under complements and countable unions.

σ-Algebras

Informal statements about Benford's Law all involve *probabilities*, therefore for mathematical precision it is necessary to reformulate these statements in the setting of rigorous probability theory.

A probability space $(\Omega, \mathcal{F}, \mathbb{P})$ consists of three objects: an outcome space Ω, a σ-algebra $\mathcal{F}$, and a probability measure $\mathbb{P}$.

A σ-algebra $\mathcal{F}$ on Ω is simply a subset of Ω such that $\emptyset \in \mathcal{F}$ and $\mathcal{F}$ is closed under complements and countable unions.

Example: The *power set* of Ω, which is the set containing all possible subsets of Ω, is the largest possible σ-algebra on Ω.

σ-Algebra Generated by a Function

For a subset C on $\mathbb{R}$ and a function $f : \Omega \to \mathbb{R}$, the *pre-image* of C under f is defined as:

σ-Algebra Generated by a Function

For a subset C on $\mathbb{R}$ and a function $f : \Omega \to \mathbb{R}$, the *pre-image* of C under f is defined as:

$$f^{-1}(C) = \{\omega \in \Omega : f(\omega) \in C\}$$

σ-Algebra Generated by a Function

For a subset C on $\mathbb{R}$ and a function $f : \Omega \to \mathbb{R}$, the *pre-image* of C under f is defined as:

$$f^{-1}(C) = \{\omega \in \Omega : f(\omega) \in C\}$$

The *σ-algebra generated by f* is:

σ-Algebra Generated by a Function

For a subset C on $\mathbb{R}$ and a function $f : \Omega \to \mathbb{R}$, the *pre-image* of C under f is defined as:

$$f^{-1}(C) = \{\omega \in \Omega : f(\omega) \in C\}$$

The *σ-algebra generated by f* is:

$$\sigma(f) = \{f^{-1}(J) : J \subset \mathbb{R}, \text{J an interval}\}$$

σ-Algebra Generated by a Function

For a subset C on $\mathbb{R}$ and a function $f : \Omega \to \mathbb{R}$, the *pre-image* of C under f is defined as:

$$f^{-1}(C) = \{\omega \in \Omega : f(\omega) \in C\}$$

The *σ-algebra generated by f* is:

$$\sigma(f) = \{f^{-1}(J) : J \subset \mathbb{R}, \text{J an interval}\}$$

$\sigma(f)$ is the *smallest* σ-algebra on Ω that contains all sets of the form $\{\omega \in \Omega : a \leq f(\omega) \leq b\}$ for every $a, b \in \mathbb{R}$.

Significand σ-Algebra

We define the significand σ-algebra $\mathcal{S}$ to be the σ-algebra generated by the significand function S, i.e. $\mathcal{S} = \mathbb{R}^+ \cap \sigma(S)$

Significand σ-Algebra

We define the significand σ-algebra $\mathcal{S}$ to be the σ-algebra generated by the significand function S, i.e. $\mathcal{S} = \mathbb{R}^+ \cap \sigma(S)$

Importance: For every event $A \in \mathcal{S}$ and every $x > 0$, knowing $S(x)$ is enough to determine whether $x \in A$ or $x \notin A$.

Significand σ-Algebra

We define the significand σ-algebra $\mathcal{S}$ to be the σ-algebra generated by the significand function S, i.e. $\mathcal{S} = \mathbb{R}^+ \cap \sigma(S)$

Importance: For every event $A \in \mathcal{S}$ and every $x > 0$, knowing $S(x)$ is enough to determine whether $x \in A$ or $x \notin A$.

Corollary: $\mathcal{S}$ is the family of events $A \in \mathbb{R}^+$ that can be described completely in-terms of their significands. For example,

Significand σ-Algebra

We define the significand σ-algebra $\mathcal{S}$ to be the σ-algebra generated by the significand function S, i.e. $\mathcal{S} = \mathbb{R}^+ \cap \sigma(S)$

Importance: For every event $A \in \mathcal{S}$ and every $x > 0$, knowing $S(x)$ is enough to determine whether $x \in A$ or $x \notin A$.

Corollary: $\mathcal{S}$ is the family of events $A \in \mathbb{R}^+$ that can be described completely in-terms of their significands. For example,

- $A_1 = \{x > 0 : D_1(x) = 1, D_3(x) \neq 7\}$

Significand σ-Algebra

We define the significand σ-algebra $\mathcal{S}$ to be the σ-algebra generated by the significand function S, i.e. $\mathcal{S} = \mathbb{R}^+ \cap \sigma(S)$

Importance: For every event $A \in \mathcal{S}$ and every $x > 0$, knowing $S(x)$ is enough to determine whether $x \in A$ or $x \notin A$.

Corollary: $\mathcal{S}$ is the family of events $A \in \mathbb{R}^+$ that can be described completely in-terms of their significands. For example,

- $A_1 = \{x > 0 : D_1(x) = 1, D_3(x) \neq 7\}$
- $A_2 = \{x > 0 : D_m(x) \in \{5, 6\}$ for all $m \in \mathbb{N}\}$

Significand σ-Algebra

We define the significand σ-algebra $\mathcal{S}$ to be the σ-algebra generated by the significand function S, i.e. $\mathcal{S} = \mathbb{R}^+ \cap \sigma(S)$

Importance: For every event $A \in \mathcal{S}$ and every $x > 0$, knowing $S(x)$ is enough to determine whether $x \in A$ or $x \notin A$.

Corollary: $\mathcal{S}$ is the family of events $A \in \mathbb{R}^+$ that can be described completely in-terms of their significands. For example,

- $A_1 = \{x > 0 : D_1(x) = 1, D_3(x) \neq 7\}$
- $A_2 = \{x > 0 : D_m(x) \in \{5, 6\}$ for all $m \in \mathbb{N}\}$
- $A_3 = \{x > 0 : S(x) \in \mathbb{Q}\}$

Significand σ-Algebra

We define the significand σ-algebra $\mathcal{S}$ to be the σ-algebra generated by the significand function S, i.e. $\mathcal{S} = \mathbb{R}^+ \cap \sigma(S)$

Importance: For every event $A \in \mathcal{S}$ and every $x > 0$, knowing $S(x)$ is enough to determine whether $x \in A$ or $x \notin A$.

Corollary: $\mathcal{S}$ is the family of events $A \in \mathbb{R}^+$ that can be described completely in-terms of their significands. For example,

- $A_1 = \{x > 0 : D_1(x) = 1, D_3(x) \neq 7\}$
- $A_2 = \{x > 0 : D_m(x) \in \{5, 6\}$ for all $m \in \mathbb{N}\}$
- $A_3 = \{x > 0 : S(x) \in \mathbb{Q}\}$

Whereas, for example, the interval $[1, 2]$ does not belong to $\mathcal{S}$.

Questions you may have...

- How do we derive the distribution function for Benford's Law?

Questions you may have...

- How do we derive the distribution function for Benford's Law?
- Which distributions of numbers follow Benford's Law?

Questions you may have...

- **How do we derive the distribution function for Benford's Law?**
- Which distributions of numbers follow Benford's Law?

Mathematical Derivation

Let $\Omega = \mathbb{R}^+$ be our sample space.

Mathematical Derivation

Let $\Omega = \mathbb{R}^+$ be our sample space.

We want to find a probability measure $P(d)$ on Ω, where $d \in [0..9]$ is the leading digit of a number in Ω.

Mathematical Derivation

Let $\Omega = \mathbb{R}^+$ be our sample space.

We want to find a probability measure $P(d)$ on Ω, where $d \in [0..9]$ is the leading digit of a number in Ω.

We can begin by defining the set of numbers in Ω with leading digit d. We can represent this in set notation as

Mathematical Derivation

Let $\Omega = \mathbb{R}^+$ be our sample space.

We want to find a probability measure $P(d)$ on Ω, where $d \in [0..9]$ is the leading digit of a number in Ω.

We can begin by defining the set of numbers in Ω with leading digit d. We can represent this in set notation as

$$S(d) = \bigcup_{n=-\infty}^{\infty} [d \cdot 10^n, (d+1) \cdot 10^n)$$

Mathematical Derivation

Let $\Omega = \mathbb{R}^+$ be our sample space.

We want to find a probability measure $P(d)$ on Ω, where $d \in [0..9]$ is the leading digit of a number in Ω.

We can begin by defining the set of numbers in Ω with leading digit d. We can represent this in set notation as

$$S(d) = \bigcup_{n=-\infty}^{\infty} [d \cdot 10^n, (d+1) \cdot 10^n)$$

Let $f(x)$ be a continuous density function on Ω, then it immediately follows that

Mathematical Derivation

Let $\Omega = \mathbb{R}^+$ be our sample space.

We want to find a probability measure $P(d)$ on Ω, where $d \in [0..9]$ is the leading digit of a number in Ω.

We can begin by defining the set of numbers in Ω with leading digit d. We can represent this in set notation as

$$S(d) = \bigcup_{n=-\infty}^{\infty} [d \cdot 10^n, (d+1) \cdot 10^n)$$

Let $f(x)$ be a continuous density function on Ω, then it immediately follows that

$$P(d) = \sum_{n=-\infty}^{\infty} \int_{d \cdot 10^n}^{(d+1) \cdot 10^n} f(x)\, dx$$

Mathematical Derivation

Let $\Omega = \mathbb{R}^+$ be our sample space.

We want to find a probability measure $P(d)$ on Ω, where $d \in [0..9]$ is the leading digit of a number in Ω.

We can begin by defining the set of numbers in Ω with leading digit d. We can represent this in set notation as

$$S(d) = \bigcup_{n=-\infty}^{\infty} [d \cdot 10^n, (d+1) \cdot 10^n)$$

Let $f(x)$ be a continuous density function on Ω, then it immediately follows that

$$P(d) = \sum_{n=-\infty}^{\infty} \int_{d \cdot 10^n}^{(d+1) \cdot 10^n} f(x)\, dx$$

Where $P(d)$ is the probability of picking a number from distribution $f(x)$ beginning with d.

Mathematical Derivation

Introducing $\Delta n = 1$, we can approximate the double integral,

Mathematical Derivation

Introducing $\Delta n = 1$, we can approximate the double integral,

$$P(d) = (\sum_{n=-\infty}^{\infty} \int_{d\cdot 10^n}^{(d+1)\cdot 10^n} f(x)\,dx)\Delta n \approx \int_{-\infty}^{\infty} (\int_{d\cdot 10^n}^{(d+1)\cdot 10^n} f(x)\,dx)\,dn$$

Mathematical Derivation

Introducing $\Delta n = 1$, we can approximate the double integral,

$$P(d) = (\sum_{n=-\infty}^{\infty} \int_{d \cdot 10^n}^{(d+1) \cdot 10^n} f(x)\, dx) \Delta n \approx \int_{-\infty}^{\infty} (\int_{d \cdot 10^n}^{(d+1) \cdot 10^n} f(x)\, dx)\, dn$$

Make the following substitutions:

$$t = d \cdot 10^n; \;\; dn = \frac{dt}{tln(10)}$$
$$x = ty; \;\; dx = tdy$$

Mathematical Derivation

Introducing $\Delta n = 1$, we can approximate the double integral,

$$P(d) = (\sum_{n=-\infty}^{\infty} \int_{d \cdot 10^n}^{(d+1)\cdot 10^n} f(x)\,dx)\Delta n \approx \int_{-\infty}^{\infty} (\int_{d\cdot 10^n}^{(d+1)\cdot 10^n} f(x)\,dx)\,dn$$

Make the following substitutions:

$$t = d \cdot 10^n; \;\; dn = \frac{dt}{tln(10)}$$
$$x = ty; \;\; dx = tdy$$

Giving

$$P(d) \approx \int_0^{\infty} \int_1^{1+\frac{1}{d}} f(ty)t\,dy \cdot \frac{dt}{t \ln 10} = \frac{1}{\ln 10} \int_0^{\infty} dt \int_1^{1+\frac{1}{d}} f(ty)\,dy$$

Mathematical Derivation

Let $f(ty) = \phi(y, t)$. By Fubini's Theorem,

Mathematical Derivation

Let $f(ty) = \phi(y,t)$. By Fubini's Theorem,

$$P(d) \approx \frac{1}{ln10} \int_0^\infty dt \int_1^{1+\frac{1}{d}} \phi(y,t)\, dy = \frac{1}{ln10} \int_1^{1+\frac{1}{d}} dy \int_0^\infty \phi(y,t)\, dt$$

Mathematical Derivation

Let $f(ty) = \phi(y,t)$. By Fubini's Theorem,

$$P(d) \approx \frac{1}{ln10} \int\limits_0^{\infty} dt \int\limits_1^{1+\frac{1}{d}} \phi(y,t)\, dy = \frac{1}{ln10} \int\limits_1^{1+\frac{1}{d}} dy \int\limits_0^{\infty} \phi(y,t)\, dt$$

$$= \frac{1}{\ln 10} \int\limits_1^{1+\frac{1}{d}} \frac{1}{y} dy \int\limits_0^{\infty} f(x)\, dx = \frac{\ln 1+\frac{1}{d}}{\ln 10} \int\limits_{\Omega} f(x)\, dx = \frac{\ln 1+\frac{1}{d}}{\ln 10}$$

Mathematical Derivation

Let $f(ty) = \phi(y,t)$. By Fubini's Theorem,

$$P(d) \approx \frac{1}{ln10} \int\limits_0^\infty dt \int\limits_1^{1+\frac{1}{d}} \phi(y,t)\, dy = \frac{1}{ln10} \int\limits_1^{1+\frac{1}{d}} dy \int\limits_0^\infty \phi(y,t)\, dt$$

$$= \frac{1}{\ln 10} \int\limits_1^{1+\frac{1}{d}} \frac{1}{y} dy \int\limits_0^\infty f(x)\, dx = \frac{\ln 1+\frac{1}{d}}{\ln 10} \int\limits_\Omega f(x)\, dx = \frac{\ln 1+\frac{1}{d}}{\ln 10}$$

By the change-of-base rule for logarithms, we are left with,

Mathematical Derivation

Let $f(ty) = \phi(y,t)$. By Fubini's Theorem,

$$P(d) \approx \frac{1}{ln10} \int\limits_0^\infty dt \int\limits_1^{1+\frac{1}{d}} \phi(y,t)\,dy = \frac{1}{ln10} \int\limits_1^{1+\frac{1}{d}} dy \int\limits_0^\infty \phi(y,t)\,dt$$

$$= \frac{1}{\ln 10} \int\limits_1^{1+\frac{1}{d}} \frac{1}{y} dy \int\limits_0^\infty f(x)\,dx = \frac{\ln 1+\frac{1}{d}}{\ln 10} \int\limits_\Omega f(x)\,dx = \frac{\ln 1+\frac{1}{d}}{\ln 10}$$

By the change-of-base rule for logarithms, we are left with,

Benford's Law

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

Benford's Law Derivation

Thus, even though many common sequences... do not follow Benford's Law, those that do are so ubiquitous that many authors have assumed that a simple explanation must exist... [however], there does not appear to be a simple derivation of Benford's Law that both offers a "correct explanation" and... provide(s) insight. —*Arno Berger (2011)*

Benford's Law Derivation

Thus, even though many common sequences... do not follow Benford's Law, those that do are so ubiquitous that many authors have assumed that a simple explanation must exist... [however], there does not appear to be a simple derivation of Benford's Law that both offers a "correct explanation" and... provide(s) insight. —*Arno Berger (2011)*

I think in statistics we need derivations, not proofs. That is, lines of reasoning from some assumptions to a formula, or a procedure, which may or may not have certain properties in a given context, but which, all going well, might provide some insight. —*Terry Speed (2009)*

Questions you may have...

- How do we derive the distribution function for Benford's Law?
- **Which distributions of numbers follow Benford's Law?**

Which distributions of numbers follow Benford's Law?

A sequence (x_n) is said to be *Benford* if,

Which distributions of numbers follow Benford's Law?

A sequence (x_n) is said to be *Benford* if,

Benford Sequence

$$\lim_{N\to+\infty} \frac{\#\{1\leq n\leq N : S(x_n)\leq t\}}{N} = \log t, \quad \text{for all } t \in [1, 10)$$

Where $\#\{\cdot\}$ denotes the number of elements in the set.

Example (Natural Numbers)

Is the sequence of natural numbers $(x_n) = n$ Benford?

Example (Natural Numbers)

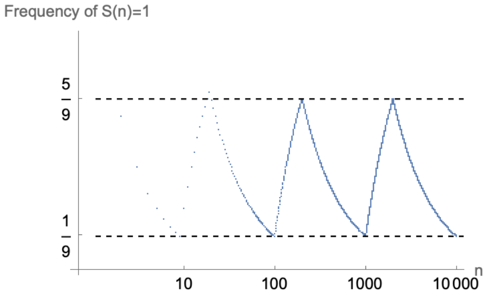
Is the sequence of natural numbers $(x_n) = n$ Benford?

Let's look at a plot of natural numbers n vs. $\frac{1}{n} \cdot \#\{N \in [1, n] : S(N) = 1\}$

Example (Natural Numbers)

Is the sequence of natural numbers $(x_n) = n$ Benford?

Let's look at a plot of natural numbers n vs. $\frac{1}{n} \cdot \#\{N \in [1, n] : S(N) = 1\}$



Example (Natural Numbers)

As we might expect, we see that,

$$\liminf N \to +\infty\left(\frac{\#\{N\in[1,n]:S(N)=1\}}{n}\right) = \frac{1}{9}$$

and,

$$\limsup N \to +\infty\left(\frac{\#\{N\in[1,n]:S(N)=1\}}{n}\right) = \frac{5}{9}$$

Example (Natural Numbers)

As we might expect, we see that,

$$\liminf N \to +\infty(\frac{\#\{N \in [1,n] : S(N)=1\}}{n}) = \frac{1}{9}$$

and,

$$\limsup N \to +\infty(\frac{\#\{N \in [1,n] : S(N)=1\}}{n}) = \frac{5}{9}$$

So the limit does not exist, and $(x_n) = n$ is not Benford!

Example (Exponential)

Is the sequence $(x_n) = 2^n$ Benford?

Example (Exponential)

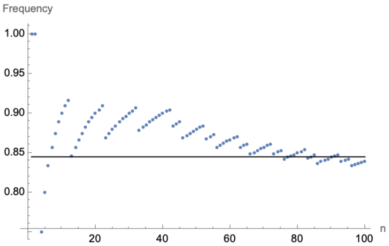
Is the sequence $(x_n) = 2^n$ Benford?

Yes. Why?

Example (Exponential)

Is the sequence $(x_n) = 2^n$ Benford?

Yes. Why?

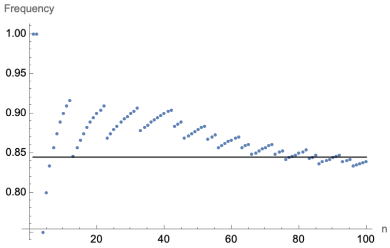


Plot of n vs. $\frac{1}{n} \cdot \#\{N \in [1, n] : S(2^N) \leq 7\}$ with line $y = \log(7)$ shown.

Example (Exponential)

Is the sequence $(x_n) = 2^n$ Benford?

Yes. Why?



Plot of n vs. $\frac{1}{n} \cdot \#\{N \in [1, n] : S(2^N) \leq 7\}$ with line $y = \log(7)$ shown.

Proof later.

Example (Exponential)

Is the sequence $(x_n) = 2^n$ Benford?

What about in **base 2**?

Example (Exponential)

Is the sequence $(x_n) = 2^n$ Benford?

What about in **base 2**?

$(x_n) = {2^n}_{\text{base } 2} = 1, 10, 100, 1000, 10000, \ldots$

Example (Exponential)

Is the sequence $(x_n) = 2^n$ Benford?

What about in **base 2**?

$(x_n) = {2^n}_{\text{base } 2} = 1, 10, 100, 1000, 10000, ...$

$Prob({D_2}^{(2)} = 0) = 1 - Prob({D_2}^{(2)} = 1) = \log_2(3) - 1 > \frac{1}{2}$

So $(x_n) = {2^n}_{\text{base } 2}$ is not Benford!

Other Benford Sequences

- The Fibonacci Sequence

Other Benford Sequences

- The Fibonacci Sequence
- $(f^n(x_0))$ where $f(x) = ax^b$ with $a > 0, b > 1$
  - Benford for almost all $x_0 > 0$, but every non-empty open interval in $\mathbb{R}^+$ contains uncountably many x_0 for which $(f^n(x_0))$ is **not** Benford.

Other Benford Sequences

- The Fibonacci Sequence
- $(f^n(x_0))$ where $f(x) = ax^b$ with $a > 0, b > 1$
  - Benford for almost all $x_0 > 0$, but every non-empty open interval in $\mathbb{R}^+$ contains uncountably many x_0 for which $(f^n(x_0))$ is **not** Benford.
- (θ^n) for any irrational θ.

Other Benford Sequences

- The Fibonacci Sequence
- $(f^n(x_0))$ where $f(x) = ax^b$ with $a > 0, b > 1$
  - Benford for almost all $x_0 > 0$, but every non-empty open interval in $\mathbb{R}^+$ contains uncountably many x_0 for which $(f^n(x_0))$ is **not** Benford.
- (θ^n) for any irrational θ.
- Prime numbers
  - "Logarithmic Benford"
  - Logarithmic density of $\{n \in \mathbb{N} : S(x_n) \leq t\} = \log(t)$

Application to Newton's Method

Newton's Method is used to approximate the roots of real-valued functions using the function,

$$N_g(x) = x - \frac{g(x)}{g'(x)}$$

Application to Newton's Method

Newton's Method is used to approximate the roots of real-valued functions using the function,

$$N_g(x) = x - \frac{g(x)}{g'(x)}$$

It can be shown that for x_0 sufficiently close to a root x' (i.e. $g(x') = 0$), that

$$\lim_{n\to\infty}({N_g}^n(x_0)) = x'$$

Application to Newton's Method

Theorem 3.1

Let the function $g : I \to \mathbb{R}$ be real-analytic with $g(x') = 0$, and assume that g is not linear.

(i) If x' is a simple root (multiplicity 1), then $(x_n - x')$ and $(x_{n+1} - x_n)$ are both Benford for almost all x_0 in a neighborhood of x'.

(ii) If x' has multiplicity ≥ 2, then $(x_n - x')$ and $(x_{n+1} - x_n)$ are Benford for all $x_0 \neq x'$ sufficiently close to x'.

Application to Newton's Method

Example: Let $g(x) = e^x - 2$, then g has a root at $x' = \ln(2)$ and $N_g(x) = x - 1 + 2e^{-x}$. By the above theorem, the sequences $(x_n - x')$ and $(x_{n+1} - x_n)$ are both Benford for almost all x_0 near x'.

Application to Newton's Method

Example: Let $g(x) = e^x - 2$, then g has a root at $x' = \ln(2)$ and $N_g(x) = x - 1 + 2e^{-x}$. By the above theorem, the sequences $(x_n - x')$ and $(x_{n+1} - x_n)$ are both Benford for almost all x_0 near x'.

Why it matters?: In computer algorithms, roundoff errors are inevitable. In computer implementations of Newton's Method, there is normally an assumption of uniformly distributed fraction parts. Such an assumption would lead to an underestimate in the average relative round-off error in the above case.

Application to Newton's Method

"[I]n order to analyze the average behavior of floating-point arithmetic algorithms, we need some statistical information that allows us to determine how often various cases arise... [If, for example, the] leading digits tend to be small [, that] makes the most obvious techniques of "average error" estimation for floating-point calculations invalid. The relative error due to rounding is usually... more than expected. —Donald Knuth, The Art of Computer Programming (1968)

Application to Newton's Method

"[I]n order to analyze the average behavior of floating-point arithmetic algorithms, we need some statistical information that allows us to determine how often various cases arise... [If, for example, the] leading digits tend to be small [, that] makes the most obvious techniques of "average error" estimation for floating-point calculations invalid. The relative error due to rounding is usually... more than expected. —Donald Knuth, The Art of Computer Programming (1968)

Question: Can Benford's Law improve current roundoff error approximation techniques in floating-point arithmetic?

Benford Random Variables

Benford Random Variables

A random variable X on probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is Benford if

$$P(S(X) \leq t) = \log(t) \text{ for all } t \in [1, 10)$$

Benford Random Variables

Benford Random Variables

A random variable X on probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is Benford if

$$P(S(X) \leq t) = \log(t) \text{ for all } t \in [1, 10)$$

Or equivalently, $S(X)$ is an absolutely continous random variable with density $f_{S(X)}(t) = t^{-1} \log(e)$

Benford Random Variables

Benford Random Variables

A random variable X on probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is Benford if

$$P(S(X) \leq t) = \log(t) \text{ for all } t \in [1, 10)$$

Or equivalently, $S(X)$ is an absolutely continous random variable with density $f_{S(X)}(t) = t^{-1}\log(e)$

Examples:

- $\mathbb{P}(D_1(X) = 1) = \mathbb{P}(1 \leq S(X) < 2) = \log(2)$

Benford Random Variables

Benford Random Variables

A random variable X on probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is Benford if

$$P(S(X) \leq t) = \log(t) \text{ for all } t \in [1, 10)$$

Or equivalently, $S(X)$ is an absolutely continous random variable with density $f_{S(X)}(t) = t^{-1}\log(e)$

Examples:

- $\mathbb{P}(D_1(X) = 1) = \mathbb{P}(1 \leq S(X) < 2) = \log(2)$
- $\mathbb{P}(D_1(X) = 9) = \log(\frac{10}{9})$

Benford Random Variables

Benford Random Variables

A random variable X on probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is Benford if

$$P(S(X) \leq t) = \log(t) \text{ for all } t \in [1, 10)$$

Or equivalently, $S(X)$ is an absolutely continous random variable with density $f_{S(X)}(t) = t^{-1}\log(e)$

Examples:

- $\mathbb{P}(D_1(X) = 1) = \mathbb{P}(1 \leq S(X) < 2) = \log(2)$
- $\mathbb{P}(D_1(X) = 9) = \log(\frac{10}{9})$
- $\mathbb{P}(D_1(X) = 3, D_2(X) = 1, D_3(X) = 4) = \log(\frac{315}{314})$

$\mathbb{N}$-Valued Random Variables

One might consider classifying $\mathbb{N}$-valued random variables (i.e. $\mathbb{P}(X \in \mathbb{N}) = 1$) as *Benford on* N if

$\mathbb{N}$-Valued Random Variables

One might consider classifying $\mathbb{N}$-valued random variables (i.e. $\mathbb{P}(X \in \mathbb{N}) = 1$) as *Benford on* N if

$$\mathbb{P}(S(X) \leq t) = \log(t)$$

$\mathbb{N}$-Valued Random Variables

One might consider classifying $\mathbb{N}$-valued random variables (i.e. $\mathbb{P}(X \in \mathbb{N}) = 1$) as *Benford on* N if

$$\mathbb{P}(S(X) \leq t) = \log(t)$$

However, no such random variable exists!

Which Distributions are Benford?

None of the standard continuous probability distributions (e.g., uniform, exponential, normal, etc.) are Benford, however their deviation from Benford's law can be quantified using the metric

Which Distributions are Benford?

None of the standard continuous probability distributions (e.g., uniform, exponential, normal, etc.) are Benford, however their deviation from Benford's law can be quantified using the metric

$$\Delta_\infty := 100 \cdot \sup_{1 \leq t < 10} |F_{S(X)}(t) - \log(t)|$$

Which Distributions are Benford?

None of the standard continuous probability distributions (e.g., uniform, exponential, normal, etc.) are Benford, however their deviation from Benford's law can be quantified using the metric

$$\Delta_\infty := 100 \cdot \sup_{1 \leq t < 10} |F_{S(X)}(t) - \log(t)|$$

Where $\Delta_\infty = 0$ if and only if X is Benford and $\Delta_\infty = 100$ if and only if $\mathbb{P}(S(X) = 1) = 1$

Example: Exponential Distribution

Consider the exponential distribution centered a 1 with cumulative distribution given by:

$$\begin{cases} 0 & \text{if } x < 0 \\ 1 - e^{-x} & \text{otherwise} \end{cases}$$

Example: Exponential Distribution

Consider the exponential distribution centered a 1 with cumulative distribution given by:

$$\begin{cases} 0 & \text{if } x < 0 \\ 1 - e^{-x} & \text{otherwise} \end{cases}$$

$$\mathbb{P}(D_1(X) = 1) = \mathbb{P}(X \in \bigcup_{k\in\mathbb{Z}} 10^k[1,2)) = \sum_{k\in\mathbb{Z}}(e^{-10^k} - e^{-2\cdot 10^k})$$

Example: Exponential Distribution

Consider the exponential distribution centered a 1 with cumulative distribution given by:

$$\begin{cases} 0 & \text{if } x < 0 \\ 1 - e^{-x} & \text{otherwise} \end{cases}$$

$$\mathbb{P}(D_1(X) = 1) = \mathbb{P}(X \in \bigcup_{k \in \mathbb{Z}} 10^k[1,2)) = \sum_{k \in \mathbb{Z}} (e^{-10^k} - e^{-2 \cdot 10^k})$$

$$\mathbb{P}(D_1(X) = 1) > (e^{\frac{1}{10}} - e^{\frac{-2}{10}}) + (e^{-1} - e^{-2}) + (e^{-10} - e^{-20}) \approx 0.3186 > \log(2)$$

Example: Exponential Distribution

Consider the exponential distribution centered a 1 with cumulative distribution given by:

$$\begin{cases} 0 & \text{if } x < 0 \\ 1 - e^{-x} & \text{otherwise} \end{cases}$$

$$\mathbb{P}(D_1(X) = 1) = \mathbb{P}(X \in \bigcup_{k\in\mathbb{Z}} 10^k[1,2)) = \sum_{k\in\mathbb{Z}}(e^{-10^k} - e^{-2\cdot 10^k})$$

$$\mathbb{P}(D_1(X) = 1) > (e^{\frac{1}{10}} - e^{\frac{-2}{10}}) + (e^{-1} - e^{-2}) + (e^{-10} - e^{-20}) \approx 0.3186 > \log(2)$$

$\Delta_\infty = 3.05$ i.e. $|\mathbb{P}(S(X) \leq t) - \log(t)|$ is small for $t \in [1, 10)$.

Other Common Distributions

Here is a table of other common distributions and how closely they follow Benford's Law:

Other Common Distributions

Here is a table of other common distributions and how closely they follow Benford's Law:

Distributions	Δ_∞
Uniform [0,1]	26.88
Exponential(1)	3.05
Pareto(1)	26.88
Arcsin	28.77
Standard Normal	6.05

Note on Uniform Distribution

Theorem 4.1

For every uniformly distributed positive random variable X,

$$\max_{1 \leq t < 10} |F_{S(X)}(t) - \log(t)| \geq \frac{1}{18} + \frac{1}{2}(\log(9) - \log(e) + \log\log(e)) \approx 0.1344$$

And this bound is *sharp*.

Note on Uniform Distribution

Theorem 4.1

For every uniformly distributed positive random variable X,

$$\max_{1\leq t<10}|F_{S(X)}(t)-\log(t)| \geq \frac{1}{18}+\frac{1}{2}(\log(9)-\log(e)+\log\log(e)) \approx 0.1344$$

And this bound is *sharp*.

Fallacy: Regularity and large spread implies Benford's Law.

Note on Uniform Distribution

Theorem 4.1

For every uniformly distributed positive random variable X,

$$\max_{1\leq t<10}|F_{S(X)}(t)-\log(t)| \geq \frac{1}{18}+\frac{1}{2}(\log(9)-\log(e)+\log\log(e)) \approx 0.1344$$

And this bound is *sharp*.

Fallacy: Regularity and large spread implies Benford's Law.

Now, this claim is clearly false. No matter how large the spread, if data follows a uniform distribution then it does not conform to Benford's Law.

Uniform Distribution Modulo 1

If a sequence is *uniformly distributed modulo 1* (u.d. mod 1), the distribution of it's fractional parts is uniform on the interval $[0, 1)$.

Uniform Distribution Modulo 1

If a sequence is *uniformly distributed modulo 1* (u.d. mod 1), the distribution of it's fractional parts is uniform on the interval $[0, 1)$.

For convenience, let $\langle x \rangle = x \mod 1$ denote the fractional part of x.

Uniform Distribution Modulo 1

If a sequence is *uniformly distributed modulo 1* (u.d. mod 1), the distribution of it's fractional parts is uniform on the interval $[0, 1)$.

For convenience, let $\langle x \rangle = x \mod 1$ denote the fractional part of x.

Uniform Distribution Modulo 1

A sequence $(x_n) = (x_1, x_2, ...)$ of real numbers is *u.d* mod 1 if

$$\lim_{N\to\infty} \frac{\#\{1 \le n \le N : \langle x_n \rangle \le s\}}{N} = s \text{ for all } s \in [0, 1)$$

Uniform Distribution Modulo 1 (Random Variables)

This definition has a natural extension to random variables, namely,

Uniform Distribution Modulo 1 (Random Variables)

A random variable (r.v.) X on a probability space $(\Omega, \sigma, \mathbb{P})$ is *u.d.* mod 1 if

$$\mathbb{P}(\langle X \rangle \leq s) = s \text{ for all } s \in [0, 1)$$

Connection to Benford's Law

Theorem 5.1

A sequence of real numbers or random variable is Benford if and only if the decimal logarithm of its absolute value is uniformly distributed modulo one.

Connection to Benford's Law

Theorem 5.1

A sequence of real numbers or random variable is Benford if and only if the decimal logarithm of its absolute value is uniformly distributed modulo one.

Importance: Theorem 5.1 is one of the main tools in the theory of Benford's law because it allows application of the powerful theory of uniform distribution modulo one.

Proof of Theorem 5.1

Let X be a random variable. Then, for all $s \in [0, 1)$,

Proof of Theorem 5.1

Let X be a random variable. Then, for all $s \in [0, 1)$,

$\mathbb{P}(\langle \log |X| \rangle \leq s)$

Proof of Theorem 5.1

Let X be a random variable. Then, for all $s \in [0,1)$,

$\mathbb{P}(\langle \log|X| \rangle \leq s) = \mathbb{P}(\log|X| \in \bigcup_{k \in \mathbb{Z}} [k, k+s])$

Proof of Theorem 5.1

Let X be a random variable. Then, for all $s \in [0, 1)$,

$\mathbb{P}(\langle \log |X| \rangle \leq s) = \mathbb{P}(\log |X| \in \bigcup_{k \in \mathbb{Z}} [k, k + s])$

$= \mathbb{P}(|X| \in \bigcup_{k \in \mathbb{Z}} [10^k, 10^{k+s}]) + \mathbb{P}(X = 0)$

Proof of Theorem 5.1

Let X be a random variable. Then, for all $s \in [0,1)$,

$\mathbb{P}(\langle \log |X| \rangle \leq s) = \mathbb{P}(\log |X| \in \bigcup_{k\in\mathbb{Z}}[k, k+s])$

$= \mathbb{P}(|X| \in \bigcup_{k\in\mathbb{Z}}[10^k, 10^{k+s}]) + \mathbb{P}(X = 0) = \mathbb{P}(S(X) \leq 10^s)$

Proof of Theorem 5.1

Let X be a random variable. Then, for all $s \in [0, 1)$,

$\mathbb{P}(\langle \log |X| \rangle \leq s) = \mathbb{P}(\log |X| \in \bigcup_{k \in \mathbb{Z}} [k, k + s])$

$= \mathbb{P}(|X| \in \bigcup_{k \in \mathbb{Z}} [10^k, 10^{k+s}]) + \mathbb{P}(X = 0) = \mathbb{P}(S(X) \leq 10^s)$

Recall that a random variable Y is Benford if

Proof of Theorem 5.1

Let X be a random variable. Then, for all $s \in [0, 1)$,

$\mathbb{P}(\langle \log |X| \rangle \leq s) = \mathbb{P}(\log |X| \in \bigcup_{k \in \mathbb{Z}} [k, k+s])$

$= \mathbb{P}(|X| \in \bigcup_{k \in \mathbb{Z}} [10^k, 10^{k+s}]) + \mathbb{P}(X = 0) = \mathbb{P}(S(X) \leq 10^s)$

Recall that a random variable Y is Benford if

$$\mathbb{P}(S(Y) \leq t) = \log(t)$$

Proof of Theorem 5.1

Let X be a random variable. Then, for all $s \in [0, 1)$,

$\mathbb{P}(\langle \log |X| \rangle \leq s) = \mathbb{P}(\log |X| \in \bigcup_{k \in \mathbb{Z}} [k, k + s])$

$= \mathbb{P}(|X| \in \bigcup_{k \in \mathbb{Z}} [10^k, 10^{k+s}]) + \mathbb{P}(X = 0) = \mathbb{P}(S(X) \leq 10^s)$

Recall that a random variable Y is Benford if

$$\mathbb{P}(S(Y) \leq t) = \log(t)$$

Hence, $\mathbb{P}(\langle \log |X| \rangle \leq s) = s \Leftrightarrow \mathbb{P}(S(X) \leq 10^s) = \log(10^s) = s$

□

Proof of Theorem 5.1

Let X be a random variable. Then, for all $s \in [0,1)$,

$\mathbb{P}(\langle \log|X| \rangle \leq s) = \mathbb{P}(\log|X| \in \bigcup_{k\in\mathbb{Z}}[k, k+s])$

$= \mathbb{P}(|X| \in \bigcup_{k\in\mathbb{Z}}[10^k, 10^{k+s}]) + \mathbb{P}(X=0) = \mathbb{P}(S(X) \leq 10^s)$

Recall that a random variable Y is Benford if

$$\mathbb{P}(S(Y) \leq t) = \log(t)$$

Hence, $\mathbb{P}(\langle \log|X| \rangle \leq s) = s \Leftrightarrow \mathbb{P}(S(X) \leq 10^s) = \log(10^s) = s$

□

The proofs for sequences are completely analagous.

Applications

Proposition: Let $(x_n) = (x_1, x_2, ...)$ be a sequence of real numbers.

If $\lim_{n\to\infty}(x_{n+1} - x_n) = \theta$ for some irrational θ, then (x_n) is *u.d.* mod 1.

Applications

Proposition: Let $(x_n) = (x_1, x_2, ...)$ be a sequence of real numbers.

If $\lim_{n\to\infty}(x_{n+1} - x_n) = \theta$ for some irrational θ, then (x_n) is *u.d.* mod 1.

Example: Consider the family of sequences $(d_n) = (n\log(\alpha))$. If $\log(\alpha)$ is irrational, i.e. $\alpha = 2$ or $\alpha = \pi$, then by the above proposition (d_n) is *u.d.* mod 1.

Applications

Proposition: Let $(x_n) = (x_1, x_2, ...)$ be a sequence of real numbers.

If $\lim_{n\to\infty}(x_{n+1} - x_n) = \theta$ for some irrational θ, then (x_n) is *u.d.* mod 1.

Example: Consider the family of sequences $(d_n) = (n\log(\alpha))$. If $\log(\alpha)$ is irrational, i.e. $\alpha = 2$ or $\alpha = \pi$, then by the above proposition (d_n) is *u.d.* mod 1.

The corresponding Benford sequences (those whose decimal logarithm is given by (d_n)) are 2^n and π^n.

Applications

Proposition: Let $(x_n) = (x_1, x_2, ...)$ be a sequence of real numbers.

If $\lim_{n\to\infty}(x_{n+1} - x_n) = \theta$ for some irrational θ, then (x_n) is *u.d.* mod 1.

Example: Consider the family of sequences $(d_n) = (n\log(\alpha))$. If $\log(\alpha)$ is irrational, i.e. $\alpha = 2$ or $\alpha = \pi$, then by the above proposition (d_n) is *u.d.* mod 1.

The corresponding Benford sequences (those whose decimal logarithm is given by (d_n)) are 2^n and π^n.

It is easy to show through this method that, for instance, θ^n is Benford for any irrational θ.

US Taxpayer Records

Although US financial data is safeguarded, forensic analysist Mark Nigiri sourced 157,518 taxpayer records from 1978 for analysis using Benford's Law.

US Taxpayer Records

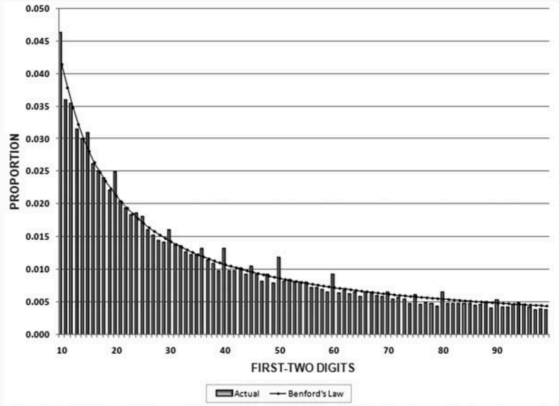
Although US financial data is safeguarded, forensic analysist Mark Nigiri sourced 157,518 taxpayer records from 1978 for analysis using Benford's Law.

Note that we can use the general Benford's Law given by

$$Prob(D_1 = d_1, D_2 = d_2, ..., D_m = d_m) = \log_{10}(1 + (\textstyle\sum_{j=1}^{m} 10^{m-j} d_j)^{-1})$$

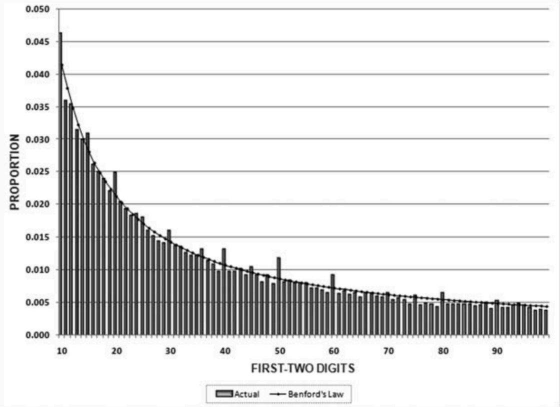
to compute the **joint probability** for the of the first *n* digits occurring.

US Taxpayer Records



Frequency of first two digits: Dividend income declared

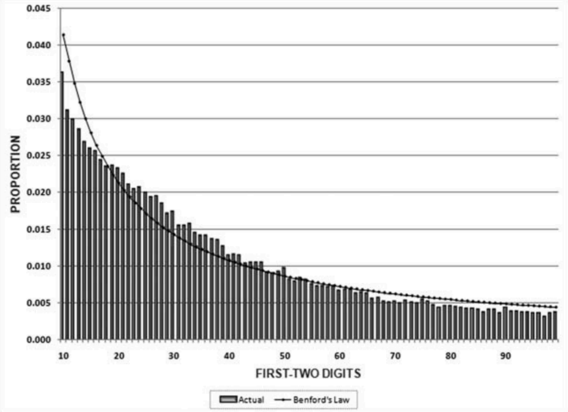
US Taxpayer Records



Frequency of first two digits: Dividend income declared

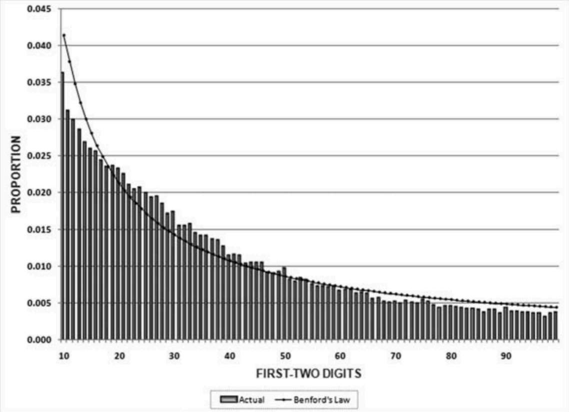
Question: Why are there spikes at multiples of 10?

US Taxpayer Records



Frequency of first two digits: Interest expense claimed

US Taxpayer Records



Frequency of first two digits: Interest expense claimed

Question: Why are the higher values supressed here?

2020 US Presidential Election

Following the 2020 US presidential election, many online debates were started due to some election data seemingly not matching Benford's Law.

2020 US Presidential Election

Following the 2020 US presidential election, many online debates were started due to some election data seemingly not matching Benford's Law.

Specifically, online threads opened about the legitimacy of the election data reported from the city of Chicago.

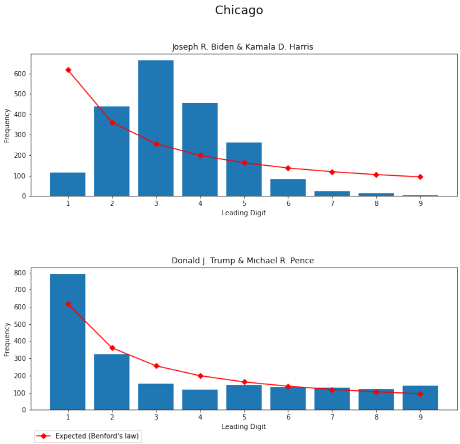
2020 US Presidential Election

Following the 2020 US presidential election, many online debates were started due to some election data seemingly not matching Benford's Law.

Specifically, online threads opened about the legitimacy of the election data reported from the city of Chicago.

The city of Chicago has 2,069 precincts which report election data. Each precinct is roughly the same size, with the smallest reporting 39 votes, and the biggest 1655, with an average of 516 and a standard deviation of 173.

2020 US Presidential Election



Plots of 2020 Chicago presidential election data by candiadate for 2,069 precincts with the predicted values by Benford's Law shown.

2020 US Presidential Election

[I]f a competitive two candidate race occurs in districts whose magnitude varies between 100 and 1000, the modal first digit for each candidate's vote will not be 1 or 2 but rather 4, 5, or 6. — Henry E. Brady (2005)

References

Berger, A., Hill, T.P. (2015). An Introduction to Benford's Law. Princeton University Press.

Newcomb, S. (1881). Note on the Frequency of Use of the Different Digits in Natural Numbers. American Journal of Mathematics, 4(1), 39-40. http://links.jstor.org/sici?sici=0002-9327%281881%294%3A1%3C39%3ANOTFOU%3E2.0.CO%3B2-V

Benford, F. (1938). The Law of Anomalous Numbers. Proceedings of the American Philosophical Society, 78(4), 551-572. https://www.jstor.org/stable/984802

Wang, L., Ma, B. Q. (2023). A concise proof of Benford's Law. Fundamental Research, 1-5. https://doi.org/10.1016/j.fmre.2023.01.002

Berger, A., Hill, T.P. Benford's Law Strikes Back: No Simple Explanation in Sight for Mathematical Gem. Math Intelligencer 33, 85–91 (2011). https://doi.org/10.1007/s00283-010-9182-3

Speed, T. (2009), ''You Want Proof?'', Bull. Inst. Math. Statistics 38, p 11.

Nigiri, M. J. (2009). Benford's Law (pp. 283-285). John Wiley Sons, Inc.

[Stand-Up Maths]. (2020, November 11). Why do Biden's votes not follow Benford's Law? [Video]. Youtube.com. https://www.youtube.com/watch?v=etx0k1nLn78ab$_channel = Stand - upMaths$

Deckert, J., Myagkov, M., Ordeshook, P. (2011). Benford's Law and the Detection of Election Fraud. Political Analysis, 19(3), 245-268. doi:10.1093/pan/mpr014