

Benford's law, or: Why the IRS cares about number theory!

Steven J Miller (Williams College)

`sjm1@williams.edu`

`http://www.williams.edu/go/math/sjmilller/`

Williams College, June 23, 2010

Summary

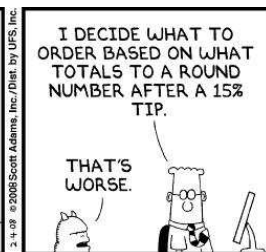
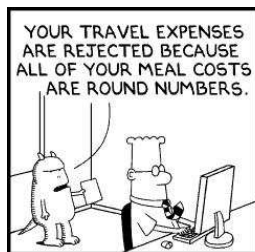
- State Benford's Law.
- Discuss examples and applications.
- Sketch proofs.
- Describe open problems.

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

Applications

- analyzing round-off errors
- determining the optimal way to store numbers
- detecting tax and image fraud, and data integrity

General Theory

Mantissas

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

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

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 $y = \log_{10} x$.

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

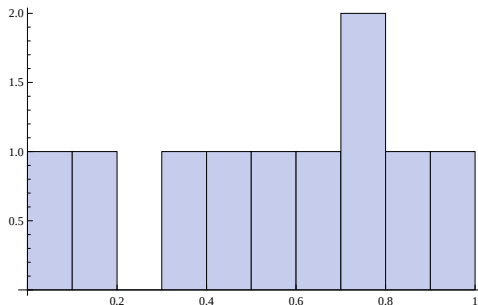
Equidistribution

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

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

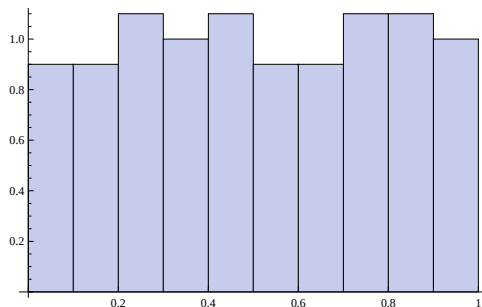
- Thm: $\beta \notin \mathbb{Q}$, $n\beta$ is equidistributed mod 1.
- Examples: $\log_{10} 2, \log_{10} \left(\frac{1+\sqrt{5}}{2}\right) \notin \mathbb{Q}$.
Proof: if rational: $2 = 10^{p/q}$.
Thus $2^q = 10^p$ or $2^{q-p} = 5^p$, impossible.

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



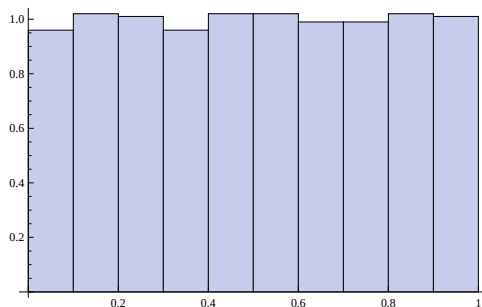
$n\sqrt{\pi} \bmod 1$ for $n \leq 10$

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



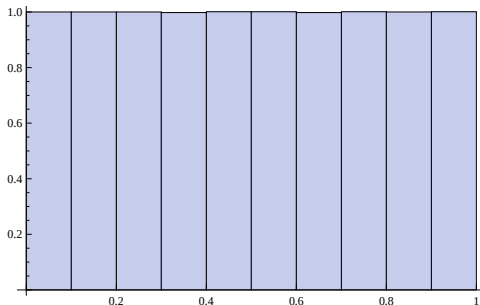
$n\sqrt{\pi} \bmod 1$ for $n \leq 100$

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

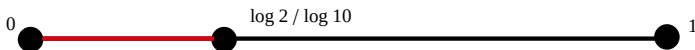
Fundamental Equivalence

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

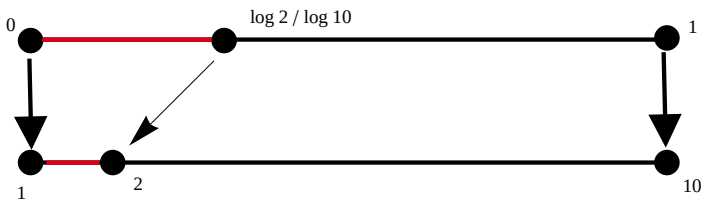
Data set $\{x_i\}$ is Benford base B if $\{y_i\}$ is equidistributed mod 1, where $y_i = \log_B x_i$.



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 $\log_{10} 2 \notin \mathbb{Q}$.

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

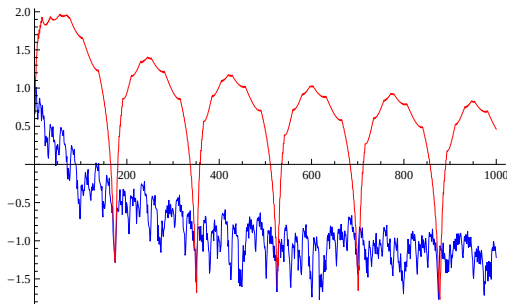
Logarithms and Benford's Law

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

$\log(\chi^2)$ vs N for π^n (red) and e^n (blue),
 $n \in \{1, \dots, N\}$. Note $\pi^{175} \approx 1.0028 \cdot 10^{87}$, (5%,
 $\log(\chi^2) \approx 2.74$).



Applications

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** (January 1, 1989, or other tax year beginning 1989, ending 1989) (OMB No. 1545-0047)

Print your name and last initial **WILLIAM J. CLINTON** Last name
 If a joint return, enter the first name and last initial **WILLIAM J. CLINTON** Last name
HILARY **RODHAM** First name and last initial (If a P.E. filer, see page 1)

1800 CENTER
 11711 E. ROCK ARKANSAS 72206

CLIN President
 Company 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 ☒

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 (only 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 (your spouse died in 1981). (See page 7 of instructions.)

Exemptions 6a ☒ Yourself? If answer "No," see page 8 of instructions. If "Yes," see page 8 of instructions. If "Yes," see page 8 of instructions.
 b ☒ Spouse
 c Dependents: (1) Name first, last, and middle (2) Social security number (3) Relationship (4) Date of birth (5) No. of your children as for whom

CHITSEA 451-43-0195 DAUGHTER 12

If more than 6 dependents, see instructions on page 8.

If you child starts the year you file is claimed as your dependent enter a pre-1985 agreement, check box ☐

Total number of exemptions claimed 3

Income 7 Wages, salaries, tips, etc. (attach Form(s) if over \$100,000) **SEE STATE 3** 84 **346,444**
 8a Taxable interest income (attach Schedule D if over \$100) **181** 84 **12,446**
 8b Tax-exempt interest income (attach Schedule D if over \$100) **181** 84 **1,266**
 9 Dividend income (attach Schedule D if over \$100) **181** 10 **11,593**
 10 Taxable refunds of state and local income taxes. If any, from worksheet on page 11 of instructions **12**
 11 Alimony received **12**
 12 Business income or loss (attach Schedule C) **12**
 13 Capital gain or loss (attach Schedule D) **14** **11,036**
 14 Capital gain distributions, not reported on line 13 **15**
 15 Other gains or losses (attach Form 4797) **15** **-1,423**
 16a Total IRA distributions **16a** 16b Rollover amount **16b**
 17a Total pensions and annuities **17a** 17b Rollover amount **17b**
 18 Rents, royalties, partnerships, estates, trusts, etc. (attach Schedule E) **18**
 19 Farm income or loss (attach Schedule F) **19**
 20 Unemployment compensation (attach Schedule G) **20**
 21a Social security benefits **21a** 21b Taxable amount **21b**
 22 Other income (state type and amount) **22** **SRR SPARTANMENT, 8** 23 **26,752**
 23 Add the amounts from lines 7 through 22. This is your total income **23** **397,651**
 24 Your IRA deduction, from applicable worksheet on page 14 or 15 **24**
 25 Spouse's IRA deduction, from applicable worksheet on page 14 or 15 **25**
 26 Self-employed health insurance deduction, from worksheet on page 15 **26**
 27 Rough retirement plan and self-employed SEP deduction **27**
 28 Penalty on early withdrawal of savings **28**
 29 Alimony paid a last year **29**
 30 Subtract line 29 from line 23. This is your adjusted gross income. If you file a line item 112, 140 and a child (over age 17), see "Family Income Credit" (line 60) on page 27 of the instructions. If you must file a return for your spouse, see page 10 of the instructions. **30** **398,483**
 Adjusted Gross Income **31** **398,483**
 Gross Income **31** **398,483**

See instructions on page 143.

Adjusted Gross Income **31** **398,483**
 Gross Income **31** **398,483**

Handwritten notes:
 "I have filed negative dividends"
 "I have filed negative dividends"
 "I have filed negative dividends"

Applications for the IRS: Detecting Fraud

10773

93-4670

1040 Department of the Treasury—Internal Revenue Service

Individual Income Tax Return 1992

OMB No. 1545-0047 Use or copy after 12-31-92 or after the date in this notice.

For the year 1992, or 1993, or other tax year beginning

1992

1-8 (OMB No. 1545-0047)

Label

Use the 1040 label. Otherwise, please print in type.

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

Your social security number

Researcher's copy of your name

For Privacy Act and
 Paperwork Reduction
 Act Notice, see page 6.

Presidential Election Campaign

Do you have \$1 to go to the fund?

If you return, does your spouse have \$1 to go to the fund?

X Yes

X Yes

No

No

Notwithstanding "Yes" or
 "No," check box if you are
 claiming a tax credit.

Filing Status

Check only one box.

1 ☒ Single
 2 ☐ Married filing joint return (attach if only one had income)
 3 ☐ Married filing separate return. Each spouse's SEI above and full name here. If
 4 ☐ each spouse's separate return, the spouse's name is printed on the back of your return, and the name here is

5 ☐ Qualifying widow(er) with dependent child. (See instructions on page 2.)

Exemptions

6 ☐ Yourself. If you are claiming an exemption for yourself, check box 1 or 2 on the 32a page.

7 ☐ Other person.

8 ☐ Dependent.

9 ☐ Other person. (See instructions on page 2.)

10 ☐ Other person.

11 ☐ Other person.

12 ☐ Other person.

13 ☐ Other person.

14 ☐ Other person.

15 ☐ Other person.

16 ☐ Other person.

17 ☐ Other person.

18 ☐ Other person.

19 ☐ Other person.

20 ☐ Other person.

21 ☐ Other person.

22 ☐ Other person.

23 ☐ Other person.

24 ☐ Other person.

25 ☐ Other person.

26 ☐ Other person.

27 ☐ Other person.

28 ☐ Other person.

29 ☐ Other person.

30 ☐ Other person.

31 ☐ Other person.

32 ☐ Other person.

33 ☐ Other person.

34 ☐ Other person.

35 ☐ Other person.

36 ☐ Other person.

37 ☐ Other person.

38 ☐ Other person.

39 ☐ Other person.

40 ☐ Other person.

41 ☐ Other person.

42 ☐ Other person.

43 ☐ Other person.

44 ☐ Other person.

45 ☐ Other person.

46 ☐ Other person.

47 ☐ Other person.

48 ☐ Other person.

49 ☐ Other person.

50 ☐ Other person.

51 ☐ Other person.

52 ☐ Other person.

53 ☐ Other person.

54 ☐ Other person.

55 ☐ Other person.

56 ☐ Other person.

57 ☐ Other person.

58 ☐ Other person.

59 ☐ Other person.

60 ☐ Other person.

See instructions on page 2.

See instructions on page 2.

See instructions on page 2.

See instructions on page 2.

See instructions

Detecting Fraud

Bank Fraud

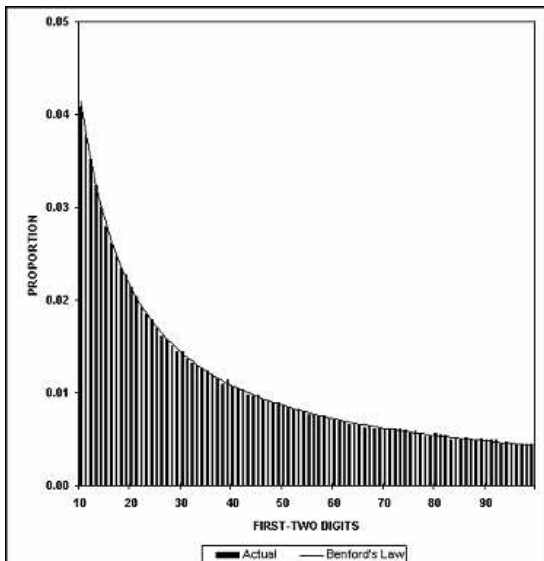
- Audit of a bank revealed huge spike of numbers starting with 48 and 49, most due to one person.

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 protests/complaints 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...).

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 $Y_{T,B} = \log_B \vec{X}_T$ (discrete/continuous):

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

Poisson Summation and Benford's Law: Definitions

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

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

- 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

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

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

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

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

Sketch of the proof (continued)

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

Sketch of the proof (continued)

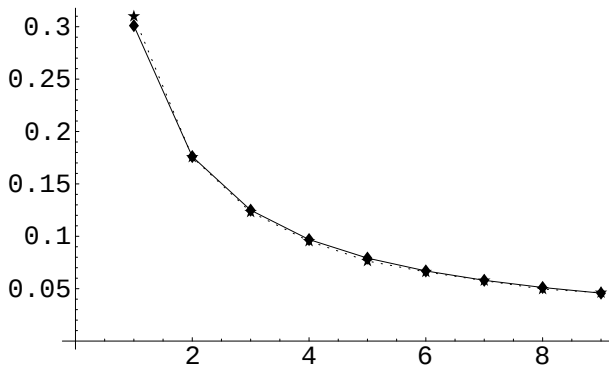
$$\begin{aligned}
 & \sum_{\ell=-\infty}^{\infty} \mathbb{P} \left(a + \ell \leq \overrightarrow{Y}_{T,B} \leq b + \ell \right) \\
 &= \sum_{|\ell| \leq Th(T)} [G_T(b + \ell) - G_T(a + \ell)] + o(1) \\
 &= \int_a^b \sum_{|\ell| \leq Th(T)} \frac{1}{T} f \left(\frac{t}{T} \right) dt + \mathcal{E}(a, b,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

Sketch of the proof (continued)

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

Riemann Zeta Function

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



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 $n = n(x)$, $T^n(x) = 1$.
- $7 \rightarrow_1 11 \rightarrow_1 17 \rightarrow_2 13 \rightarrow_3 5 \rightarrow_4 1 \rightarrow_2 1$,
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 $\log(3/4) < 1$.

Structure Theorem: Sinai, Kontorovich-Sinai

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

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

Theorem (Sinai, Kontorovich-Sinai)

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

Structure Theorem: Sinai, Kontorovich-Sinai

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

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

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

Theorem (Sinai, Kontorovich-Sinai)

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

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

Structure Theorem: Sinai, Kontorovich-Sinai

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

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

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

Theorem (Sinai, Kontorovich-Sinai)

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

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

Structure Theorem: Sinai, Kontorovich-Sinai

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

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

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

Theorem (Sinai, Kontorovich-Sinai)

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

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

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.

Sketch of the proof

- Failed Proof: lattices, bad errors.

- CLT: $(S_m - 2m)/\sqrt{2m} \rightarrow N(0, 1)$:

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

- Quantified Equidistribution: $I_\ell = \{\ell M, \dots, (\ell + 1)M - 1\}$,
 $M = m^c$, $c < 1/2$

$$k_1, k_2 \in I_\ell: \left| \eta\left(\frac{k_1}{\sqrt{m}}\right) - \eta\left(\frac{k_2}{\sqrt{m}}\right) \right| \text{ small}$$

$C = \log_B 2$ of irrationality type $\kappa < \infty$:

$$\#\{k \in I_\ell : \overline{kC} \in [a, b]\} = M(b - a) + O(M^{1+\epsilon-1/\kappa}).$$

Sketch of the proof: Irrationality Type

Irrationality type

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

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

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

Sketch of the proof: 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 $B \geq 4$,

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

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 $\kappa < 1.2 \cdot 10^{602}$:

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

Sketch of the proof : Quantified Equidistribution

Theorem (Erdős-Turan)

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

There is a C such that for all m :

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

Sketch of the proof : Proof of Erdős-Turan

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

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

$3x + 1$ 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

Conclusions

Conclusions and Future Investigations

- See many different systems exhibit Benford behavior.
- Ingredients of proofs (logarithms, equidistribution).
- Applications to fraud detection / data integrity.
- **Future work:**
 - ◇ Study digits of other systems.
 - ◇ Develop more sophisticated tests for fraud

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.
-  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, *When the Cramér-Rao Inequality provides no information*, Communications in Information and Systems **7** (2007), no. 3, 265–272.



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.

Products of Random Variables

Preliminaries

- $X_1 \cdots X_n \Leftrightarrow Y_1 + \cdots + Y_n \bmod 1$, $Y_i = \log_B X_i$
- Density Y_i is g_i , density $Y_i + Y_j$ is

$$(g_i * g_j)(y) = \int_0^1 g_i(t)g_j(y - t)dt.$$

- $h_n = g_1 * \cdots * g_n$, $\widehat{g}(\xi) = \widehat{g}_1(\xi) \cdots \widehat{g}_n(\xi)$.

Modulo 1 Central Limit Theorem

Theorem (M– and Nigrini 2007)

$\{Y_m\}$ independent continuous random variables on $[0, 1)$ (not necc. i.i.d.), densities $\{g_m\}$.

$Y_1 + \cdots + Y_M \bmod 1$ converges to the uniform distribution as $M \rightarrow \infty$ in $L^1([0, 1])$ if and only if for all $n \neq 0$, $\lim_{M \rightarrow \infty} \widehat{g}_1(n) \cdots \widehat{g}_M(n) = 0$.

◇ Gives info on rate of convergence.

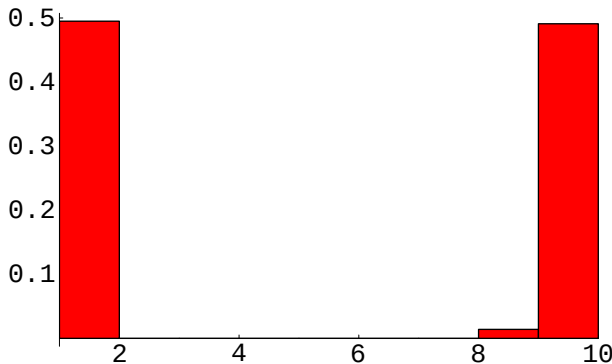
Generalizations

- Levy proved for i.i.d.r.v. just one year after Benford's paper.
- Generalized to other compact groups, with estimates on the rate of convergence.
 - ◇ Stromberg: n -fold convolution of a regular probability measure on a compact Hausdorff group G converges to normalized Haar measure in weak-star topology iff support of the distribution not contained in a coset of a proper normal closed subgroup of G .

Distribution of digits (base 10) of 1000 products

$X_1 \cdots X_{1000}$, where $g_{10,m} = \phi_{11^m}$.

$\phi_m(x) = m$ if $|x - 1/8| \leq 1/2m$ (0 otherwise).



Proof under stronger conditions

- Use standard CLT to show $Y_1 + \cdots + Y_M$ tends to a Gaussian.
- Use Poisson Summation to show the Gaussian tends to the uniform modulo 1.

Proof under stronger conditions

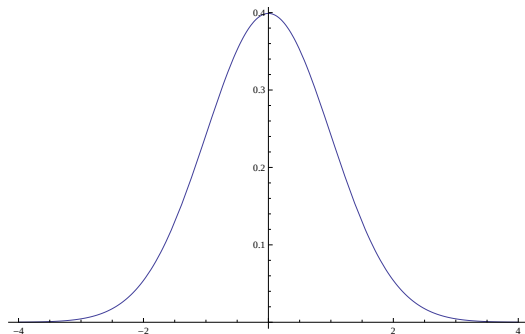


Figure: Plot of normal (mean 0, stdev 1).

Proof under stronger conditions

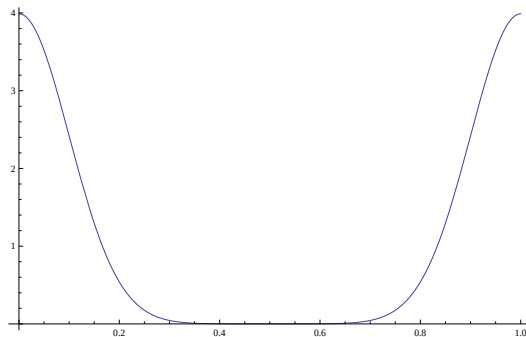


Figure: Plot of normal (mean 0, stdev .1) modulo 1.

Proof under stronger conditions

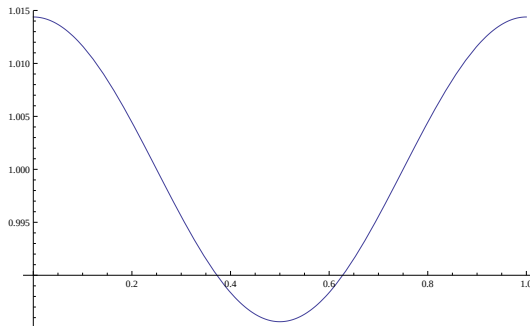


Figure: Plot of normal (mean 0, stdev .5) modulo 1.

Inputs

Poisson Summation Formula

f nice:

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

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

Lemma

$$\frac{2}{\sqrt{2\pi\sigma^2}} \int_{\sigma^{1+\delta}}^{\infty} e^{-x^2/2\sigma^2} dx \ll e^{-\sigma^{2\delta}/2}.$$

Proof Under Weaker Conditions

Lemma

As $N \rightarrow \infty$, $p_N(x) = \frac{e^{-\pi x^2/N}}{\sqrt{N}}$ becomes equidistributed modulo 1.

- $\int_{\substack{x=-\infty \\ x \bmod 1 \in [a,b]}}^{\infty} p_N(x) dx = \frac{1}{\sqrt{N}} \sum_{n \in \mathbb{Z}} \int_{x=a}^b e^{-\pi(x+n)^2/N} dx.$
- $e^{-\pi(x+n)^2/N} = e^{-\pi n^2/N} + O\left(\frac{\max(1, |n|)}{N} e^{-n^2/N}\right).$
- Can restrict sum to $|n| \leq N^{5/4}.$
- $\frac{1}{\sqrt{N}} \sum_{n \in \mathbb{Z}} e^{-\pi n^2/N} = \sum_{n \in \mathbb{Z}} e^{-\pi n^2 N}.$

Proof Under Weaker Conditions

$$\begin{aligned}
 & \frac{1}{\sqrt{N}} \sum_{|n| \leq N^{5/4}} \int_{x=a}^b e^{-\pi(x+n)^2/N} dx \\
 &= \frac{1}{\sqrt{N}} \sum_{|n| \leq N^{5/4}} \int_{x=a}^b \left[e^{-\pi n^2/N} + O\left(\frac{\max(1, |n|)}{N} e^{-n^2/N}\right) \right] dx \\
 &= \frac{b-a}{\sqrt{N}} \sum_{|n| \leq N^{5/4}} e^{-\pi n^2/N} + O\left(\frac{1}{N} \sum_{n=0}^{N^{5/4}} \frac{n+1}{\sqrt{N}} e^{-\pi(n/\sqrt{N})^2}\right) \\
 &= \frac{b-a}{\sqrt{N}} \sum_{|n| \leq N^{5/4}} e^{-\pi n^2/N} + O\left(\frac{1}{N} \int_{w=0}^{N^{3/4}} (w+1) e^{-\pi w^2} \sqrt{N} dw\right) \\
 &= \frac{b-a}{\sqrt{N}} \sum_{|n| \leq N^{5/4}} e^{-\pi n^2/N} + O\left(N^{-1/2}\right).
 \end{aligned}$$

Proof Under Weaker Conditions

Extend sums to $n \in \mathbb{Z}$, apply Poisson
Summation:

$$\frac{1}{\sqrt{N}} \sum_{n \in \mathbb{Z}} \int_{x=a}^b e^{-\pi(x+n)^2/N} dx \approx (b-a) \cdot \sum_{n \in \mathbb{Z}} e^{-\pi n^2 N}.$$

For $n = 0$ the right hand side is $b - a$.

For all other n , we trivially estimate the sum:

$$\sum_{n \neq 0} e^{-\pi n^2 N} \leq 2 \sum_{n \geq 1} e^{-\pi n N} \leq \frac{2e^{-\pi N}}{1 - e^{-\pi N}},$$

which is less than $4e^{-\pi N}$ for N sufficiently large.

Proof in General Case: Fourier input

- Fejér kernel:

$$F_N(x) = \sum_{n=-N}^N \left(1 - \frac{|n|}{N}\right) e^{2\pi i n x}.$$

- Fejér series $T_N f(x)$ equals

$$(f * F_N)(x) = \sum_{n=-N}^N \left(1 - \frac{|n|}{N}\right) \widehat{f}(n) e^{2\pi i n x}.$$

- Lebesgue's Theorem: $f \in L^1([0, 1])$. As $N \rightarrow \infty$, $T_N f$ converges to f in $L^1([0, 1])$.
- $T_N(f * g) = (T_N f) * g$: convolution assoc.

Proof of Modulo 1 CLT

- Density of sum is $h_\ell = g_1 * \cdots * g_\ell$.
- Suffices show $\forall \epsilon: \lim_{M \rightarrow \infty} \int_0^1 |h_M(x) - 1| dx < \epsilon$.
- Lebesgue's Theorem: N large,

$$\|h_1 - T_N h_1\|_1 = \int_0^1 |h_1(x) - T_N h_1(x)| dx < \frac{\epsilon}{2}.$$

- Claim: above holds for h_M for all M .

Proof of Modulo 1 CLT : Proof of Claim

$$T_N h_{M+1} = T_N(h_M * g_{M+1}) = (T_N h_M) * g_{M+1}$$

$$\begin{aligned}
 \|h_{M+1} - T_N h_{M+1}\|_1 &= \int_0^1 |h_{M+1}(x) - T_N h_{M+1}(x)| dx \\
 &= \int_0^1 |(h_M * g_{M+1})(x) - (T_N h_M) * g_{M+1}(x)| dx \\
 &= \int_0^1 \left| \int_0^1 (h_M(y) - T_N h_M(y)) g_{M+1}(x - y) dy \right| dx \\
 &\leq \int_0^1 \int_0^1 |h_M(y) - T_N h_M(y)| g_{M+1}(x - y) dx dy \\
 &= \int_0^1 |h_M(y) - T_N h_M(y)| dy \cdot 1 < \frac{\epsilon}{2}.
 \end{aligned}$$

Proof of Modulo 1 CLT

Show $\lim_{M \rightarrow \infty} \|h_M - 1\|_1 = 0$.

Triangle inequality:

$$\|h_M - 1\|_1 \leq \|h_M - T_N h_M\|_1 + \|T_N h_M - 1\|_1.$$

Choices of N and ϵ :

$$\|h_M - T_N h_M\|_1 < \epsilon/2.$$

Show $\|T_N h_M - 1\|_1 < \epsilon/2$.

Proof of Modulo 1 CLT

$$\begin{aligned} \|T_N h_M - 1\|_1 &= \int_0^1 \left| \sum_{\substack{n=-N \\ n \neq 0}}^N \left(1 - \frac{|n|}{N}\right) \widehat{h}_M(n) e^{2\pi i n x} \right| dx \\ &\leq \sum_{\substack{n=-N \\ n \neq 0}}^N \left(1 - \frac{|n|}{N}\right) |\widehat{h}_M(n)| \end{aligned}$$

$$\widehat{h}_M(n) = \widehat{g}_1(n) \cdots \widehat{g}_M(n) \longrightarrow_{M \rightarrow \infty} 0.$$

For fixed N and ϵ , choose M large so that $|\widehat{h}_M(n)| < \epsilon/4N$ whenever $n \neq 0$ and $|n| \leq N$.

Products and Chains of Random Variables

Key Ingredients

- Mellin transform and Fourier transform related by **logarithmic** change of variable.
- Poisson summation from collapsing to modulo 1 random variables.

Preliminaries

- $\Xi_1, \dots, \Xi_n$ nice independent r.v.'s on $[0, \infty)$.
- Density $\Xi_1 \cdot \Xi_2$:

$$\int_0^\infty f_2\left(\frac{x}{t}\right) f_1(t) \frac{dt}{t}$$

Preliminaries

- $\Xi_1, \dots, \Xi_n$ nice independent r.v.'s on $[0, \infty)$.
- Density $\Xi_1 \cdot \Xi_2$:

$$\int_0^\infty f_2\left(\frac{x}{t}\right) f_1(t) \frac{dt}{t}$$

◇ Proof: $\text{Prob}(\Xi_1 \cdot \Xi_2 \in [0, x])$:

$$\begin{aligned} & \int_{t=0}^\infty \text{Prob}\left(\Xi_2 \in \left[0, \frac{x}{t}\right]\right) f_1(t) dt \\ &= \int_{t=0}^\infty F_2\left(\frac{x}{t}\right) f_1(t) dt, \end{aligned}$$

differentiate.

Mellin Transform

$$(\mathcal{M}f)(s) = \int_0^{\infty} f(x) x^s \frac{dx}{x}$$

$$(\mathcal{M}^{-1}g)(x) = \frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} g(s) x^{-s} ds$$

$$g(s) = (\mathcal{M}f)(s), f(x) = (\mathcal{M}^{-1}g)(x).$$

$$(f_1 \star f_2)(x) = \int_0^{\infty} f_2\left(\frac{x}{t}\right) f_1(t) \frac{dt}{t}$$

$$(\mathcal{M}(f_1 \star f_2))(s) = (\mathcal{M}f_1)(s) \cdot (\mathcal{M}f_2)(s).$$

Mellin Transform Formulation: Products Random Variables

Theorem

X_i 's independent, densities f_i . $\Xi_n = X_1 \cdots X_n$,

$$h_n(x_n) = (f_1 \star \cdots \star f_n)(x_n)$$

$$(\mathcal{M}h_n)(s) = \prod_{m=1}^n (\mathcal{M}f_m)(s).$$

As $n \rightarrow \infty$, Ξ_n becomes Benford: $Y_n = \log_B \Xi_n$,
 $|\text{Prob}(Y_n \bmod 1 \in [a, b]) - (b - a)| \leq$

$$(b - a) \cdot \sum_{\ell \neq 0, \ell = -\infty}^{\infty} \prod_{m=1}^n (\mathcal{M}f_i) \left(1 - \frac{2\pi i \ell}{\log B} \right).$$

Proof of Kossovsky's Chain Conjecture for certain densities

Conditions

- $\{\mathcal{D}_i(\theta)\}_{i \in I}$: one-parameter distributions, densities $f_{\mathcal{D}_i(\theta)}$ on $[0, \infty)$.
- $p : \mathbb{N} \rightarrow I$, $X_1 \sim \mathcal{D}_{p(1)}(1)$, $X_m \sim \mathcal{D}_{p(m)}(X_{m-1})$.
- $m \geq 2$,

$$f_m(x_m) = \int_0^\infty f_{\mathcal{D}_{p(m)}(1)}\left(\frac{x_m}{x_{m-1}}\right) f_{m-1}(x_{m-1}) \frac{dx_{m-1}}{x_{m-1}}$$

•

$$\lim_{n \rightarrow \infty} \sum_{\substack{\ell = -\infty \\ \ell \neq 0}}^{\infty} \prod_{m=1}^n (\mathcal{M} f_{\mathcal{D}_{p(m)}(1)}) \left(1 - \frac{2\pi i \ell}{\log B}\right) = 0$$

Chains of Random Variables

Return to street problem: chain of uniforms.

Let $\mathcal{D}_{\text{unif}}(\theta)$ be the density of a uniform random variable on $[0, \theta]$.

Let $X_1 \sim \mathcal{D}_{\text{unif}}(1)$ and $X_{n+1} \sim \mathcal{D}_{\text{unif}}(X_n)$.

Proof of Kossovsky's Chain Conjecture for certain densities

Theorem (JKKKM)

- *If conditions hold, as $n \rightarrow \infty$ the distribution of leading digits of X_n tends to Benford's law.*
- *The error is a nice function of the Mellin transforms: if $Y_n = \log_B X_n$, then*

$$\begin{aligned}
 & |\text{Prob}(Y_n \bmod 1 \in [a, b]) - (b - a)| \leq \\
 & \left| (b - a) \cdot \sum_{\substack{\ell=-\infty \\ \ell \neq 0}}^{\infty} \prod_{m=1}^n (\mathcal{M}f_{\mathcal{D}_{p(m)}(1)}) \left(1 - \frac{2\pi i \ell}{\log B}\right) \right|
 \end{aligned}$$

Example: All $X_i \sim \text{Exp}(1)$

- $X_i \sim \text{Exp}(1)$, $Y_n = \log_B \Xi_n$.
- Needed ingredients:
 - ◊ $\int_0^\infty \exp(-x)x^{s-1}dx = \Gamma(s)$.
 - ◊ $|\Gamma(1 + ix)| = \sqrt{\pi x / \sinh(\pi x)}$, $x \in \mathbb{R}$.
- $|P_n(s) - \log_{10}(s)| \leq$

$$\log_B s \sum_{\ell=1}^{\infty} \left(\frac{2\pi^2 \ell / \log B}{\sinh(2\pi^2 \ell / \log B)} \right)^{n/2}.$$

Example: All $X_i \sim \text{Exp}(1)$

Bounds on the error

- $|P_n(s) - \log_{10} s| \leq$
 - ◇ $3.3 \cdot 10^{-3} \log_B s$ if $n = 2$,
 - ◇ $1.9 \cdot 10^{-4} \log_B s$ if $n = 3$,
 - ◇ $1.1 \cdot 10^{-5} \log_B s$ if $n = 5$, and
 - ◇ $3.6 \cdot 10^{-13} \log_B s$ if $n = 10$.
- Error at most

$$\log_{10} s \sum_{\ell=1}^{\infty} \left(\frac{17.148\ell}{\exp(8.5726\ell)} \right)^{n/2} \leq .057^n \log_{10} s$$