

Why the IRS cares about the Riemann Zeta Function and Number Theory (and why you should too!)

Steven J. Miller
sjm1@williams.edu

[http://web.williams.edu/Mathematics/
sjmiller/public_html/](http://web.williams.edu/Mathematics/sjmiller/public_html/)

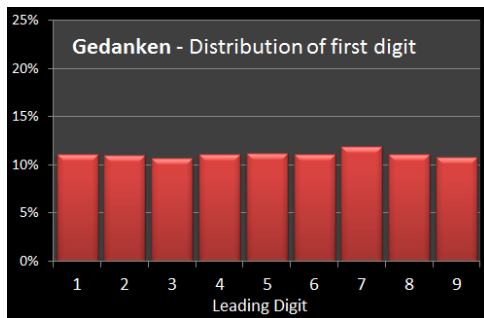
Faculty Lecture Series, Williams College, 2/12/15

Interesting Question

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

Interesting Question

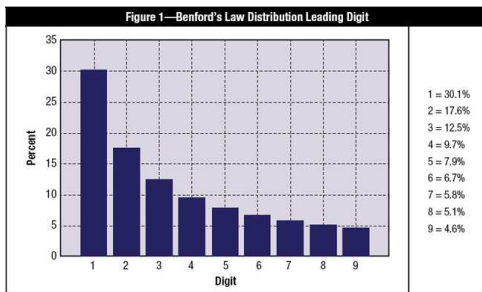
Motivating Question: For a nice data set, such as the Fibonacci numbers, stock prices, street addresses of Williams employees and students, ..., what percent of the leading digits are 1?



Natural guess: 10% (but immediately correct to 11%!).

Interesting Question

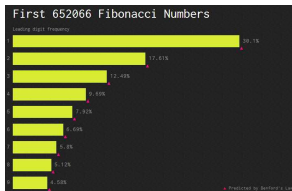
Motivating Question: For a nice data set, such as the Fibonacci numbers, stock prices, street addresses of Williams employees and students, ..., what percent of the leading digits are 1?



Answer: Benford's law!

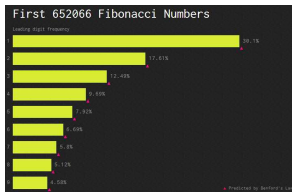
Examples with First Digit Bias

Fibonacci numbers

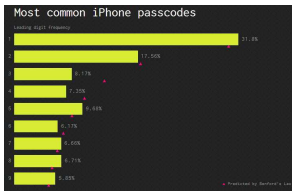


Examples with First Digit Bias

Fibonacci numbers

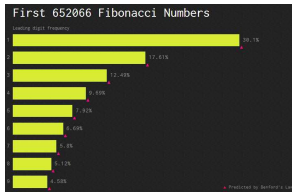


Most common iPhone passcodes

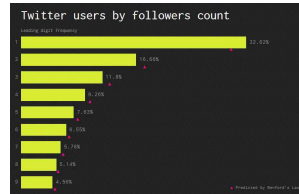


Examples with First Digit Bias

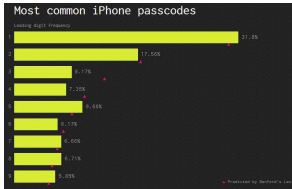
Fibonacci numbers



Twitter users by # followers

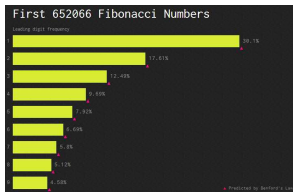


Most common iPhone passcodes

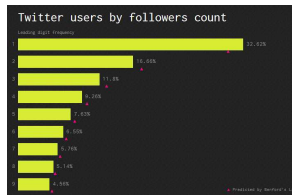


Examples with First Digit Bias

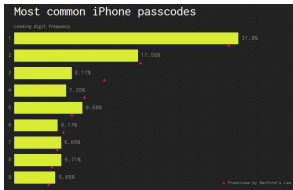
Fibonacci numbers



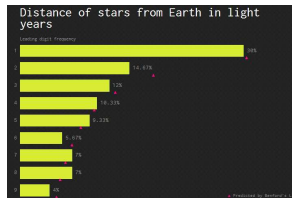
Twitter users by # followers



Most common iPhone passcodes



Distance of stars from Earth



Summary

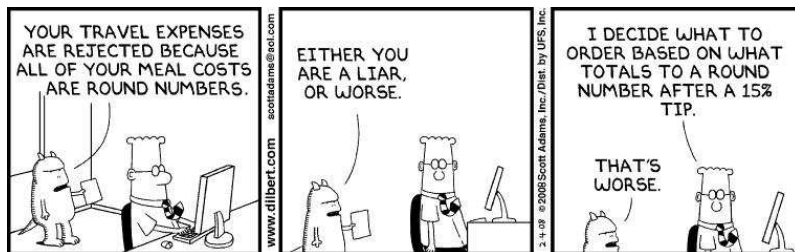
- Explain Benford's Law.
- Discuss examples and applications.
- Sketch proofs.
- Describe open problems.

Caveats!

- A math test indicating fraud is *not* proof of fraud: unlikely events, alternate reasons.

Caveats!

- A math test indicating fraud is *not* proof of fraud: unlikely events, alternate reasons.



Examples

- recurrence relations
- special functions (such as $n!$)
- iterates of power, exponential, rational maps
- products of random variables
- L -functions, characteristic polynomials
- iterates of the $3x + 1$ map
- differences of order statistics
- hydrology and financial data
- many hierarchical Bayesian models

Applications

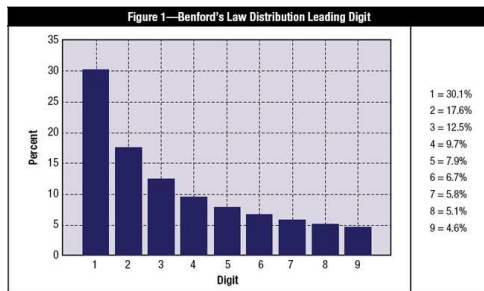
- Analyzing round-off errors.
- Determining the optimal way to store numbers.
- Detecting tax and image fraud, and data integrity.

General Theory

Benford's Law: Newcomb (1881), Benford (1938)

Statement

For many data sets, probability of observing a first digit of d base B is $\log_B \left(\frac{d+1}{d} \right)$; base 10 about 30% are 1s.



Benford's Law (probabilities)

Background Material

- Modulo: $a = b \bmod c$ if $a - b$ is an integer times c ; thus $17 = 5 \bmod 12$, and $4.5 = .5 \bmod 1$.

Background Material

- Modulo: $a = b \bmod c$ if $a - b$ is an integer times c ; thus $17 = 5 \bmod 12$, and $4.5 = .5 \bmod 1$.
- Significand: $x = S_{10}(x) \cdot 10^k$, k integer, $1 \leq S_{10}(x) < 10$.

Background Material

- Modulo: $a = b \bmod c$ if $a - b$ is an integer times c ; thus $17 = 5 \bmod 12$, and $4.5 = .5 \bmod 1$.
- Significand: $x = S_{10}(x) \cdot 10^k$, k integer, $1 \leq S_{10}(x) < 10$.
- $S_{10}(x) = S_{10}(\tilde{x})$ if and only if x and $\tilde{x}$ have the same leading digits. Note $\log_{10} x = \log_{10} S_{10}(x) + k$.

Background Material

- Modulo: $a = b \bmod c$ if $a - b$ is an integer times c ; thus $17 = 5 \bmod 12$, and $4.5 = .5 \bmod 1$.
- Significand: $x = S_{10}(x) \cdot 10^k$, k integer, $1 \leq S_{10}(x) < 10$.
- $S_{10}(x) = S_{10}(\tilde{x})$ if and only if x and $\tilde{x}$ have the same leading digits. Note $\log_{10} x = \log_{10} S_{10}(x) + k$.
- **Key observation:** $\log_{10}(x) = \log_{10}(\tilde{x}) \bmod 1$ if and only if x and $\tilde{x}$ have the same leading digits.

Thus often study $y = \log_{10} x \bmod 1$.

Advanced: $e^{2\pi i u} = e^{2\pi i (u \bmod 1)}$.

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 $b - a$:

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

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 $b - a$:

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

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

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 $b - a$:

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

- Thm: $\beta \notin \mathbb{Q}$, $n\beta$ is equidistributed mod 1.
- Examples: $\log_{10} 2, \log_{10} \left(\frac{1+\sqrt{5}}{2} \right) \notin \mathbb{Q}$.

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 $b - a$:

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

- Thm: $\beta \notin \mathbb{Q}$, $n\beta$ is equidistributed mod 1.
- Examples: $\log_{10} 2, \log_{10} \left(\frac{1+\sqrt{5}}{2}\right) \notin \mathbb{Q}$.
Proof: if rational: $2 = 10^{p/q}$.

Equidistribution and Benford's Law

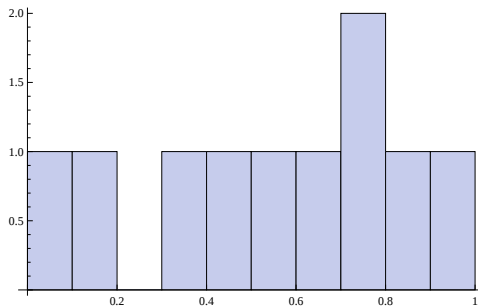
Equidistribution

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

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

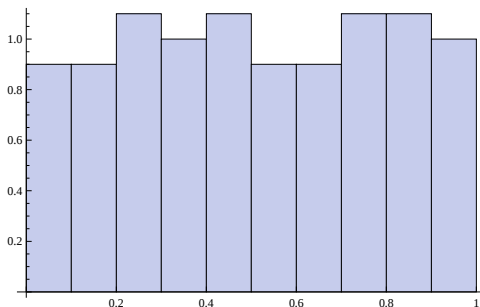
- Thm: $\beta \notin \mathbb{Q}$, $n\beta$ is equidistributed mod 1.
- Examples: $\log_{10} 2, \log_{10} \left(\frac{1+\sqrt{5}}{2}\right) \notin \mathbb{Q}$.
Proof: if rational: $2 = 10^{p/q}$.
 Thus $2^q = 10^p$ or $2^{q-p} = 5^p$, impossible.

Example of Equidistribution: $n\sqrt{\pi} \bmod 1$



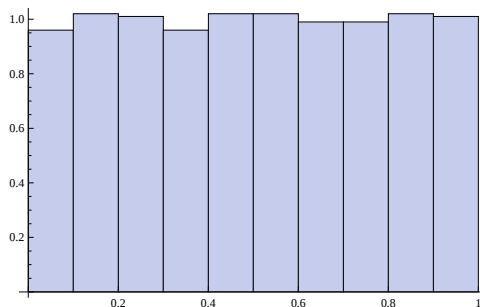
$n\sqrt{\pi} \bmod 1$ for $n \leq 10$

Example of Equidistribution: $n\sqrt{\pi} \bmod 1$



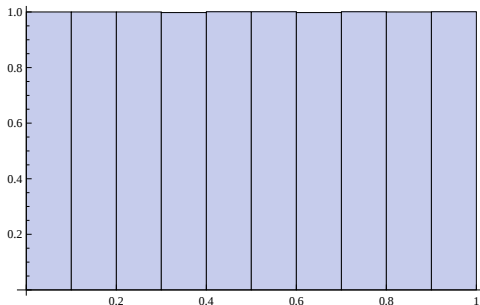
$n\sqrt{\pi} \bmod 1$ for $n \leq 100$

Example of Equidistribution: $n\sqrt{\pi} \bmod 1$



$n\sqrt{\pi} \bmod 1$ for $n \leq 1000$

Example of Equidistribution: $n\sqrt{\pi} \bmod 1$



$n\sqrt{\pi} \bmod 1$ for $n \leq 10,000$

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 $y_i = \log_B x_i$.

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

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

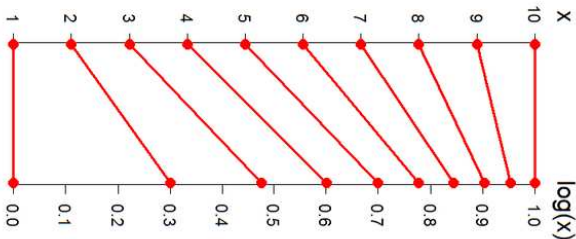
Logarithms and Benford's Law

Fundamental Equivalence

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

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

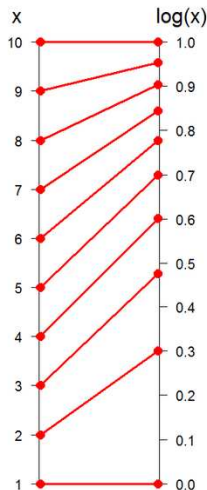
$$\log_{10} x = \log_{10} S_{10}(x) + k = \log_{10} S_{10}x \bmod 1.$$



Logarithms and Benford's Law

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

Have Benford's law $\leftrightarrow$
 mantissa of logarithms
 of data are uniformly
 distributed



Examples

- 2^n is Benford base 10 as $\log_{10} 2 \notin \mathbb{Q}$.

Examples

- Fibonacci numbers are Benford base 10.

Examples

- Fibonacci numbers are Benford base 10.

$$a_{n+1} = a_n + a_{n-1}.$$

Examples

- Fibonacci numbers are Benford base 10.

$$a_{n+1} = a_n + a_{n-1}.$$

Guess $a_n = r^n$: $r^{n+1} = r^n + r^{n-1}$ or $r^2 = r + 1$.

Examples

- Fibonacci numbers are Benford base 10.

$$a_{n+1} = a_n + a_{n-1}.$$

Guess $a_n = r^n$: $r^{n+1} = r^n + r^{n-1}$ or $r^2 = r + 1$.

Roots $r = (1 \pm \sqrt{5})/2$.

Examples

- Fibonacci numbers are Benford base 10.

$$a_{n+1} = a_n + a_{n-1}.$$

Guess $a_n = r^n$: $r^{n+1} = r^n + r^{n-1}$ or $r^2 = r + 1$.

Roots $r = (1 \pm \sqrt{5})/2$.

General solution: $a_n = c_1 r_1^n + c_2 r_2^n$.

Examples

- Fibonacci numbers are Benford base 10.

$$a_{n+1} = a_n + a_{n-1}.$$

Guess $a_n = r^n$: $r^{n+1} = r^n + r^{n-1}$ or $r^2 = r + 1$.

Roots $r = (1 \pm \sqrt{5})/2$.

General solution: $a_n = c_1 r_1^n + c_2 r_2^n$.

$$\text{Binet: } a_n = \frac{1}{\sqrt{5}} \left(\frac{1+\sqrt{5}}{2} \right)^n - \frac{1}{\sqrt{5}} \left(\frac{1-\sqrt{5}}{2} \right)^n.$$

Examples

- Fibonacci numbers are Benford base 10.

$$a_{n+1} = a_n + a_{n-1}.$$

Guess $a_n = r^n$: $r^{n+1} = r^n + r^{n-1}$ or $r^2 = r + 1$.

Roots $r = (1 \pm \sqrt{5})/2$.

General solution: $a_n = c_1 r_1^n + c_2 r_2^n$.

$$\text{Binet: } a_n = \frac{1}{\sqrt{5}} \left(\frac{1+\sqrt{5}}{2} \right)^n - \frac{1}{\sqrt{5}} \left(\frac{1-\sqrt{5}}{2} \right)^n.$$

- Most linear recurrence relations Benford:

Examples

- Fibonacci numbers are Benford base 10.

$$a_{n+1} = a_n + a_{n-1}.$$

Guess $a_n = r^n$: $r^{n+1} = r^n + r^{n-1}$ or $r^2 = r + 1$.

Roots $r = (1 \pm \sqrt{5})/2$.

General solution: $a_n = c_1 r_1^n + c_2 r_2^n$.

$$\text{Binet: } a_n = \frac{1}{\sqrt{5}} \left(\frac{1+\sqrt{5}}{2} \right)^n - \frac{1}{\sqrt{5}} \left(\frac{1-\sqrt{5}}{2} \right)^n.$$

- Most linear recurrence relations Benford:

$$\diamond a_{n+1} = 2a_n$$

Examples

- Fibonacci numbers are Benford base 10.

$$a_{n+1} = a_n + a_{n-1}.$$

Guess $a_n = r^n$: $r^{n+1} = r^n + r^{n-1}$ or $r^2 = r + 1$.

Roots $r = (1 \pm \sqrt{5})/2$.

General solution: $a_n = c_1 r_1^n + c_2 r_2^n$.

$$\text{Binet: } a_n = \frac{1}{\sqrt{5}} \left(\frac{1+\sqrt{5}}{2} \right)^n - \frac{1}{\sqrt{5}} \left(\frac{1-\sqrt{5}}{2} \right)^n.$$

- Most linear recurrence relations Benford:

$$\diamond a_{n+1} = 2a_n - a_{n-1}$$

Examples

- **Fibonacci numbers are Benford base 10.**

$$a_{n+1} = a_n + a_{n-1}.$$

Guess $a_n = r^n$: $r^{n+1} = r^n + r^{n-1}$ or $r^2 = r + 1$.

Roots $r = (1 \pm \sqrt{5})/2$.

General solution: $a_n = c_1 r_1^n + c_2 r_2^n$.

$$\text{Binet: } a_n = \frac{1}{\sqrt{5}} \left(\frac{1+\sqrt{5}}{2} \right)^n - \frac{1}{\sqrt{5}} \left(\frac{1-\sqrt{5}}{2} \right)^n.$$

- **Most linear recurrence relations Benford:**

$$\diamond a_{n+1} = 2a_n - a_{n-1}$$

$\diamond$ take $a_0 = a_1 = 1$ or $a_0 = 0, a_1 = 1$.

Digits of 2^n

First 60 values of 2^n (only displaying 30)

			digit	#	Obs Prob	Benf Prob
1	1024	1048576				
2	2048	2097152	1	18	.300	.301
4	4096	4194304	2	12	.200	.176
8	8192	8388608	3	6	.100	.125
16	16384	16777216	4	6	.100	.097
32	32768	33554432	5	6	.100	.079
64	65536	67108864	6	4	.067	.067
128	131072	134217728	7	2	.033	.058
256	262144	268435456	8	5	.083	.051
512	524288	536870912	9	1	.017	.046

Digits of 2^n

First 60 values of 2^n (only displaying 30)

			digit	#	Obs Prob	Benf Prob
1	1024	1048576				
2	2048	2097152	1	18	.300	.301
4	4096	4194304	2	12	.200	.176
8	8192	8388608	3	6	.100	.125
16	16384	16777216	4	6	.100	.097
32	32768	33554432	5	6	.100	.079
64	65536	67108864	6	4	.067	.067
128	131072	134217728	7	2	.033	.058
256	262144	268435456	8	5	.083	.051
512	524288	536870912	9	1	.017	.046

Digits of 2^n

First 60 values of 2^n (only displaying 30): $2^{10} = 1024 \approx 10^3$.

			digit	#	Obs Prob	Benf Prob
1	1024	1048576				
2	2048	2097152	1	18	.300	.301
4	4096	4194304	2	12	.200	.176
8	8192	8388608	3	6	.100	.125
16	16384	16777216	4	6	.100	.097
32	32768	33554432	5	6	.100	.079
64	65536	67108864	6	4	.067	.067
128	131072	134217728	7	2	.033	.058
256	262144	268435456	8	5	.083	.051
512	524288	536870912	9	1	.017	.046

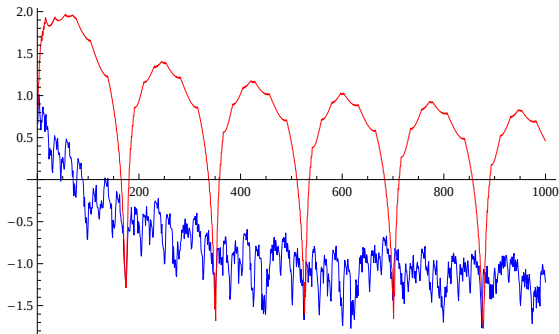
Logarithms and Benford's Law

χ^2 values for α^n , $1 \leq n \leq N$ (5% 15.5).

N	$\chi^2(\gamma)$	$\chi^2(e)$	$\chi^2(\pi)$
100	0.72	0.30	46.65
200	0.24	0.30	8.58
400	0.14	0.10	10.55
500	0.08	0.07	2.69
700	0.19	0.04	0.05
800	0.04	0.03	6.19
900	0.09	0.09	1.71
1000	0.02	0.06	2.90

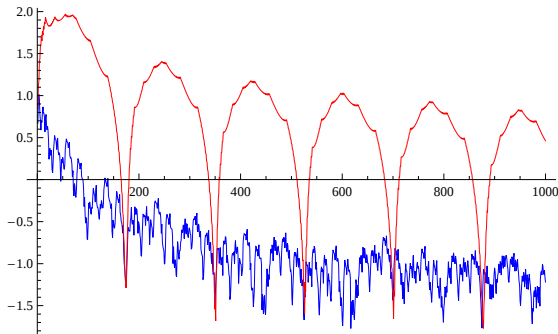
Logarithms and Benford's Law: Base 10 (5%: $\log(\chi^2) \approx 2.74$)

$\log(\chi^2)$ vs N for π^n (red) and e^n (blue),
 $n \in \{1, \dots, N\}$.



Logarithms and Benford's Law: Base 10 (5%: $\log(\chi^2) \approx 2.74$)

$\log(\chi^2)$ vs N for π^n (red) and e^n (blue),
 $n \in \{1, \dots, N\}$. Note $\pi^{175} \approx 1.0028 \cdot 10^{87}$.



Why Benford's Law?

Streets

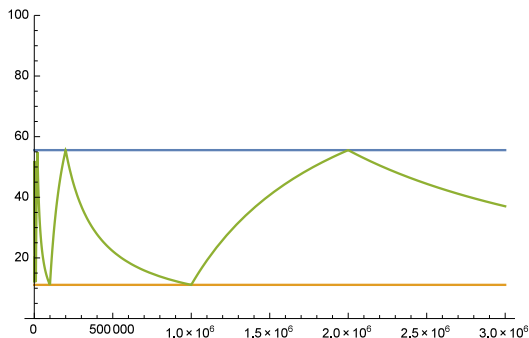
Not all data sets satisfy Benford's Law.

- Long street $[1, L]$: $L = 199$ versus $L = 999$.
- Oscillates b/w $1/9$ and $5/9$ with first digit 1.

Streets

Not all data sets satisfy Benford's Law.

- Long street $[1, L]$: $L = 199$ versus $L = 999$.
- Oscillates b/w $1/9$ and $5/9$ with first digit 1.

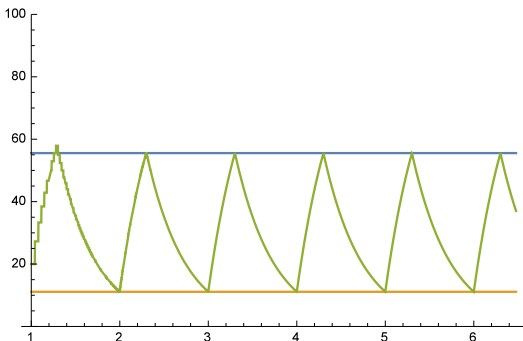


Probability first digit 1 versus street length L .

Streets

Not all data sets satisfy Benford's Law.

- Long street $[1, L]$: $L = 199$ versus $L = 999$.
- Oscillates b/w $1/9$ and $5/9$ with first digit 1.

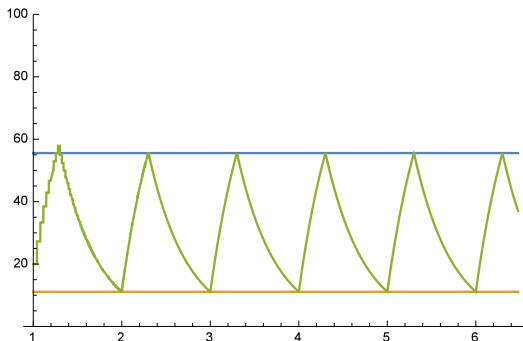


Probability first digit 1 versus $\log(\text{street length } L)$.

Streets

Not all data sets satisfy Benford's Law.

- Long street $[1, L]$: $L = 199$ versus $L = 999$.
- Oscillates b/w $1/9$ and $5/9$ with first digit 1.

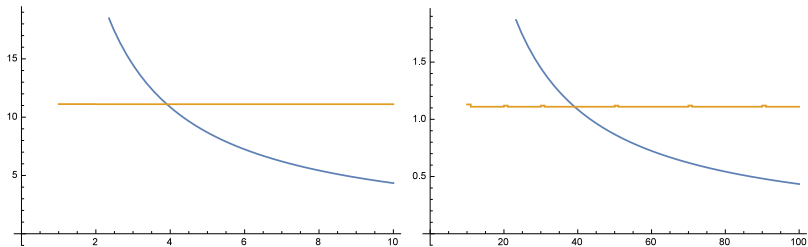


Probability first digit 1 versus $\log(\text{street length } L)$.

What if we have many streets of different lengths?

Amalgamating Streets

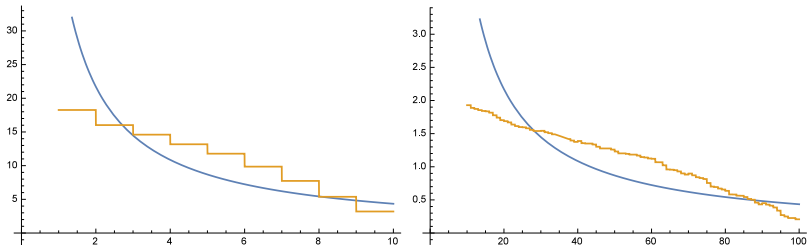
All houses: 100 Streets, each from 1 to 10000.



First digit and first two digits vs Benford.

Amalgamating Streets

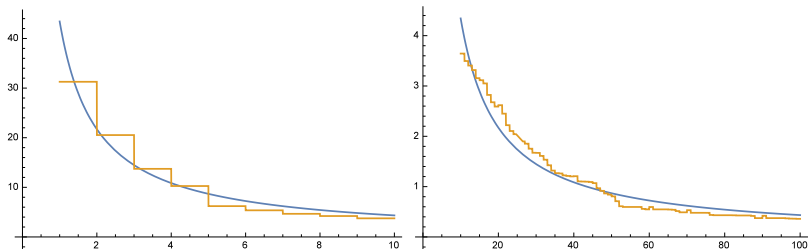
All houses: 100 Streets, each from 1 to rand(10000).



First digit and first two digits vs Benford.

Amalgamating Streets

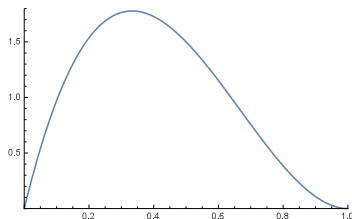
All houses: 100 Streets, each 1 to $\text{rand}(\text{rand}(10000))$.



First digit and first two digits vs Benford.

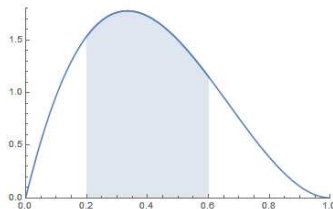
Conclusion: More processes, closer to Benford.

Probability Review



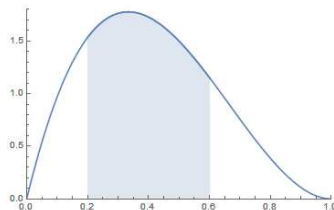
- Let X be random variable with density $p(x)$:
 - ◇ $p(x) \geq 0$; $\int_{-\infty}^{\infty} p(x)dx = 1$;
 - ◇ $\text{Prob}(a \leq X \leq b) = \int_a^b p(x)dx$.

Probability Review



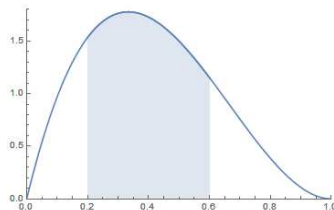
- Let X be random variable with density $p(x)$:
 - ◇ $p(x) \geq 0$; $\int_{-\infty}^{\infty} p(x)dx = 1$;
 - ◇ $\text{Prob}(a \leq X \leq b) = \int_a^b p(x)dx$.
- Mean $\mu = \int_{-\infty}^{\infty} xp(x)dx$.

Probability Review



- Let X be random variable with density $p(x)$:
 - ◇ $p(x) \geq 0$; $\int_{-\infty}^{\infty} p(x)dx = 1$;
 - ◇ $\text{Prob}(a \leq X \leq b) = \int_a^b p(x)dx$.
- Mean $\mu = \int_{-\infty}^{\infty} xp(x)dx$.
- Variance $\sigma^2 = \int_{-\infty}^{\infty} (x - \mu)^2 p(x)dx$.

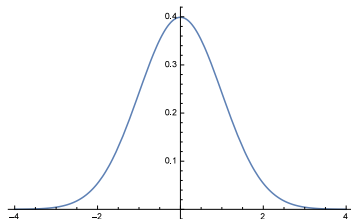
Probability Review



- Let X be random variable with density $p(x)$:
 - ◇ $p(x) \geq 0$; $\int_{-\infty}^{\infty} p(x)dx = 1$;
 - ◇ $\text{Prob}(a \leq X \leq b) = \int_a^b p(x)dx$.
- Mean $\mu = \int_{-\infty}^{\infty} xp(x)dx$.
- Variance $\sigma^2 = \int_{-\infty}^{\infty} (x - \mu)^2 p(x)dx$.
- Independence: knowledge of one random variable gives no knowledge of the other.

Central Limit Theorem

$$\text{Normal } N(\mu, \sigma^2) : p(x) = e^{-(x-\mu)^2/2\sigma^2} / \sqrt{2\pi\sigma^2}.$$



Theorem

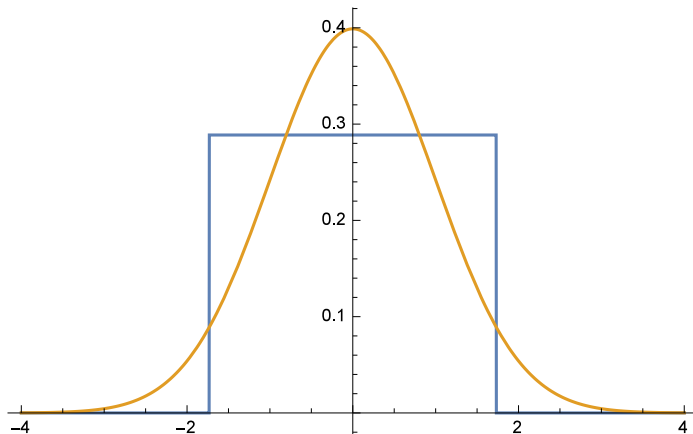
If $X_1, X_2, \dots$ independent, identically distributed random variables (mean μ , variance σ^2 , finite moments) then

$$S_N := \frac{X_1 + \dots + X_N - N\mu}{\sigma\sqrt{N}} \text{ converges to } N(0, 1).$$

Central Limit Theorem: Sums of Uniform Random Variables

$X_i \sim \text{Unif}(-1/2, 1/2)$

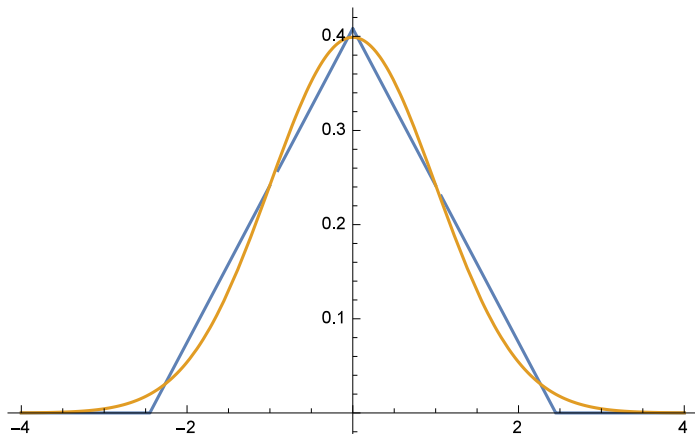
$Y_1 = X_1/\sigma_{X_1}$ vs $N(0, 1)$.



Central Limit Theorem: Sums of Uniform Random Variables

$X_i \sim \text{Unif}(-1/2, 1/2)$

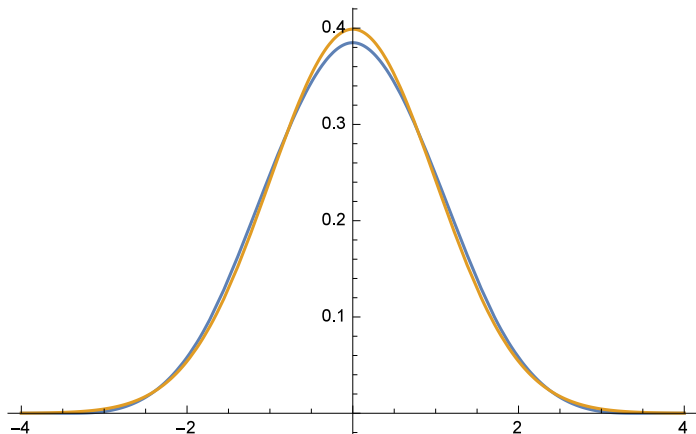
$Y_2 = (X_1 + X_2)/\sigma_{X_1+X_2}$ vs $N(0, 1)$.



Central Limit Theorem: Sums of Uniform Random Variables

$X_i \sim \text{Unif}(-1/2, 1/2)$

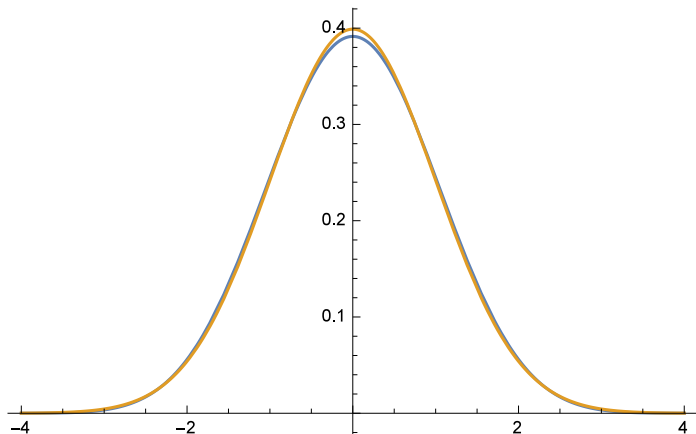
$$Y_4 = (X_1 + X_2 + X_3 + X_4) / \sigma_{X_1 + X_2 + X_3 + X_4} \text{ vs } N(0, 1).$$



Central Limit Theorem: Sums of Uniform Random Variables

$X_i \sim \text{Unif}(-1/2, 1/2)$

$Y_8 = (X_1 + \cdots + X_8) / \sigma_{X_1 + \cdots + X_8}$ vs $N(0, 1)$.



Central Limit Theorem: Sums of Uniform Random Variables

$$X_i \sim \text{Unif}(-1/2, 1/2)$$

Density of $Y_4 = (X_1 + \dots + X_4)/\sigma_{X_1+\dots+X_4}$.

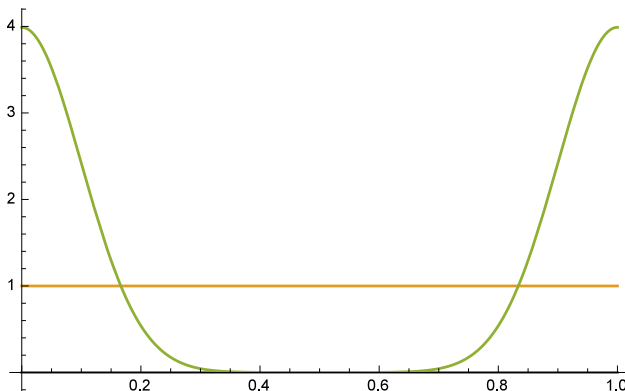
$$\left[\begin{array}{ll} \frac{1}{27} (18 + 9 \sqrt{3} y - \sqrt{3} y^3) & y = 0 \\ \frac{1}{18} (12 - 6 y^2 - \sqrt{3} y^3) & -\sqrt{3} < y < 0 \\ \frac{1}{54} (72 - 36 \sqrt{3} y + 18 y^2 - \sqrt{3} y^3) & \sqrt{3} < y < 2 \sqrt{3} \\ \frac{1}{54} (18 \sqrt{3} y - 18 y^2 + \sqrt{3} y^3) & y = \sqrt{3} \\ \frac{1}{18} (12 - 6 y^2 + \sqrt{3} y^3) & 0 < y < \sqrt{3} \\ \frac{1}{54} (72 + 36 \sqrt{3} y + 18 y^2 + \sqrt{3} y^3) & -2 \sqrt{3} < y \leq -\sqrt{3} \\ 0 & \text{True} \end{array} \right]$$

$$\sqrt{3}$$

(Don't even think of asking to see Y_8 's!)

Normal Distributions Mod 1

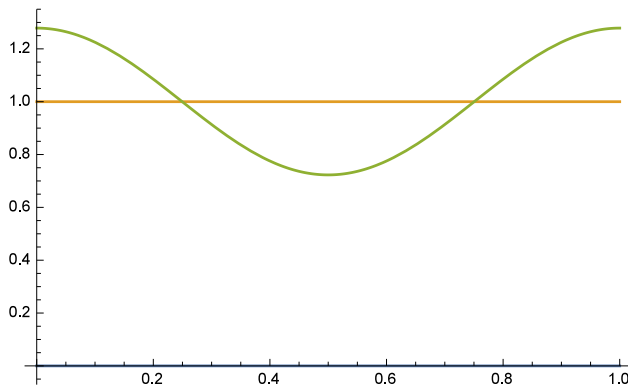
As $\sigma \rightarrow \infty$, $N(0, \sigma^2) \bmod 1 \rightarrow \text{Unif}(0, 1)$.



Variance is .01.

Normal Distributions Mod 1

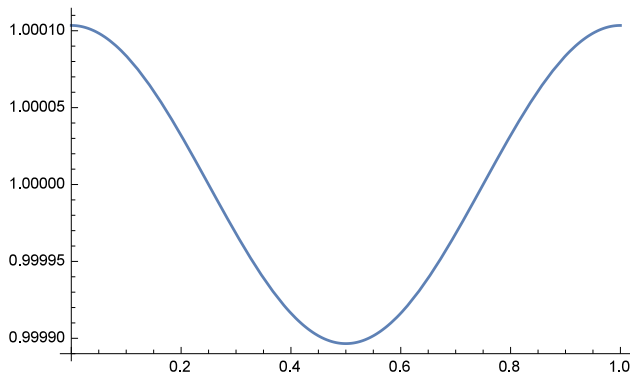
As $\sigma \rightarrow \infty$, $N(0, \sigma^2) \bmod 1 \rightarrow \text{Unif}(0, 1)$.



Variance is .1.

Normal Distributions Mod 1

As $\sigma \rightarrow \infty$, $N(0, \sigma^2) \bmod 1 \rightarrow \text{Unif}(0, 1)$.



Variance is .5.

Products and Benford's Law

Pavlovian Response: See a product, take a logarithm.

Products and Benford's Law

Pavlovian Response: See a product, take a logarithm.

$X_1, X_2, \dots$ nice, $W_N = X_1 \cdot X_2 \cdots X_N$.

Products and Benford's Law

Pavlovian Response: See a product, take a logarithm.

$X_1, X_2, \dots$ nice, $W_N = X_1 \cdot X_2 \cdots X_N$.

$Y_i = \log_{10} X_i$, $V_N := \log_{10} W_N$.

Products and Benford's Law

Pavlovian Response: See a product, take a logarithm.

$X_1, X_2, \dots$ nice, $W_N = X_1 \cdot X_2 \cdots X_N$.

$Y_i = \log_{10} X_i$, $V_N := \log_{10} W_N$.

$$V_N = \log_{10}(X_1 \cdot X_2 \cdots X_N)$$

Products and Benford's Law

Pavlovian Response: See a product, take a logarithm.

$X_1, X_2, \dots$ nice, $W_N = X_1 \cdot X_2 \cdots X_N$.

$Y_i = \log_{10} X_i$, $V_N := \log_{10} W_N$.

$$\begin{aligned} V_N &= \log_{10}(X_1 \cdot X_2 \cdots X_N) \\ &= \log_{10} X_1 + \log_{10} X_2 + \cdots + \log_{10} X_N \end{aligned}$$

Products and Benford's Law

Pavlovian Response: See a product, take a logarithm.

$X_1, X_2, \dots$ nice, $W_N = X_1 \cdot X_2 \cdots X_N$.

$Y_i = \log_{10} X_i$, $V_N := \log_{10} W_N$.

$$\begin{aligned} V_N &= \log_{10}(X_1 \cdot X_2 \cdots X_N) \\ &= \log_{10} X_1 + \log_{10} X_2 + \cdots + \log_{10} X_N \\ &= Y_1 + Y_2 + \cdots + Y_N. \end{aligned}$$

Need distribution of $V_N \bmod 1$, which by CLT becomes uniform,
implying Benfordness!

Applications

Applications for the IRS: Detecting Fraud



A Tale of Two Steve Millers....

Applications for the IRS: Detecting Fraud

Department of the Treasury - Internal Revenue Service
1040 U.S. Individual Income Tax Return 1989
 For the year **1989** or other tax year beginning **1989**, ending **1989** SSN **429-92-9947**

For the year **1989** or other tax year beginning **1989**, ending **1989** SSN **429-92-9947**

WILLIAM J. CLINTON Last name
WILLIAM J. CLINTON First name
WILLIAM J. CLINTON Middle name
WILLIAM J. CLINTON Suffix
WILLIAM J. CLINTON Address
WILLIAM J. CLINTON City, State, and ZIP code
WILLIAM J. CLINTON Country

1800 CENTER Street
WILLIAM J. CLINTON City, State, and ZIP code
WILLIAM J. CLINTON Country

CLINTON President
CLINTON Vice President
CLINTON Secretary
CLINTON Treasurer
CLINTON Director

Do you want \$1 to go to this fund? **Yes** **No**
 If you return, does your spouse want \$1 to go to this fund? **Yes** **No**

Filing Status
 1 ☐ Single
 2 ☒ Married filing joint return (even if only one had income)
 3 ☐ Married filing separate returns (Enter spouse's social security number above and full name here)
 4 ☐ Head of household (Enter qualifying person's name here)
 5 ☐ Qualifying widow(er) with dependent child (Enter spouse's name here)

Exemptions
 6a ☒ Yourself
 6b ☐ Spouse
 6c ☐ Dependents
 6d ☐ Other

Income
 7 Wages, salaries, tips, etc. (Attach Form(s) 1042)
 8 Taxable interest income (Enter Schedule D or other form)
 9 Dividend income (Enter each Schedule D or other form)
 10 Taxable refunds of state and local income taxes (If any, from worksheet on page 11 of instructions)
 11 Annuity received
 12 Business income or loss (Enter Schedule C)
 13 Capital gain or loss (Enter Schedule D)
 14 Other gains or losses (Enter Form 4797)
 15a Total IRA distributions
 15b Total pension and annuities
 16a Total IRA distributions
 16b Total pension and annuities
 17a Total IRA distributions
 17b Total pension and annuities
 18a Total IRA distributions
 18b Total pension and annuities
 19a Total IRA distributions
 19b Total pension and annuities
 20a Total IRA distributions
 20b Total pension and annuities
 21a Total IRA distributions
 21b Total pension and annuities
 22a Total IRA distributions
 22b Total pension and annuities
 23a Total IRA distributions
 23b Total pension and annuities
 24a Total IRA distributions
 24b Total pension and annuities
 25a Total IRA distributions
 25b Total pension and annuities
 26a Total IRA distributions
 26b Total pension and annuities
 27a Total IRA distributions
 27b Total pension and annuities
 28a Total IRA distributions
 28b Total pension and annuities
 29a Total IRA distributions
 29b Total pension and annuities
 30a Total IRA distributions
 30b Total pension and annuities
 31a Total IRA distributions
 31b Total pension and annuities

Adjusted Gross Income
 31a Total IRA distributions
 31b Total pension and annuities

Other
 32a Total IRA distributions
 32b Total pension and annuities

Adjusted Gross Income
 33a Total IRA distributions
 33b Total pension and annuities

Other
 34a Total IRA distributions
 34b Total pension and annuities

Adjusted Gross Income
 35a Total IRA distributions
 35b Total pension and annuities

Other
 36a Total IRA distributions
 36b Total pension and annuities

Adjusted Gross Income
 37a Total IRA distributions
 37b Total pension and annuities

Other
 38a Total IRA distributions
 38b Total pension and annuities

Adjusted Gross Income
 39a Total IRA distributions
 39b Total pension and annuities

Other
 40a Total IRA distributions
 40b Total pension and annuities

Adjusted Gross Income
 41a Total IRA distributions
 41b Total pension and annuities

Other
 42a Total IRA distributions
 42b Total pension and annuities

Adjusted Gross Income
 43a Total IRA distributions
 43b Total pension and annuities

Other
 44a Total IRA distributions
 44b Total pension and annuities

Adjusted Gross Income
 45a Total IRA distributions
 45b Total pension and annuities

Other
 46a Total IRA distributions
 46b Total pension and annuities

Adjusted Gross Income
 47a Total IRA distributions
 47b Total pension and annuities

Other
 48a Total IRA distributions
 48b Total pension and annuities

Adjusted Gross Income
 49a Total IRA distributions
 49b Total pension and annuities

Other
 50a Total IRA distributions
 50b Total pension and annuities

Adjusted Gross Income
 51a Total IRA distributions
 51b Total pension and annuities

Other
 52a Total IRA distributions
 52b Total pension and annuities

Adjusted Gross Income
 53a Total IRA distributions
 53b Total pension and annuities

Other
 54a Total IRA distributions
 54b Total pension and annuities

Adjusted Gross Income
 55a Total IRA distributions
 55b Total pension and annuities

Other
 56a Total IRA distributions
 56b Total pension and annuities

Adjusted Gross Income
 57a Total IRA distributions
 57b Total pension and annuities

Other
 58a Total IRA distributions
 58b Total pension and annuities

Adjusted Gross Income
 59a Total IRA distributions
 59b Total pension and annuities

Other
 60a Total IRA distributions
 60b Total pension and annuities

Adjusted Gross Income
 61a Total IRA distributions
 61b Total pension and annuities

Other
 62a Total IRA distributions
 62b Total pension and annuities

Adjusted Gross Income
 63a Total IRA distributions
 63b Total pension and annuities

Other
 64a Total IRA distributions
 64b Total pension and annuities

Adjusted Gross Income
 65a Total IRA distributions
 65b Total pension and annuities

Other
 66a Total IRA distributions
 66b Total pension and annuities

Adjusted Gross Income
 67a Total IRA distributions
 67b Total pension and annuities

Other
 68a Total IRA distributions
 68b Total pension and annuities

Adjusted Gross Income
 69a Total IRA distributions
 69b Total pension and annuities

Other
 70a Total IRA distributions
 70b Total pension and annuities

Adjusted Gross Income
 71a Total IRA distributions
 71b Total pension and annuities

Other
 72a Total IRA distributions
 72b Total pension and annuities

Adjusted Gross Income
 73a Total IRA distributions
 73b Total pension and annuities

Other
 74a Total IRA distributions
 74b Total pension and annuities

Adjusted Gross Income
 75a Total IRA distributions
 75b Total pension and annuities

Other
 76a Total IRA distributions
 76b Total pension and annuities

Adjusted Gross Income
 77a Total IRA distributions
 77b Total pension and annuities

Other
 78a Total IRA distributions
 78b Total pension and annuities

Adjusted Gross Income
 79a Total IRA distributions
 79b Total pension and annuities

Other
 80a Total IRA distributions
 80b Total pension and annuities

Adjusted Gross Income
 81a Total IRA distributions
 81b Total pension and annuities

Other
 82a Total IRA distributions
 82b Total pension and annuities

Adjusted Gross Income
 83a Total IRA distributions
 83b Total pension and annuities

Other
 84a Total IRA distributions
 84b Total pension and annuities

Adjusted Gross Income
 85a Total IRA distributions
 85b Total pension and annuities

Other
 86a Total IRA distributions
 86b Total pension and annuities

Adjusted Gross Income
 87a Total IRA distributions
 87b Total pension and annuities

Other
 88a Total IRA distributions
 88b Total pension and annuities

Adjusted Gross Income
 89a Total IRA distributions
 89b Total pension and annuities

Other
 90a Total IRA distributions
 90b Total pension and annuities

Adjusted Gross Income
 91a Total IRA distributions
 91b Total pension and annuities

Other
 92a Total IRA distributions
 92b Total pension and annuities

Adjusted Gross Income
 93a Total IRA distributions
 93b Total pension and annuities

Other
 94a Total IRA distributions
 94b Total pension and annuities

Adjusted Gross Income
 95a Total IRA distributions
 95b Total pension and annuities

Other
 96a Total IRA distributions
 96b Total pension and annuities

Adjusted Gross Income
 97a Total IRA distributions
 97b Total pension and annuities

Other
 98a Total IRA distributions
 98b Total pension and annuities

Adjusted Gross Income
 99a Total IRA distributions
 99b Total pension and annuities

Other
 100a Total IRA distributions
 100b Total pension and annuities

Applications for the IRS: Detecting Fraud

93-4670

1040 Department of the Treasury—Internal Revenue Service
U.S. Individual Income Tax Return 1992

OMB Use only—Do not write on this line
OMB No. 1545-0047

For the year 1992, 1-1-1992, or other tax year beginning 1992, ending

Label
WILLIAM J. CLINTON
HILLARY RODHAM CLINTON
THE WHITE HOUSE
1600 PENNSYLVANIA AVENUE N.W.
WASHINGTON, DC 20500

Use the IRS label. Otherwise, place name in type.

For Privacy Act and Paperwork Reduction Act Notice, see page 4.

Presidential Election Campaign
Do you want \$1 to go to this fund? ☒ Yes ☐ No
If you return, does your spouse want \$1 to go to the fund? ☒ Yes ☐ No

Filing Status
Check only one box.
1 ☒ Married (file joint return) (even if only one had income)
2 ☐ Married filing separate return. Enter spouse's SSN above and full name here. If
3 ☐ Single
4 ☐ Head of household. Enter name of dependent child on line 2b or page 2.
5 ☐ Qualifying surviving spouse. Enter name of deceased spouse on line 2b or page 2.

Exemptions
6 a ☒ Yourself. If your spouse or dependent child was a dependent on this return, do not check this box. See instructions for this box on line 2b or page 2.
b ☐ Other.
c ☐ Dependent. If a dependent child, enter child's name on line 2b or page 2.
d ☐ Other. If a dependent child, enter child's name on line 2b or page 2.
e ☐ Other. If a dependent child, enter child's name on line 2b or page 2.
f ☐ Other. If a dependent child, enter child's name on line 2b or page 2.
g ☐ Other. If a dependent child, enter child's name on line 2b or page 2.
h ☐ Other. If a dependent child, enter child's name on line 2b or page 2.
i ☐ Other. If a dependent child, enter child's name on line 2b or page 2.
j ☐ Other. If a dependent child, enter child's name on line 2b or page 2.
k ☐ Other. If a dependent child, enter child's name on line 2b or page 2.
l ☐ Other. If a dependent child, enter child's name on line 2b or page 2.
m ☐ Other. If a dependent child, enter child's name on line 2b or page 2.
n ☐ Other. If a dependent child, enter child's name on line 2b or page 2.
o ☐ Other. If a dependent child, enter child's name on line 2b or page 2.
p ☐ Other. If a dependent child, enter child's name on line 2b or page 2.
q ☐ Other. If a dependent child, enter child's name on line 2b or page 2.
r ☐ Other. If a dependent child, enter child's name on line 2b or page 2.
s ☐ Other. If a dependent child, enter child's name on line 2b or page 2.
t ☐ Other. If a dependent child, enter child's name on line 2b or page 2.
u ☐ Other. If a dependent child, enter child's name on line 2b or page 2.
v ☐ Other. If a dependent child, enter child's name on line 2b or page 2.
w ☐ Other. If a dependent child, enter child's name on line 2b or page 2.
x ☐ Other. If a dependent child, enter child's name on line 2b or page 2.
y ☐ Other. If a dependent child, enter child's name on line 2b or page 2.
z ☐ Other. If a dependent child, enter child's name on line 2b or page 2.

Income
7 ☐ Wages, salaries, tips, etc. Attach Form W-2.
8 ☐ Taxable interest income. Attach Schedule B if over \$400.
9 ☐ Tax-exempt interest income. Do not include on this line.
10 ☐ Dividend income. Attach Schedule B if over \$400.
11 ☐ Taxable refunds, credits, or offsets of state and local income taxes.
12 ☐ Alimony received.
13 ☐ Business income or loss. Attach Schedule C or C-EZ.
14 ☐ Capital gain or loss. Attach Schedule D.
15 ☐ Other gains or losses. Attach Form 4797.
16 ☐ Total IRA distributions.
17 ☐ Total pensions and annuities.
18 ☐ Rents, royalties, partnerships, estates, trusts, etc. Attach Schedule E.
19 ☐ Farm income or loss. Attach Schedule F.
20 ☐ Unemployment compensation.
21 ☐ Social Security benefits.
22 ☐ Other income. 1099-MISC FORMS-IL. EIGHTSON
23 ☐ Add the amounts in the far right column for lines 7 through 22. This is your total income.

Adjustments to income
24 ☐ Your IRA deduction.
25 ☐ Spouse's IRA deduction.
26 ☐ Overhead or self-employment tax.
27 ☐ Self-employed health insurance deduction.
28 ☐ Keogh retirement plan and self-employed SEP deduction.
29 ☐ Penalty on early withdrawal of savings.
30 ☐ Alimony paid. Recipient's SSN: _____

AGI
31 ☐ Subtract line 30 from line 23. This is your adjusted gross income.

Totals
32 ☐ Total income. 237,659.
33 ☐ Total adjustments. 7,269.
34 ☐ Total income after adjustments. 230,390.
35 ☐ Total income after adjustments. 230,390.

1073

Detecting Fraud

Bank Fraud

- Audit of a bank revealed huge spike of numbers starting with

Detecting Fraud

Bank Fraud

- Audit of a bank revealed huge spike of numbers starting with 4

Detecting Fraud

Bank Fraud

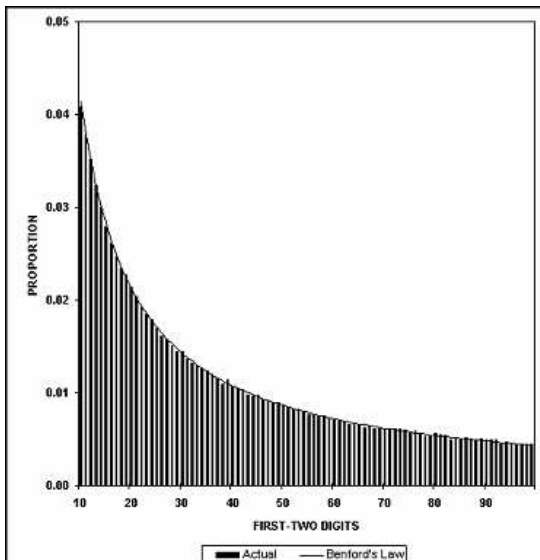
- Audit of a bank revealed huge spike of numbers starting with 48 and 49, most due to one person.

Detecting Fraud

Bank Fraud

- Audit of a bank revealed huge spike of numbers starting with 48 and 49, most due to one person.
- Write-off limit of \$5,000. Officer had friends applying for credit cards, ran up balances just under \$5,000 then he would write the debts off.

Data Integrity: Stream Flow Statistics: 130 years, 457,440 records



Election Fraud: Iran 2009

Numerous questions over Iran's 2009 elections.

Lot of analysis; data moderately suspicious:

- First and second leading digits;
- Last two digits (should almost be uniform);
- Last two digits differing by at least 2.

Warning: enough tests, even if nothing wrong will find a suspicious result (but when all tests are on the boundary...).

The Riemann Zeta Function $\zeta(s)$ and Benford's Law

Riemann Zeta Function (for real part of s greater than 1)

$$\zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s}$$

Riemann Zeta Function (for real part of s greater than 1)

$$\zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s} = \prod_{p \text{ prime}} \left(1 - \frac{1}{p^s}\right)^{-1}, \quad \operatorname{Re}(s) > 1.$$

Riemann Zeta Function (for real part of s greater than 1)

$$\zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s} = \prod_{p \text{ prime}} \left(1 - \frac{1}{p^s}\right)^{-1}, \quad \operatorname{Re}(s) > 1.$$

Geometric Series Formula: $(1 - x)^{-1} = 1 + x + x^2 + \dots$

Unique Factorization: $n = p_1^{r_1} \dots p_m^{r_m}$.

Riemann Zeta Function (for real part of s greater than 1)

$$\zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s} = \prod_{p \text{ prime}} \left(1 - \frac{1}{p^s}\right)^{-1}, \quad \text{Re}(s) > 1.$$

Geometric Series Formula: $(1 - x)^{-1} = 1 + x + x^2 + \dots$

Unique Factorization: $n = p_1^{r_1} \dots p_m^{r_m}$.

$$\begin{aligned} \prod_p \left(1 - \frac{1}{p^s}\right)^{-1} &= \left[1 + \frac{1}{2^s} + \left(\frac{1}{2^s}\right)^2 + \dots\right] \left[1 + \frac{1}{3^s} + \left(\frac{1}{3^s}\right)^2 + \dots\right] \dots \\ &= \sum_n \frac{1}{n^s}. \end{aligned}$$

Riemann Zeta Function (cont)

$$\zeta(s) = \sum_n \frac{1}{n^s} = \prod_p \left(1 - \frac{1}{p^s}\right)^{-1}, \quad \text{Re}(s) > 1$$

$$\pi(x) = \#\{p : p \text{ is prime}, p \leq x\}$$

Properties of $\zeta(s)$ and Primes:

Riemann Zeta Function (cont)

$$\zeta(s) = \sum_n \frac{1}{n^s} = \prod_p \left(1 - \frac{1}{p^s}\right)^{-1}, \quad \operatorname{Re}(s) > 1$$

$$\pi(x) = \#\{p : p \text{ is prime}, p \leq x\}$$

Properties of $\zeta(s)$ and Primes:

- $\lim_{s \rightarrow 1^+} \zeta(s) = \infty, \pi(x) \rightarrow \infty.$

Riemann Zeta Function (cont)

$$\zeta(s) = \sum_n \frac{1}{n^s} = \prod_p \left(1 - \frac{1}{p^s}\right)^{-1}, \quad \operatorname{Re}(s) > 1$$

$$\pi(x) = \#\{p : p \text{ is prime}, p \leq x\}$$

Properties of $\zeta(s)$ and Primes:

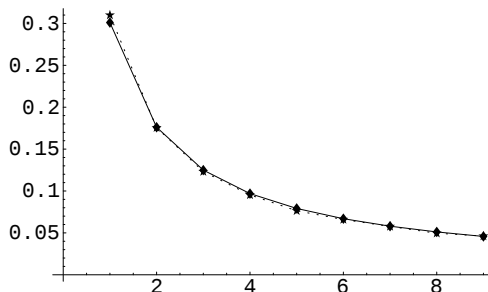
- $\lim_{s \rightarrow 1^+} \zeta(s) = \infty, \pi(x) \rightarrow \infty.$
- $\zeta(2) = \frac{\pi^2}{6}, \pi(x) \rightarrow \infty.$

The Riemann Zeta Function and Benford's Law

$$\left| \zeta \left(\frac{1}{2} + i\frac{k}{4} \right) \right|, k \in \{0, 1, \dots, 65535\}.$$

The Riemann Zeta Function and Benford's Law

$$\left| \zeta \left(\frac{1}{2} + i \frac{k}{4} \right) \right|, k \in \{0, 1, \dots, 65535\}.$$



First digits of $\left| \zeta \left(\frac{1}{2} + i \frac{k}{4} \right) \right|$ versus Benford's law.

The $3x + 1$ Problem and Benford's Law

$3x + 1$ Problem

- Kakutani (conspiracy), Erdős (not ready).
- x odd, $T(x) = \frac{3x+1}{2^k}$, $2^k || 3x + 1$.
- Conjecture: for some $n = n(x)$, $T^n(x) = 1$.

$3x + 1$ Problem

- Kakutani (conspiracy), Erdős (not ready).
- x odd, $T(x) = \frac{3x+1}{2^k}$, $2^k \parallel 3x + 1$.
- Conjecture: for some $n = n(x)$, $T^n(x) = 1$.
- 7

$3x + 1$ Problem

- Kakutani (conspiracy), Erdős (not ready).
- x odd, $T(x) = \frac{3x+1}{2^k}$, $2^k \parallel 3x + 1$.
- Conjecture: for some $n = n(x)$, $T^n(x) = 1$.
- $7 \rightarrow_1 11$

$3x + 1$ Problem

- Kakutani (conspiracy), Erdős (not ready).
- x odd, $T(x) = \frac{3x+1}{2^k}$, $2^k || 3x + 1$.
- Conjecture: for some $n = n(x)$, $T^n(x) = 1$.
- $7 \rightarrow_1 11 \rightarrow_1 17$

$3x + 1$ Problem

- Kakutani (conspiracy), Erdős (not ready).
- x odd, $T(x) = \frac{3x+1}{2^k}$, $2^k \parallel 3x + 1$.
- Conjecture: for some $n = n(x)$, $T^n(x) = 1$.
- $7 \rightarrow_1 11 \rightarrow_1 17 \rightarrow_2 13$

$3x + 1$ Problem

- Kakutani (conspiracy), Erdős (not ready).
- x odd, $T(x) = \frac{3x+1}{2^k}$, $2^k \parallel 3x + 1$.
- Conjecture: for some $n = n(x)$, $T^n(x) = 1$.
- $7 \rightarrow_1 11 \rightarrow_1 17 \rightarrow_2 13 \rightarrow_3 5$

$3x + 1$ Problem

- Kakutani (conspiracy), Erdős (not ready).
- x odd, $T(x) = \frac{3x+1}{2^k}$, $2^k \parallel 3x + 1$.
- Conjecture: for some $n = n(x)$, $T^n(x) = 1$.
- $7 \rightarrow_1 11 \rightarrow_1 17 \rightarrow_2 13 \rightarrow_3 5 \rightarrow_4 1$

$3x + 1$ Problem

- Kakutani (conspiracy), Erdős (not ready).
- x odd, $T(x) = \frac{3x+1}{2^k}$, $2^k \parallel 3x + 1$.
- Conjecture: for some $n = n(x)$, $T^n(x) = 1$.
- $7 \rightarrow_1 11 \rightarrow_1 17 \rightarrow_2 13 \rightarrow_3 5 \rightarrow_4 1 \rightarrow_2 1$,

$3x + 1$ Problem

- Kakutani (conspiracy), Erdős (not ready).
- x odd, $T(x) = \frac{3x+1}{2^k}$, $2^k \parallel 3x + 1$.
- Conjecture: for some $n = n(x)$, $T^n(x) = 1$.
- $7 \rightarrow_1 11 \rightarrow_1 17 \rightarrow_2 13 \rightarrow_3 5 \rightarrow_4 1 \rightarrow_2 1$,
2-path $(1, 1)$, 5-path $(1, 1, 2, 3, 4)$.
 m -path: $(k_1, \dots, k_m)$.

Heuristic Proof of $3x + 1$ Conjecture

$$\begin{aligned}
 a_{n+1} &= T(a_n) \\
 \mathbb{E}[\log a_{n+1}] &\approx \sum_{k=1}^{\infty} \frac{1}{2^k} \log \left(\frac{3a_n}{2^k} \right) \\
 &= \log a_n + \log 3 - \log 2 \sum_{k=1}^{\infty} \frac{k}{2^k} \\
 &= \log a_n + \log \left(\frac{3}{4} \right).
 \end{aligned}$$

Geometric Brownian Motion, drift $\log(3/4) < 1$.

$3x + 1$ and Benford

Theorem (Kontorovich and M–, 2005)

As $m \rightarrow \infty$, $x_m / (3/4)^m x_0$ is Benford.

Theorem (Lagarias-Soundararajan, 2006)

$X \geq 2^N$, for all but at most $c(B)N^{-1/36}X$ initial seeds the distribution of the first N iterates of the $3x + 1$ map are within $2N^{-1/36}$ of the Benford probabilities.

$3x + 1$ Data: random 10,000 digit number, $2^k \parallel 3x + 1$

80,514 iterations ($(4/3)^n = a_0$ predicts 80,319);
 $\chi^2 = 13.5$ (5% 15.5).

Digit	Number	Observed	Benford
1	24251	0.301	0.301
2	14156	0.176	0.176
3	10227	0.127	0.125
4	7931	0.099	0.097
5	6359	0.079	0.079
6	5372	0.067	0.067
7	4476	0.056	0.058
8	4092	0.051	0.051
9	3650	0.045	0.046

$3x + 1$ Data: random 10,000 digit number, $2|3x + 1$

241,344 iterations, $\chi^2 = 11.4$ (5% 15.5).

Digit	Number	Observed	Benford
1	72924	0.302	0.301
2	42357	0.176	0.176
3	30201	0.125	0.125
4	23507	0.097	0.097
5	18928	0.078	0.079
6	16296	0.068	0.067
7	13702	0.057	0.058
8	12356	0.051	0.051
9	11073	0.046	0.046

$5x + 1$ Data: random 10,000 digit number, $2^k \parallel 5x + 1$

27,004 iterations, $\chi^2 = 1.8$ (5% 15.5).

Digit	Number	Observed	Benford
1	8154	0.302	0.301
2	4770	0.177	0.176
3	3405	0.126	0.125
4	2634	0.098	0.097
5	2105	0.078	0.079
6	1787	0.066	0.067
7	1568	0.058	0.058
8	1357	0.050	0.051
9	1224	0.045	0.046

$5x + 1$ Data: random 10,000 digit number, $2|5x + 1$

241,344 iterations, $\chi^2 = 3 \cdot 10^{-4}$ (5% 15.5).

Digit	Number	Observed	Benford
1	72652	0.301	0.301
2	42499	0.176	0.176
3	30153	0.125	0.125
4	23388	0.097	0.097
5	19110	0.079	0.079
6	16159	0.067	0.067
7	13995	0.058	0.058
8	12345	0.051	0.051
9	11043	0.046	0.046

Stick Decomposition

Fixed Proportion Decomposition Process

Decomposition Process

- 1 Consider a stick of length $\mathcal{L}$.

Fixed Proportion Decomposition Process

Decomposition Process

- 1 Consider a stick of length $\mathcal{L}$.
- 2 Uniformly choose a proportion $p \in (0, 1)$.

Fixed Proportion Decomposition Process

Decomposition Process

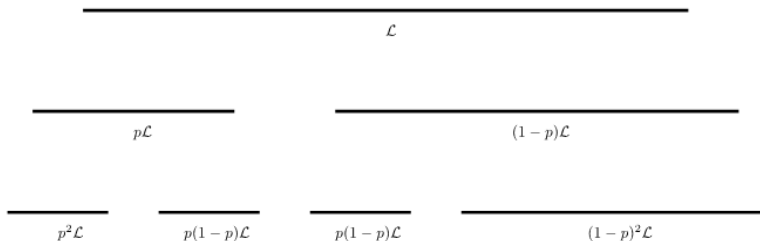
- 1 Consider a stick of length $\mathcal{L}$.
- 2 Uniformly choose a proportion $p \in (0, 1)$.
- 3 Break the stick into two pieces—lengths $p\mathcal{L}$ and $(1 - p)\mathcal{L}$.

Fixed Proportion Decomposition Process

Decomposition Process

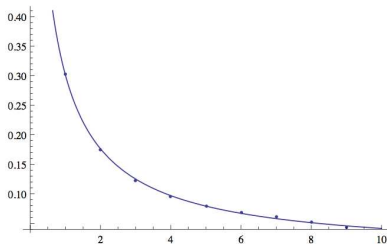
- 1 Consider a stick of length $\mathcal{L}$.
- 2 Uniformly choose a proportion $p \in (0, 1)$.
- 3 Break the stick into two pieces—lengths $p\mathcal{L}$ and $(1 - p)\mathcal{L}$.
- 4 Repeat N times (using the same proportion).

Fixed Proportion Decomposition Process

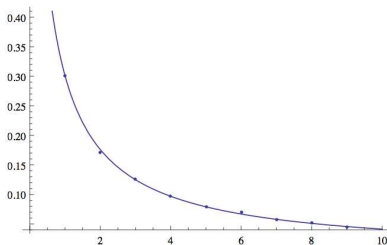


Fixed Proportion Conjecture (Joy Jing '13)

Conjecture: The above decomposition process is Benford as $N \rightarrow \infty$ for any $p \in (0, 1)$, $p \neq \frac{1}{2}$.



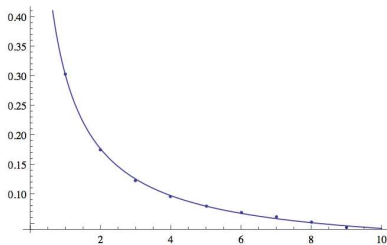
(B) $p = 0.51$ and $N = 10000$.



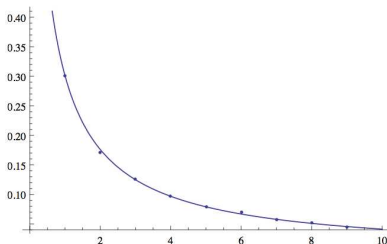
(B) $p = 0.99$ and $N = 50000$. Benford distribution overlaid.

Fixed Proportion Conjecture (Joy Jing '13)

Conjecture: The above decomposition process is Benford as $N \rightarrow \infty$ for any $p \in (0, 1)$, $p \neq \frac{1}{2}$.



(B) $p = 0.51$ and $N = 10000$.



(B) $p = 0.99$ and $N = 50000$. Benford distribution overlaid.

Counterexample (SMALL '13): $p = \frac{1}{11}$, $1 - p = \frac{10}{11}$.

Benford Analysis

At N^{th} level,

- 2^N sticks
- $N + 1$ distinct lengths:

$p^N \left(\frac{1-p}{p} \right)^j$, $j \in \{0, \dots, N\}$, have $\binom{N}{j}$ times.

Benford Analysis

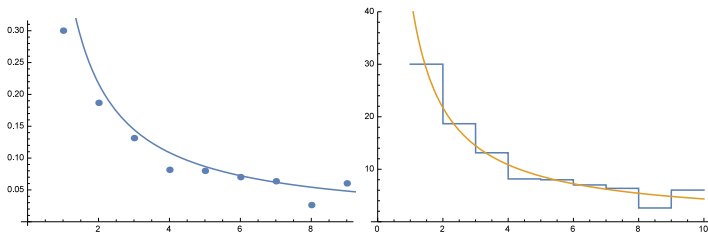
At N^{th} level,

- 2^N sticks
- $N + 1$ distinct lengths:

$$p^N \left(\frac{1-p}{p} \right)^j, \quad j \in \{0, \dots, N\}, \text{ have } \binom{N}{j} \text{ times.}$$

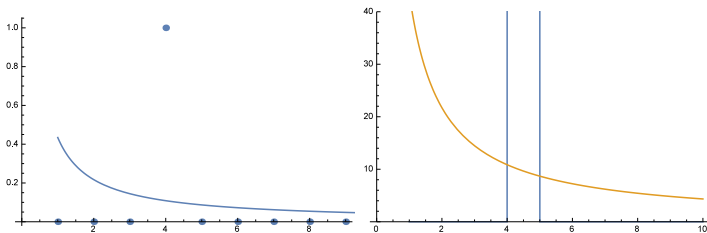
(Weighted) Geometric with ratio $\frac{1-p}{p} = 10^y$;
behavior depends on irrationality of y !

Examples



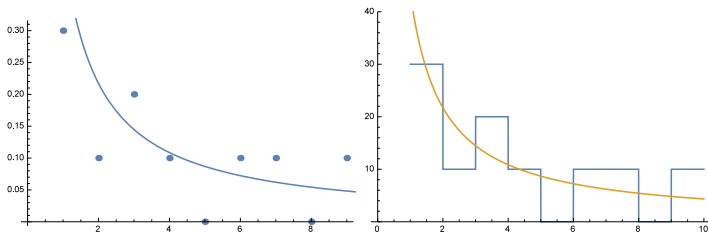
$p = 3/11$, 1000 levels; $y = \log_{10}(8/3) \notin \mathbb{Q}$
(irrational)

Examples



$p = 1/11$, 1000 levels; $y = 1 \in \mathbb{Q}$
(rational)

Examples



$p = 1/(1 + 10^{33/10})$, 1000 levels; $y = 33/10 \in \mathbb{Q}$
(rational)

Conclusions

Current / Future Investigations

- Develop more sophisticated tests for fraud.
- Study digits of other systems.
 - ◊ Break rod of fixed length a variable number of times.
 - ◊ Break rods of variable length a variable number of times.
 - ◊ Break rods of variable length, each piece then breaks with given probability.
 - ◊ Break rods of variable integer length, each piece breaks until is a prime, or a square,

Conclusions and Future Investigations

- See many different systems exhibit Benford behavior.
- Ingredients of proofs (logarithms, equidistribution).
- Applications to fraud detection / data integrity.

References

-  A. K. Adhikari, *Some results on the distribution of the most significant digit*, Sankhyā: The Indian Journal of Statistics, Series B **31** (1969), 413–420.
-  A. K. Adhikari and B. P. Sarkar, *Distribution of most significant digit in certain functions whose arguments are random variables*, Sankhyā: The Indian Journal of Statistics, Series B **30** (1968), 47–58.
-  R. N. Bhattacharya, *Speed of convergence of the n -fold convolution of a probability measure on a compact group*, Z. Wahrscheinlichkeitstheorie verw. Geb. **25** (1972), 1–10.
-  F. Benford, *The law of anomalous numbers*, Proceedings of the American Philosophical Society **78** (1938), 551–572.
-  A. Berger, Leonid A. Bunimovich and T. Hill, *One-dimensional dynamical systems and Benford's Law*, Trans. Amer. Math. Soc. **357** (2005), no. 1, 197–219.

-  A. Berger and T. Hill, *Newton's method obeys Benford's law*, The Amer. Math. Monthly **114** (2007), no. 7, 588-601.
-  A. Berger and T. Hill, *Benford on-line bibliography*, <http://www.benfordonline.net/>.
-  J. Boyle, *An application of Fourier series to the most significant digit problem* Amer. Math. Monthly **101** (1994), 879–886.
-  J. Brown and R. Duncan, *Modulo one uniform distribution of the sequence of logarithms of certain recursive sequences*, Fibonacci Quarterly **8** (1970) 482–486.
-  P. Diaconis, *The distribution of leading digits and uniform distribution mod 1*, Ann. Probab. **5** (1979), 72–81.
-  W. Feller, *An Introduction to Probability Theory and its Applications, Vol. II*, second edition, John Wiley & Sons, Inc., 1971.



R. W. Hamming, *On the distribution of numbers*, Bell Syst. Tech. J. **49** (1970), 1609-1625.



T. Hill, *The first-digit phenomenon*, American Scientist **86** (1996), 358–363.



T. Hill, *A statistical derivation of the significant-digit law*, Statistical Science **10** (1996), 354–363.



P. J. Holewijn, *On the uniform distribuiton of sequences of random variables*, Z. Wahrscheinlichkeitstheorie verw. Geb. **14** (1969), 89–92.



W. Hurlimann, *Benford's Law from 1881 to 2006: a bibliography*, <http://arxiv.org/abs/math/0607168>.



D. Jang, J. U. Kang, A. Kruckman, J. Kudo and S. J. Miller, *Chains of distributions, hierarchical Bayesian models and Benford's Law*, Journal of Algebra, Number Theory: Advances and Applications, volume 1, number 1 (March 2009), 37–60.

-  E. Janvresse and T. de la Rue, *From uniform distribution to Benford's law*, Journal of Applied Probability **41** (2004) no. 4, 1203–1210.
-  A. Kontorovich and S. J. Miller, *Benford's Law, Values of L-functions and the $3x + 1$ Problem*, Acta Arith. **120** (2005), 269–297.
-  D. Knuth, *The Art of Computer Programming, Volume 2: Seminumerical Algorithms*, Addison-Wesley, third edition, 1997.
-  J. Lagarias and K. Soundararajan, *Benford's Law for the $3x + 1$ Function*, J. London Math. Soc. (2) **74** (2006), no. 2, 289–303.
-  S. Lang, *Undergraduate Analysis*, 2nd edition, Springer-Verlag, New York, 1997.



P. Levy, *L'addition des variables aléatoires définies sur une circonférence*, Bull. de la S. M. F. **67** (1939), 1–41.



E. Ley, *On the peculiar distribution of the U.S. Stock Indices Digits*, The American Statistician **50** (1996), no. 4, 311–313.



R. M. Loynes, *Some results in the probabilistic theory of asymptotic uniform distributions modulo 1*, Z. Wahrscheinlichkeitstheorie verw. Geb. **26** (1973), 33–41.



S. J. Miller, *Theory and Applications of Benford's Law*, Princeton University Press, in press, expected publication date 2015.



S. J. Miller and M. Nigrini, *The Modulo 1 Central Limit Theorem and Benford's Law for Products*, International Journal of Algebra **2** (2008), no. 3, 119–130.



S. J. Miller and M. Nigrini, *Order Statistics and Benford's law*, International Journal of Mathematics and Mathematical Sciences, Volume 2008 (2008), Article ID 382948, 19 pages.



S. J. Miller and R. Takloo-Bighash, *An Invitation to Modern Number Theory*, Princeton University Press, Princeton, NJ, 2006.



S. Newcomb, *Note on the frequency of use of the different digits in natural numbers*, Amer. J. Math. **4** (1881), 39-40.



M. Nigrini, *Digital Analysis and the Reduction of Auditor Litigation Risk*. Pages 69–81 in *Proceedings of the 1996 Deloitte & Touche / University of Kansas Symposium on Auditing Problems*, ed. M. Ettredge, University of Kansas, Lawrence, KS, 1996.



M. Nigrini, *The Use of Benford's Law as an Aid in Analytical Procedures*, Auditing: A Journal of Practice & Theory, **16** (1997), no. 2, 52–67.



M. Nigrini and S. J. Miller, *Benford's Law applied to hydrology data – results and relevance to other geophysical data*, Mathematical Geology **39** (2007), no. 5, 469–490.

-  M. Nigrini and S. J. Miller, *Data diagnostics using second order tests of Benford's Law*, Auditing: A Journal of Practice and Theory **28** (2009), no. 2, 305–324.
-  R. Pinkham, *On the Distribution of First Significant Digits*, The Annals of Mathematical Statistics **32**, no. 4 (1961), 1223–1230.
-  R. A. Raimi, *The first digit problem*, Amer. Math. Monthly **83** (1976), no. 7, 521–538.
-  H. Robbins, *On the equidistribution of sums of independent random variables*, Proc. Amer. Math. Soc. **4** (1953), 786–799.
-  H. Sakamoto, *On the distributions of the product and the quotient of the independent and uniformly distributed random variables*, Tôhoku Math. J. **49** (1943), 243–260.
-  P. Schatte, *On sums modulo 2π of independent random variables*, Math. Nachr. **110** (1983), 243–261.



P. Schatte, *On the asymptotic uniform distribution of sums reduced mod 1*, Math. Nachr. **115** (1984), 275–281.



P. Schatte, *On the asymptotic logarithmic distribution of the floating-point mantissas of sums*, Math. Nachr. **127** (1986), 7–20.



E. Stein and R. Shakarchi, *Fourier Analysis: An Introduction*, Princeton University Press, 2003.



M. D. Springer and W. E. Thompson, *The distribution of products of independent random variables*, SIAM J. Appl. Math. **14** (1966) 511–526.



K. Stromberg, *Probabilities on a compact group*, Trans. Amer. Math. Soc. **94** (1960), 295–309.



P. R. Turner, *The distribution of leading significant digits*, IMA J. Numer. Anal. **2** (1982), no. 4, 407–412.