Benford's Law, Values of L-Functions and the $3x+1$ Problem, or: Why the IRS should care about Number Theory

Steven J Miller
Mount Holyoke / Smith / Williams Colleges

sjm1@williams.edu

http://web.williams.edu/Mathematics/sjmiller/public_html/index.htm

Five Colleges Number Theory Seminar
January 31, 2012

Introduction

Interesting Question

For a nice data set, such as the Fibonacci numbers, stock prices, street addresses of Five Colleges Number Theory Seminar participants, ..., what percent of the leading digits are 1?

Plausible answers:

Introduction

Interesting Question

For a nice data set, such as the Fibonacci numbers, stock prices, street addresses of Five Colleges Number Theory Seminar participants, ..., what percent of the leading digits are 1?

Plausible answers: 10%

Introduction

Interesting Question

For a nice data set, such as the Fibonacci numbers, stock prices, street addresses of Five Colleges Number Theory Seminar participants, ..., what percent of the leading digits are 1?

Plausible answers: 10%, 11%

Introduction

Interesting Question

For a nice data set, such as the Fibonacci numbers, stock prices, street addresses of Five Colleges Number Theory Seminar participants, ..., what percent of the leading digits are 1?

Plausible answers: 10%, 11%, about 30%.

Summary

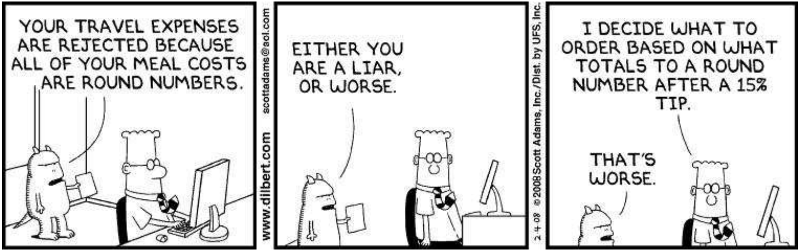
- State Benford’s Law.
- Discuss examples and applications.
- Sketch proofs.
- Describe open problems.

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.



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: 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.

- Not all data sets satisfy Benford's Law.

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.

- Not all data sets satisfy Benford's Law.
  ◇ Long street $[1, L]$: $L = 199$ versus $L = 999$.

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.

- Not all data sets satisfy Benford's Law.
  - ◇ Long street $[1, L]$: $L = 199$ versus $L = 999$.
  - ◇ Oscillates between $1/9$ and $5/9$ with first digit 1.

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.

- Not all data sets satisfy Benford's Law.
  - ◇ Long street $[1, L]$: $L = 199$ versus $L = 999$.
  - ◇ Oscillates between $1/9$ and $5/9$ with first digit 1.
  - ◇ Many streets of different sizes: close to Benford.

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

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

Mantissas (or Significands)

Mantissa: $x = M_{10}(x) \cdot 10^k$, k integer.

$M_{10}(x) = M_{10}(\widetilde{x})$ if and only if x and $\widetilde{x}$ have the same leading digits.

Key observation: $\log_{10}(x) = \log_{10}(\widetilde{x}) \bmod 1$ if and only if x and $\widetilde{x}$ have the same leading digits. Thus often study $y = \log_{10} x$.

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 $b-a$:

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

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 $b-a$:

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

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

Equidistribution and Benford's Law

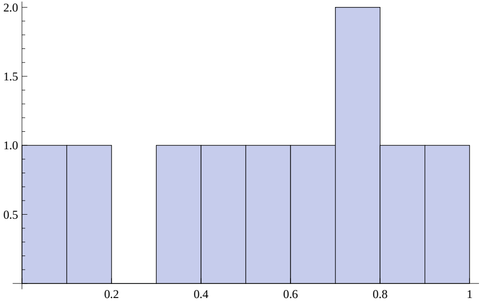
Equidistribution

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

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

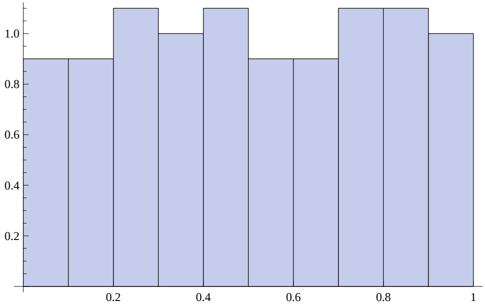
- Thm: $\beta \notin \mathbb{Q}$, $n\beta$ is equidistributed mod 1.
- Examples: $\log_{10} 2, \log_{10}\left(\frac{1+\sqrt{5}}{2}\right) \notin \mathbb{Q}$.

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



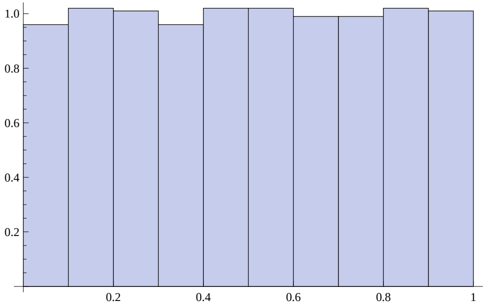
$n\sqrt{\pi} \bmod 1$ for $n \le 10$

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



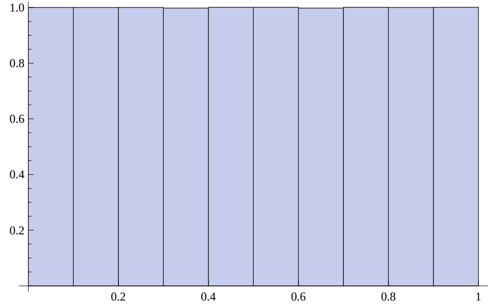
$n\sqrt{\pi}$ mod 1 for $n \leq 100$

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



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

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



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

Logarithms and Benford's Law

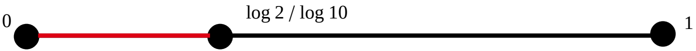
Fundamental Equivalence

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

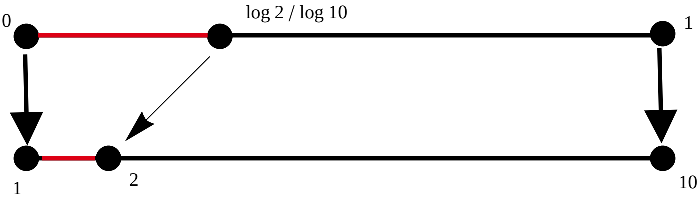
Data set $\{x_i\}$ is Benford base B if $\{y_i\}$ is equidistributed mod 1, where $y_i = \log_B x_i$.



Logarithms and Benford’s Law

Fundamental Equivalence

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



Examples

- 2^n is Benford base 10 as $\log_{10} 2 \not\in \mathbb{Q}$.

Examples

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

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

  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

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

- Fibonacci numbers are Benford base 10.

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

  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.

  $a_{n+1} = 2a_n$

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

  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)

1	1024	1048576
2	2048	2097152
4	4096	4194304
8	8192	8388608
16	16384	16777216
32	32768	33554432
64	65536	67108864
128	131072	134217728
256	262144	268435456
512	524288	536870912

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

Digits of 2^n

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

1	1024	1048576
2	2048	2097152
4	4096	4194304
8	8192	8388608
16	16384	16777216
32	32768	33554432
64	65536	67108864
128	131072	134217728
256	262144	268435456
512	524288	536870912

digit	#	Obs Prob	Benf Prob
1	18	.300	.301
2	12	.200	.176
3	6	.100	.125
4	6	.100	.097
5	6	.100	.079
6	4	.067	.067
7	2	.033	.058
8	5	.083	.051
9	1	.017	.046

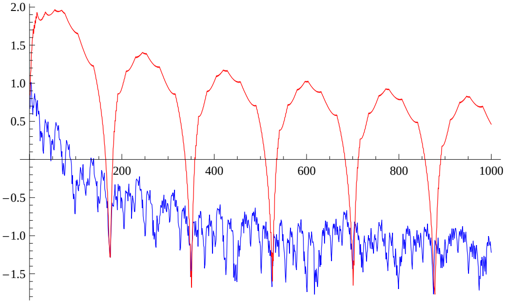
Logarithms and Benford's Law

χ^2 values for α^n, $1 \le n \le N$ (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

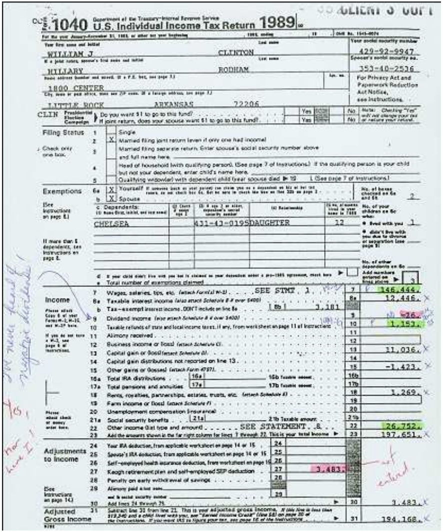
Logarithms and Benford's Law: Base 10

$\log_{10}(\chi^2)$ vs N for π^n (red) and e^n (blue), $n \in \{1, \dots, N\}$. Note $\pi^{175} \approx 1.0028 \cdot 10^{87}$, (5% and 8 d.f., $\log_{10}(\chi^2) \approx .44$).



Applications

Applications for the IRS: Detecting Fraud



Form 1040 Department of the Treasury—Internal Revenue Service
U.S. Individual Income Tax Return 1989

CLIENT'S COPY

Your first name and initial: WILLIAM J	Last name: CLINTON	Your social security number: 429-92-9947
If a joint return, spouse's first name and initial: HILLARY	Last name: RODHAM	Spouse's social security no.: 353-40-2536
Home address: 1800 CENTER		For Privacy Act and Paperwork Reduction Act Notice, see Instructions.
City: LITTLE ROCK	ARKANSAS 72206	

CLIN Presidential Election Campaign ▶ Do you want $1 to go to this fund? Yes No
If joint return, does your spouse want $1 to go to this fund? Yes No

Filling Status — Check only one box.
1 ☐ Single
2 ☒ Married filing joint return (even if only one had income)
3 ☐ Married filing separate return. Enter spouse's social security number above and full name here.
4 ☐ Head of household (with qualifying person). (See page 7 of Instructions.) If the qualifying person is your child but not your dependent, enter child's name here.
5 ☐ Qualifying widow(er) with dependent child (year spouse died ▶ 19). (See page 7 of Instructions.)

Exemptions
6a ☒ Yourself
b ☒ Spouse
No. of boxes checked on 6a and 6b: 2

c. Dependents: (1) Name	(2)	(3) Dependent's social security number	(4) Relationship	(5) No. of months lived in your home in 1989
CHELSEA		431-43-0195	DAUGHTER	12

No. of your children on 6c who: lived with you 1
e Total number of exemptions claimed — Add numbers entered on lines above ▶ 3

Income

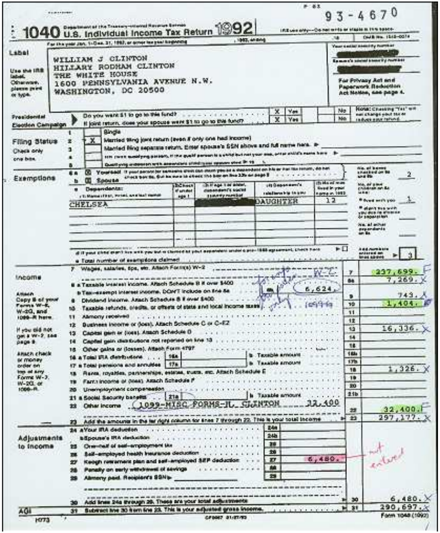
Line		Amount
7	Wages, salaries, tips, etc. (attach Form(s) W-2) SEE STMT 1	146,444.
8a	Taxable interest income	12,446.
8b	Tax-exempt interest income. DON'T include on line 8a — 8b 1,101	
9	Dividend income	-26.
10	Taxable refunds of state and local income taxes, if any, from worksheet on page 11 of Instructions	1,153.
11	Alimony received	
12	Business income or (loss) (attach Schedule C)	
13	Capital gain or (loss) (attach Schedule D)	11,036.
14	Capital gain distributions not reported on line 13	
15	Other gains or (losses) (attach Form 4797)	-1,423.
16a	Total IRA distributions — 16b Taxable amount	
17a	Total pensions and annuities — 17b Taxable amount	
18	Rents, royalties, partnerships, estates, trusts, etc. (attach Schedule E)	1,269.
19	Farm income or (loss) (attach Schedule F)	
20	Unemployment compensation (insurance)	
21a	Social security benefits — 21b Taxable amount	
22	Other income (list type and amount) SEE STATEMENT 8	26,752.
23	Add the amounts shown in the far right column for lines 7 through 22. This is your total income ▶	197,651.

Adjustments to Income

Line			Amount
24	Your IRA deduction, from applicable worksheet on page 14 or 15		
25	Spouse's IRA deduction, from applicable worksheet on page 14 or 15		
26	Self-employed health insurance deduction, from worksheet on page 15		
27	Keogh retirement plan and self-employed SEP deduction	3,483.	
28	Penalty on early withdrawal of savings		
29	Alimony paid		
30	Add lines 24 through 29. ▶		3,483.
31	Subtract line 30 from line 23. This is your adjusted gross income. ▶		194,168.

Yes,

Applications for the IRS: Detecting Fraud



P 83 93-4670

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

For the year Jan. 1–Dec. 31, 1992, or other tax year beginning , 1992, ending

Label

Use the IRS label. Otherwise, please print or type.

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

Your social security number [redacted]

Spouse's social security number [redacted]

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

Presidential Election Campaign

Do you want $1 to go to this fund? X Yes No

If joint return, does your spouse want $1 to go to this fund? X Yes No

Filing Status Check only one box.

1 ☐ Single

2 ☒ Married filing joint return (even if only one had income)

3 ☐ Married filing separate return. Enter spouse's SSN above and full name here. ▶

4 ☐

5 ☐

Exemptions

6a ☒ Yourself

b ☒ Spouse

c Dependents:

(1) Name (first, initial, and last name)	(2) Check if under age 1	(3) If age 1 or older, dependent's social security number	(4) Dependent's relationship to you	(5) No. of months lived in your home in 1992
CHELSEA		[redacted]	DAUGHTER	12

No. of boxes checked on 6a and 6b: 2

No. of your children on 6c who: lived with you: 1

e Total number of exemptions claimed

Add numbers entered on lines above ▶ 3

Income

Attach Copy B of your Forms W-2, W-2G, and 1099-R here.

If you did not get a W-2, see page 9.

Attach check or money order on top of any Forms W-2, W-2G, or 1099-R.

7	Wages, salaries, tips, etc. Attach Form(s) W-2		7	237,699. F
8a	Taxable interest income. Attach Schedule B if over $400		8a	7,269. X
b	Tax-exempt interest income. DON'T include on line 8a	6,624.		
9	Dividend income. Attach Schedule B if over $400		9	743. X
10	Taxable refunds, credits, or offsets of state and local income taxes		10	1,404.
11	Alimony received		11	
12	Business income or (loss). Attach Schedule C or C-EZ		12	
13	Capital gain or (loss). Attach Schedule D		13	16,336. X
14	Capital gain distributions not reported on line 13		14	
15	Other gains or (losses). Attach Form 4797		15	
16a	Total IRA distributions 16a	b Taxable amount	16b	
17a	Total pensions and annuities 17a	b Taxable amount	17b	
18	Rents, royalties, partnerships, estates, trusts, etc. Attach Schedule E		18	1,326. X
19	Farm income or (loss). Attach Schedule F		19	
20	Unemployment compensation		20	
21a	Social Security benefits 21a	b Taxable amount	21b	
22	Other income 1099-MISC FORMS-H. CLINTON 32,400		22	32,400. F
23	Add the amounts in the far right column for lines 7 through 22. This is your total income ▶		23	297,177. X

W-2 to total only 1099

Adjustments to Income

24a	Your IRA deduction	24a	
b	Spouse's IRA deduction	24b	
25	One-half of self-employment tax	25	
26	Self-employed health insurance deduction	26	
27	Keogh retirement plan and self-employed SEP deduction	27	6,480.
28	Penalty on early withdrawal of savings	28	
29	Alimony paid. Recipient's SSN ▶	29	

not entered

30	Add lines 24a through 29. These are your total adjustments ▶	30	6,480. X
31	**AGI** Subtract line 30 from line 23. This is your adjusted gross income. ▶	31	290,697. X

H073 CF0067 01/27/93 Form 1040 (1992)

Detecting Fraud

Bank Fraud

- Audit of a bank revealed huge spike of numbers

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

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.

Benford Good Processes

Poisson Summation and Benford’s Law: Definitions

- Feller, Pinkham (often exact processes)

Poisson Summation and Benford’s Law: Definitions

- Feller, Pinkham (often exact processes)
- data $Y_{T,B} = \log_B \overrightarrow{X}_T$ (discrete/continuous):

$$\mathbb{P}(A) \ = \ \lim_{T\to\infty} \frac{\#\{n \in A : n \le T\}}{T}$$

Poisson Summation and Benford's Law: Definitions

- Feller, Pinkham (often exact processes)
- data $Y_{T,B} = \log_B \overrightarrow{X}_T$ (discrete/continuous):

$$\mathbb{P}(A) \ = \ \lim_{T\to\infty} \frac{\#\{n \in A : n \le T\}}{T}$$

- Poisson Summation Formula: f nice:

$$\sum_{\ell=-\infty}^{\infty} f(\ell) \ = \ \sum_{\ell=-\infty}^{\infty} \widehat{f}(\ell),$$

$$\text{Fourier transform } \widehat{f}(\xi) \ = \ \int_{-\infty}^{\infty} f(x) e^{-2\pi i x \xi} dx.$$

Benford Good Process

X_T is Benford Good if there is a nice f st

$$\mathrm{CDF}_{\vec{Y}_{T,B}}(y) = \int_{-\infty}^{y} \frac{1}{T} f\left(\frac{t}{T}\right) dt + E_T(y) := G_T(y)$$

and monotonically increasing h ($h(|T|) \to \infty$):

Benford Good Process

X_T is Benford Good if there is a nice f st

$$\mathrm{CDF}_{\vec{Y}_{T,B}}(y) = \int_{-\infty}^{y} \frac{1}{T} f\left(\frac{t}{T}\right) dt + E_T(y) := G_T(y)$$

and monotonically increasing h ($h(|T|) \to \infty$):

- Small tails: $G_T(\infty) - G_T(Th(T)) = o(1)$, $G_T(-Th(T)) - G_T(-\infty) = 0(1)$.

Benford Good Process

X_T is **Benford Good** if there is a nice f st

$$\mathrm{CDF}_{\vec{Y}_{T,B}}(y) \ = \ \int_{-\infty}^{y} \frac{1}{T} f\left(\frac{t}{T}\right) dt + E_T(y) \ := \ G_T(y)$$

and monotonically increasing h ($h(|T|) \to \infty$):

- Small tails: $G_T(\infty) - G_T(Th(T)) = o(1)$, $G_T(-Th(T)) - G_T(-\infty) = 0(1)$.
- Decay of the Fourier Transform: $\sum_{\ell \neq 0} \left| \frac{\widehat{f}(T\ell)}{\ell} \right| = o(1)$.

Benford Good Process

X_T is Benford Good if there is a nice f st

$$\mathrm{CDF}_{\vec{Y}_{T,B}}(y) \ = \ \int_{-\infty}^{y} \frac{1}{T} f\left(\frac{t}{T}\right) dt + E_T(y) \ := \ G_T(y)$$

and monotonically increasing h ($h(|T|) \to \infty$):

- Small tails: $G_T(\infty) - G_T(Th(T)) = o(1)$, $G_T(-Th(T)) - G_T(-\infty) = 0(1)$.
- Decay of the Fourier Transform: $\sum_{\ell \neq 0} \left| \frac{\widehat{f}(T\ell)}{\ell} \right| = o(1)$.
- Small translated error: $\mathcal{E}(a,b,T)) = \sum_{|\ell| \leq Th(T)} [E_T(b+\ell) - E_T(a+\ell)] = o(1)$.

Main Theorem

Theorem (Kontorovich and M–, 2005)

X_T converging to X as $T \to \infty$ (think spreading Gaussian). If X_T is Benford good, then X is Benford.

Main Theorem

Theorem (Kontorovich and M–, 2005)

X_T converging to X as $T \to \infty$ (think spreading Gaussian). If X_T is Benford good, then X is Benford.

- Examples
  - ◇ L-functions
  - ◇ characteristic polynomials (RMT)
  - ◇ $3x + 1$ problem
  - ◇ geometric Brownian motion.

Sketch of the proof

- Structure Theorem:
  - ◇ main term is something nice spreading out
  - ◇ apply Poisson summation

Sketch of the proof

- Structure Theorem:
  - ◇ main term is something nice spreading out
  - ◇ apply Poisson summation
- Control translated errors:
  - ◇ hardest step
  - ◇ techniques problem specific

Sketch of the proof (continued)

$$\sum_{\ell=-\infty}^{\infty} \mathbb{P}\left(a+\ell \le \overrightarrow{Y}_{T,B} \le b+\ell\right)$$

Sketch of the proof (continued)

$$\begin{aligned} &\sum_{\ell=-\infty}^{\infty} \mathbb{P}\left(a+\ell \leq \overrightarrow{Y}_{T,B} \leq b+\ell\right) \\ = &\sum_{|\ell| \leq Th(T)} \left[G_T(b+\ell) - G_T(a+\ell)\right] + o(1) \end{aligned}$$

Sketch of the proof (continued)

$$\begin{aligned}
&\sum_{\ell=-\infty}^{\infty} \mathbb{P}\left(a+\ell \leq \overrightarrow{Y}_{T,B} \leq b+\ell\right) \\
=\ & \sum_{|\ell| \leq Th(T)} \left[G_T(b+\ell) - G_T(a+\ell)\right] + o(1) \\
=\ & \int_a^b \sum_{|\ell| \leq Th(T)} \frac{1}{T} f\left(\frac{t}{T}\right) dt + \mathcal{E}(a,b,T) + o(1)
\end{aligned}$$

Sketch of the proof (continued)

$$\begin{aligned}
&\sum_{\ell=-\infty}^{\infty} \mathbb{P}\left(a+\ell \le \overrightarrow{Y}_{T,B} \le b+\ell\right) \\
=\ & \sum_{|\ell| \le Th(T)} \left[G_T(b+\ell) - G_T(a+\ell)\right] + o(1) \\
=\ & \int_a^b \sum_{|\ell| \le Th(T)} \frac{1}{T} f\left(\frac{t}{T}\right) dt + \mathcal{E}(a,b,T) + o(1) \\
=\ & \widehat{f}(0) \cdot (b-a) + \sum_{\ell \ne 0} \widehat{f}(T\ell) \frac{e^{2\pi i b\ell} - e^{2\pi i a\ell}}{2\pi i \ell} + o(1).
\end{aligned}$$

L-functions and Random Matrix Theory

Good L-functions ($\zeta(s)$, full level cusp form)

L-function is good if:

- Euler product:

$$L(s,f) = \sum_{n=1}^{\infty} \frac{a_f(n)}{n^s} = \prod_{p \text{ prime}} \prod_{j=1}^{d} \left(1 - \alpha_{f,j}(p) p^{-s}\right)^{-1}.$$

- meromorphic continuation to $\mathbb{C}$, of finite order, at most finitely many poles (all on the line $\mathrm{Re}(s) = 1$).
- Functional equation: $\omega \in \mathbb{R}$, $G(s)$ prod Γ-fns:

$$e^{\mathrm{i}\omega} G(s) L(s,f) = e^{-\mathrm{i}\omega} \overline{G(1-\overline{s}) L(1-\overline{s})}.$$

Good L-functions

- For some $\aleph > 0$, $c \in \mathbb{C}$, $x \geq 2$:

$$\sum_{p \leq x} \frac{|a_f(p)|^2}{p} = \aleph \log\log x + c + O\left(\frac{1}{\log x}\right).$$

- The $\alpha_{f,j}(p)$ are (Ramanujan-Petersson) tempered: $|\alpha_{f,j}(p)| \leq 1$.
- $N(\sigma, T) = \#\{\rho : L(\rho, f) = 0, \mathrm{Re}(\rho) \geq \sigma, \mathrm{Im}(\rho) \in [0, T]\}$. $\exists \beta > 0$

$$N(\sigma, T) = O\left(T^{1-\beta\left(\sigma-\frac{1}{2}\right)} \log T\right).$$

Log-Normal Law (Hejhal, Laurinčikas, Selberg)

Log-Normal Law

$$\frac{\mu\left(\{t\in[T,2T]:\log\left|L\left(\sigma+it,f\right)\right|\in[a,b]\}\right)}{T}=$$

$$\frac{1}{\sqrt{\psi(\sigma,T)}}\int_a^b e^{-\pi u^2/\psi(\sigma,T)}du+\mathrm{Error}$$

$$\psi(\sigma,T)=\aleph\log\left[\min\left(\log T,\frac{1}{\sigma-\frac{1}{2}}\right)\right]+O(1)$$

$$\frac{1}{2}\leq\sigma\leq\frac{1}{2}+\frac{1}{\log^\delta T},\quad\delta\in(0,1).$$

Values of L-functions and Benford's Law

Theorem (Kontorovich and M–, 2005)

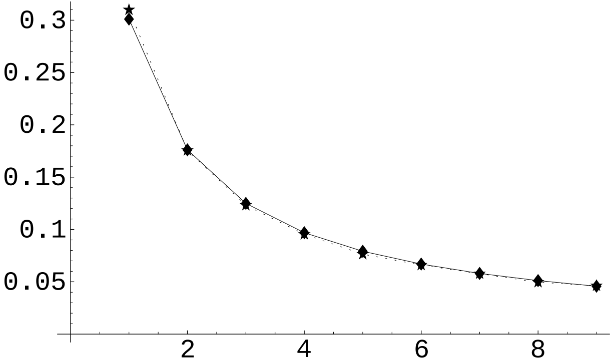
$L(s, f)$ a good L-function, as $T \to \infty$, $L(\sigma_T + it, f)$ is Benford.

Ingredients

- Approximate $\log L(\sigma_T + it, f)$ with $\sum_{n \le x} \frac{c(n)\Lambda(n)}{\log n} \frac{1}{n^{\sigma_T + it}}$.
- study moments $\int_T^{2T} |\cdot|$, $k \le \log^{1-\delta} T$.
- Montgomery-Vaughan: $\int_T^{2T} \sum a_n n^{-it} \overline{\sum b_m m^{-it}} dt = H \sum a_n \overline{b}_n + O(1)\sqrt{\sum n|a_n|^2 \sum n|b_n|^2}$.

Riemann Zeta Function

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



Random Matrices: Preliminaries

- $N \times N$ unitary matrices U (Haar measure):

$$p_N(U) = \frac{1}{(2\pi)^N N!} \prod_{1 \le j < m \le N} \left| e^{\mathrm{i}\theta_j} - e^{\mathrm{i}\theta_m} \right|.$$

- characteristic polynomial:

$$Z(U,\theta) = \det(I - Ue^{-\mathrm{i}\theta}) = \prod \left(1 - e^{\mathrm{i}(\theta_n - \theta)} \right).$$

- $\rho_N(x)$ the probability density for $\log|Z(U,\theta)|$:

$$\widetilde{\rho}_N(x) \ = \ \sqrt{Q_2(N)}\, \rho_N(\sqrt{Q_2(N)}\, x),$$

variance $Q_2(N) \sim (\log N)/2$.

Random Matrices and Benford's Law

Theorem (Kontorovich and M–, 2005)

As $N \to \infty$, the distribution of digits of the absolute values of the characteristic polynomials of $N \times N$ unitary matrices (with respect to Haar measure) converges to the Benford probabilities.

- Key Ingredient: Keating-Snaith:

$$\widetilde{\rho}_N(x)dx \ = \ \frac{1}{\sqrt{2\pi}}e^{-x^2/2}dx+O\left((\log N)^{-3/2}dx\right).$$

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

$3x + 1$ Problem

- Kakutani (conspiracy), Erdös (not ready).
- x odd, $T(x) = \frac{3x+1}{2^k}$, $2^k || 3x + 1$.

$3x+1$ **Problem**

- Kakutani (conspiracy), Erdös (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), Erdös (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

$3x+1$ Problem

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

$3x+1$ **Problem**

- Kakutani (conspiracy), Erdös (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), Erdös (not ready).
- x odd, $T(x) = \frac{3x+1}{2^k}$, $2^k || 3x + 1$.
- Conjecture: for some $n = n(x)$, $T^n(x) = 1$.
- $7 \to_1 11 \to_1 17 \to_2 13$

$3x + 1$ Problem

- Kakutani (conspiracy), Erdös (not ready).
- x odd, $T(x) = \frac{3x+1}{2^k}$, $2^k || 3x + 1$.
- Conjecture: for some $n = n(x)$, $T^n(x) = 1$.
- $7 \to_1 11 \to_1 17 \to_2 13 \to_3 5$

$3x + 1$ Problem

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

$3x + 1$ Problem

- Kakutani (conspiracy), Erdös (not ready).
- x odd, $T(x) = \frac{3x+1}{2^k}$, $2^k || 3x + 1$.
- Conjecture: for some $n = n(x)$, $T^n(x) = 1$.
- $7 \to_1 11 \to_1 17 \to_2 13 \to_3 5 \to_4 1 \to_2 1$,

$3x + 1$ **Problem**

- Kakutani (conspiracy), Erdös (not ready).
- x odd, $T(x) = \frac{3x+1}{2^k}$, $2^k || 3x + 1$.
- Conjecture: for some $n = n(x)$, $T^n(x) = 1$.
- $7 \to_1 11 \to_1 17 \to_2 13 \to_3 5 \to_4 1 \to_2 1$,
  2-path $(1, 1)$, 5-path $(1, 1, 2, 3, 4)$.
  m-path: $(k_1, \dots, k_m)$.

Heuristic Proof of $3x + 1$ Conjecture

$$a_{n+1} \;=\; T(a_n)$$

Heuristic Proof of $3x+1$ Conjecture

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

Heuristic Proof of $3x+1$ Conjecture

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

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

Structure Theorem: Sinai, Kontorovich-Sinai

$$\mathbb{P}(A) = \lim_{N\to\infty} \frac{\#\{n\le N : n\equiv 1,5 \bmod 6, n\in A\}}{\#\{n\le N : n\equiv 1,5 \bmod 6\}}.$$

Structure Theorem: Sinai, Kontorovich-Sinai

$\mathbb{P}(A) = \lim_{N\to\infty} \frac{\#\{n \le N : n \equiv 1,5 \bmod 6, n \in A\}}{\#\{n \le N : n \equiv 1,5 \bmod 6\}}$.

$(k_1, \dots, k_m)$: two full arithm progressions:

$6 \cdot 2^{k_1+\cdots+k_m} p + q$.

Structure Theorem: Sinai, Kontorovich-Sinai

$$\mathbb{P}(A) = \lim_{N\to\infty} \frac{\#\{n\le N : n\equiv 1,5 \bmod 6, n\in A\}}{\#\{n\le N : n\equiv 1,5 \bmod 6\}}.$$

$(k_1,\dots,k_m)$: two full arithm progressions:

$6\cdot 2^{k_1+\cdots+k_m}p + q$.

Theorem (Sinai, Kontorovich-Sinai)

k_i-values are i.i.d.r.v. (geometric, $1/2$):

$$\mathbb{P}\left(\frac{\log_2\left[\frac{x_m}{\left(\frac{3}{4}\right)^m x_0}\right]}{\sqrt{2m}} \le a\right) = \mathbb{P}\left(\frac{S_m - 2m}{\sqrt{2m}} \le a\right)$$

Structure Theorem: Sinai, Kontorovich-Sinai

$$\mathbb{P}(A) = \lim_{N\to\infty} \frac{\#\{n \le N : n \equiv 1,5 \bmod 6, n \in A\}}{\#\{n \le N : n \equiv 1,5 \bmod 6\}}.$$

$(k_1, \dots, k_m)$: two full arithm progressions: $6 \cdot 2^{k_1+\cdots+k_m} p + q$.

Theorem (Sinai, Kontorovich-Sinai)

k_i-values are i.i.d.r.v. (geometric, $1/2$):

$$\mathbb{P}\left(\frac{\log_2\left[\frac{x_m}{\left(\frac{3}{4}\right)^m x_0}\right]}{(\log_2 B)\sqrt{2m}} \le a\right) = \mathbb{P}\left(\frac{S_m - 2m}{(\log_2 B)\sqrt{2m}} \le a\right)$$

Structure Theorem: Sinai, Kontorovich-Sinai

$$\mathbb{P}(A) = \lim_{N\to\infty} \frac{\#\{n\le N : n\equiv 1,5 \bmod 6, n\in A\}}{\#\{n\le N : n\equiv 1,5 \bmod 6\}}.$$

$(k_1, \dots, k_m)$: two full arithm progressions: $6\cdot 2^{k_1+\cdots+k_m}p+q$.

Theorem (Sinai, Kontorovich-Sinai)

k_i-values are i.i.d.r.v. (geometric, $1/2$):

$$\mathbb{P}\left(\frac{\log_B\left[\frac{x_m}{\left(\frac{3}{4}\right)^m x_0}\right]}{\sqrt{2m}} \le a\right) = \mathbb{P}\left(\frac{\frac{(S_m-2m)}{\log_2 B}}{\sqrt{2m}} \le a\right)$$

$3x+1$ and Benford

Theorem (Kontorovich and M–, 2005)

As $m \to \infty$, $x_m/(3/4)^m x_0$ 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.

Sketch of the proof

- Failed Proof: lattices, bad errors.

Sketch of the proof

- Failed Proof: lattices, bad errors.
- CLT: $(S_m - 2m)/\sqrt{2m} \to N(0,1)$:

$$\mathbb{P}\left(S_m - 2m = k\right) = \frac{\eta(k/\sqrt{m})}{\sqrt{m}} + O\left(\frac{1}{g(m)\sqrt{m}}\right).$$

Sketch of the proof

- Failed Proof: lattices, bad errors.
- CLT: $(S_m - 2m)/\sqrt{2m} \to N(0,1)$:
$$\mathbb{P}\left(S_m - 2m = k\right) = \frac{\eta(k/\sqrt{m})}{\sqrt{m}} + O\left(\frac{1}{g(m)\sqrt{m}}\right).$$
- Quantified Equidistribution:
$I_\ell = \{\ell M, \dots, (\ell+1)M - 1\}$, $M = m^c$, $c < 1/2$

Sketch of the proof

- Failed Proof: lattices, bad errors.
- CLT: $(S_m - 2m)/\sqrt{2m} \to N(0,1)$:
$$\mathbb{P}\left(S_m - 2m = k\right) = \frac{\eta(k/\sqrt{m})}{\sqrt{m}} + O\left(\frac{1}{g(m)\sqrt{m}}\right).$$
- Quantified Equidistribution:
$I_\ell = \{\ell M, \dots, (\ell+1)M-1\}$, $M = m^c$, $c < 1/2$
$k_1, k_2 \in I_\ell$: $\left|\eta\left(\frac{k_1}{\sqrt{m}}\right) - \eta\left(\frac{k_2}{\sqrt{m}}\right)\right|$ small

Sketch of the proof

- Failed Proof: lattices, bad errors.
- CLT: $(S_m - 2m)/\sqrt{2m} \to N(0,1)$:
$$\mathbb{P}\left(S_m - 2m = k\right) = \frac{\eta(k/\sqrt{m})}{\sqrt{m}} + O\left(\frac{1}{g(m)\sqrt{m}}\right).$$
- Quantified Equidistribution:
$I_\ell = \{\ell M, \dots, (\ell+1)M - 1\}$, $M = m^c$, $c < 1/2$
$k_1, k_2 \in I_\ell$: $\left|\eta\left(\frac{k_1}{\sqrt{m}}\right) - \eta\left(\frac{k_2}{\sqrt{m}}\right)\right|$ small
$C = \log_B 2$ of irrationality type $\kappa < \infty$:
$\#\{k \in I_\ell : \overline{kC} \in [a,b]\} = M(b-a) + O(M^{1+\epsilon-1/\kappa})$.

Irrationality Type

Irrationality type

α has irrationality type κ if κ is the supremum of all γ with

$$\underline{\lim}_{q\to\infty} q^{\gamma+1} \min_p \left| \alpha - \frac{p}{q} \right| = 0.$$

- Algebraic irrationals: type 1 (Roth's Thm).
- Theory of Linear Forms: $\log_B 2$ of finite type.

Linear Forms

Theorem (Baker)

$\alpha_1, \dots, \alpha_n$ *algebraic numbers height* $A_j \geq 4$,
$\beta_1, \dots, \beta_n \in \mathbb{Q}$ *with height at most* $B \geq 4$,

$$\Lambda = \beta_1 \log \alpha_1 + \cdots + \beta_n \log \alpha_n.$$

If $\Lambda \neq 0$ *then* $|\Lambda| > B^{-C\Omega \log \Omega'}$, *with*
$d = [\mathbb{Q}(\alpha_i, \beta_j) : \mathbb{Q}]$, $C = (16nd)^{200n}$,
$\Omega = \prod_j \log A_j$, $\Omega' = \Omega / \log A_n$.

Gives $\log_{10} 2$ of finite type, with $\kappa < 1.2 \cdot 10^{602}$:

$$\left|\log_{10} 2 - p/q\right| = \left|q \log 2 - p \log 10\right| / q \log 10.$$

Quantified Equidistribution

Theorem (Erdös-Turan)

$$D_N = \frac{\sup_{[a,b]} |N(b-a) - \#\{n \le N : x_n \in [a,b]\}|}{N}$$

There is a C such that for all m:

$$D_N \le C \cdot \left(\frac{1}{m} + \sum_{h=1}^{m} \frac{1}{h} \left| \frac{1}{N} \sum_{n=1}^{N} e^{2\pi i h x_n} \right| \right)$$

Proof of Erdös-Turan

Consider special case $x_n = n\alpha$, $\alpha \notin \mathbb{Q}$.

- Exponential sum $\le \frac{1}{|\sin(\pi h\alpha)|} \le \frac{1}{2||h\alpha||}$.
- Must control $\sum_{h=1}^{m} \frac{1}{h||h\alpha||}$, see irrationality type enter.
- type κ, $\sum_{h=1}^{m} \frac{1}{h||h\alpha||} = O\left(m^{\kappa-1+\epsilon}\right)$, take $m = \lfloor N^{1/\kappa} \rfloor$.

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

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

Conclusions

Conclusions and Future Investigations

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