The Amusing Law of Benford

Last updated: 2021/10/06, 20:26:04 EDT

Principles of Complex Systems, Vols. 1 & 2
CSYS/MATH 300 and 303, 2021–2022| @pocsvox

Prof. Peter Sheridan Dodds | @peterdodds

Computational Story Lab | Vermont Complex Systems Center
Vermont Advanced Computing Core | University of Vermont

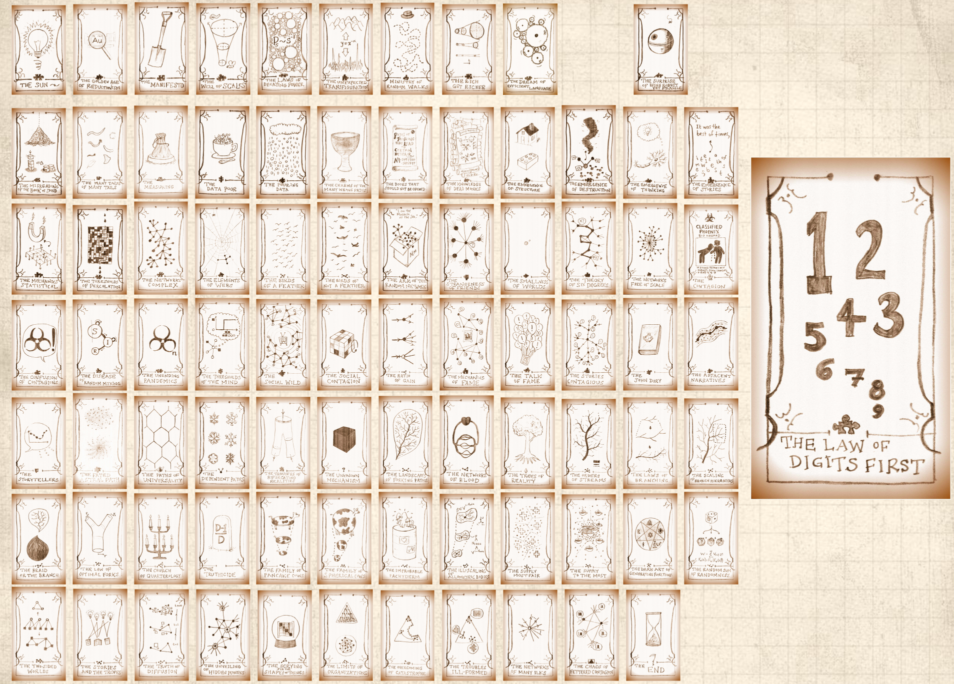

These slides are brought to you by:

These slides are also brought to you by:

Special Guest Executive Producer

On Instagram at pratchett_the_cat

Outline


THE SUN
THE GOLDEN AGE OF REDUCTIONISM
THE MANIFESTO
THE WELL OF SCALES
THE LAWS OF DECAYING POWER
THE UNEXPECTED TRANSFIGURATION
MINISTRY OF RANDOM WALKS
THE RICH GET RICHER
THE DREAM OF EFFICIENT LANGUAGE
THE SURPRISE OF BEING ROBUST YET FRAGILE
THE MISREADING OF THE BOOK OF SAND
THE MANY TALES OF MANY TAILS
THE MEASURING
THE DATA POOR
THE POURING DATA
THE CHARMS OF THE MANY WRONG PATHS
THE DOORS THAT SHOULD NOT BE OPENED
THE KNOWLEDGE OF DEAD MONKS
THE EMERGENCE OF STRUCTURE
THE EMERGENCE OF DESTRUCTION
THE EMERGENCE OF THINKING
It was the best of times,
THE EMERGENCE OF STORIES
THE MECHANICS STATISTICAL
THE THRESHOLDS OF PERCOLATION
THE NETWORKS COMPLEX
THE ELEMENTS OF WEBS
THE BIRDS OF A FEATHER
THE BIRDS OF NOT A FEATHER
THE ARK OF THE RANDOM NETWORK
THE STRANGENESS OF FRIENDS
THE SMALLNESS OF WORLDS
THE THEORY OF SIX DEGREES
THE NETWORKS FREE OF SCALE
CLASSIFIED
PHOENIX
BIO HAZARD
THE CONTAGION
THE CONFUSION OF CONTAGIONS
THE DISEASE OF RANDOM MIXING
THE UNENDING PANDEMICS
THE THRESHOLDS OF THE MIND
THE SOCIAL WILD
THE SOCIAL CONTAGION
THE RATIO OF GAIN
THE MECHANICS OF FAME
THE TALK OF FAME
THE STORIES CONTAGIOUS
THE JOHN DORY
THE ADJACENT NARRATIVES
THE STORYTELLERS
THE FATED ASTRAL PATH
THE PATHS OF UNIVERSALITY
THE DEPENDENT PATHS
THE TROUSERS OF BIFURCATING REALITIES
THE UNKNOWN MECHANISM
THE LANDSCAPES OF FORKING PATHS
THE NETWORKS OF BLOOD
THE TREES OF REALITY
THE ORDERS OF STREAMS
THE LAWS OF BRANCHING
THE SCALING OF BROKEN HIERARCHIES
THE BRAID OR THE BRANCH
THE LAW OF OPTIMAL FORKS
THE CHURCH OF QUARTEROLOGY
THE TRUTHICIDE
THE FAMILY OF PANCAKE COWS
THE FAMILY OF SPHERICAL COWS
THE IMPROBABLE PACHYDERM
THE ILL-SCALING OF ALLOMETRIC BODIES
THE SUPPLY MOST FAIR
THE SUPPLY TO THE MOST
THE DARK ART OF GENERATING FUNCTIONS
THE RANDOM SUM OF RANDOMNESS
THE TWO-SIDED WORLDS
THE STORIES AND THE TROPES
THE TRUTH OF DIFFUSION
THE UNVEILING OF HIDDEN POWERS
THE SCRYING FOR THE SHAPES OF THINGS
THE LIMITS OF ORGANIZATIONS
THE OVERCOMING OF CATASTROPHE
THE TROUBLES ILL-FORMED
THE NETWORKS OF MANY ILKS
THE CHAOS OF FETTERED CONTAGION
THE END
1 2 4 3 5 6 7 8 9
THE LAW OF DIGITS FIRST

Benford's Law—The Law of First Digits

$$P(\text{first digit} = d) \propto \log_b \left(1 + \frac{1}{d}\right)$$

for certain sets of 'naturally' occurring numbers in base b

Benford's Law—The Law of First Digits

- $$P(\text{first digit} = d) \propto \log_b \left(1 + \frac{1}{d}\right)$$
  for certain sets of 'naturally' occurring numbers in base b
- Around 30.1% of first digits are '1',
  compared to only 4.6% for '9'.

Benford's Law—The Law of First Digits

- $$P(\text{first digit} = d) \propto \log_b \left(1 + \frac{1}{d}\right)$$

  for certain sets of 'naturally' occurring numbers in base b
- Around 30.1% of first digits are '1', compared to only 4.6% for '9'.
- First observed by Simon Newcomb [3] in 1881 "Note on the Frequency of Use of the Different Digits in Natural Numbers"

Benford's Law—The Law of First Digits

- $$P(\text{first digit} = d) \propto \log_b \left(1 + \frac{1}{d}\right)$$

  for certain sets of 'naturally' occurring numbers in base b
- Around 30.1% of first digits are '1', compared to only 4.6% for '9'.
- First observed by Simon Newcomb [3] in 1881 "Note on the Frequency of Use of the Different Digits in Natural Numbers"
- Independently discovered in 1938 by Frank Benford.

Benford's Law—The Law of First Digits

$$P(\text{first digit} = d) \propto \log_b \left(1 + \frac{1}{d}\right)$$

for certain sets of 'naturally' occurring numbers in base b

- Around 30.1% of first digits are '1', compared to only 4.6% for '9'.
- First observed by Simon Newcomb [3] in 1881 "Note on the Frequency of Use of the Different Digits in Natural Numbers"
- Independently discovered in 1938 by Frank Benford.
- Newcomb almost always noted but Benford gets the stamp, according to Stigler's Law of Eponymy.

Benford's Law—The Law of First Digits

Observed for

- Fundamental constants (electron mass, charge, etc.)
- Utility bills
- Numbers on tax returns (ha!)
- Death rates
- Street addresses
- Numbers in newspapers

Benford's Law—The Law of First Digits

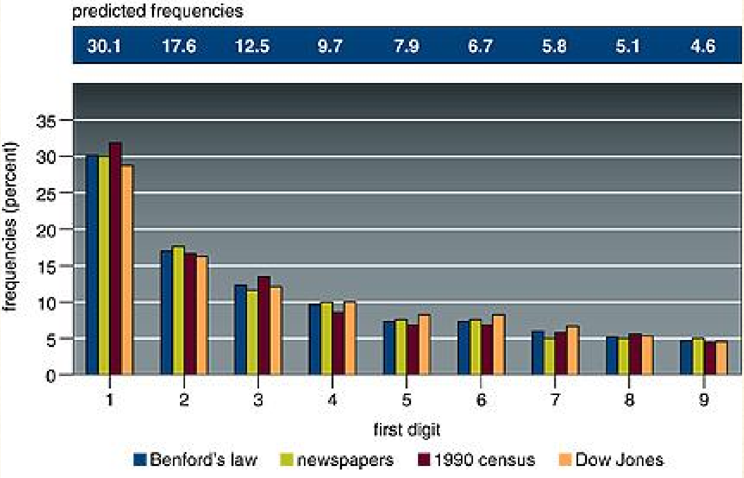
Observed for

- Fundamental constants (electron mass, charge, etc.)
- Utility bills
- Numbers on tax returns (ha!)
- Death rates
- Street addresses
- Numbers in newspapers

- Cited as evidence of fraud in the 2009 Iranian elections.

Benford's Law—The Law of First Digits

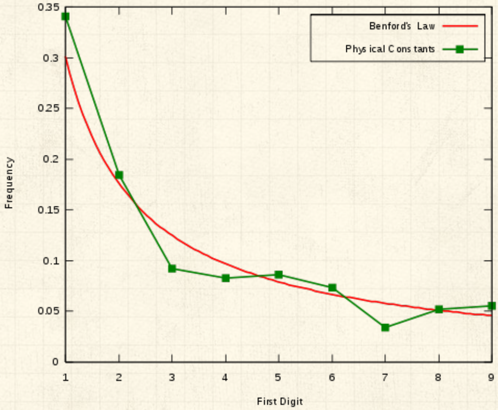
Real data:



From 'The First-Digit Phenomenon' by T. P. Hill (1998) [1]

Benford's Law—The Law of First Digits

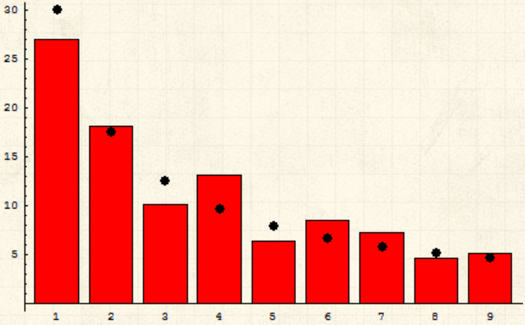
Physical constants of the universe:



Taken from here.

Benford's Law—The Law of First Digits

Population of countries:



Taken from here.

Essential story

$$P(\text{first digit} = d) \propto \log_b \left(1 + \frac{1}{d}\right)$$

Essential story

$$P(\text{first digit} = d) \propto \log_b \left(1 + \frac{1}{d}\right)$$

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

Essential story

$$P(\text{first digit} = d) \propto \log_b\left(1 + \frac{1}{d}\right)$$

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

$$= \log_b(d+1) - \log_b(d)$$

Essential story

$$P(\text{first digit} = d) \propto \log_b \left(1 + \frac{1}{d}\right)$$

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

$$= \log_b (d+1) - \log_b (d)$$

- Observe this distribution if numbers are distributed uniformly in log-space:

$$P(\log_e x)\,\mathrm{d}(\log_e x) \propto 1 \cdot \mathrm{d}(\log_e x)$$

Essential story

$$P(\text{first digit} = d) \propto \log_b\left(1 + \frac{1}{d}\right)$$

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

$$= \log_b(d+1) - \log_b(d)$$

- Observe this distribution if numbers are distributed uniformly in log-space:

$$P(\log_e x)\,\mathrm{d}(\log_e x) \propto 1 \cdot \mathrm{d}(\log_e x) = x^{-1}\,\mathrm{d}x$$

Essential story

- $$P(\text{first digit} = d) \propto \log_b\left(1 + \frac{1}{d}\right)$$
  $$= \log_b\left(\frac{d+1}{d}\right)$$
  $$= \log_b(d+1) - \log_b(d)$$

- Observe this distribution if numbers are distributed uniformly in log-space:

  $$P(\log_e x)\,\mathrm{d}(\log_e x) \propto 1\cdot\mathrm{d}(\log_e x) = x^{-1}\,\mathrm{d}x = P(x)\,\mathrm{d}x$$

Essential story

$$P(\text{first digit} = d) \propto \log_b\left(1 + \frac{1}{d}\right)$$

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

$$= \log_b(d+1) - \log_b(d)$$

- Observe this distribution if numbers are distributed uniformly in log-space:

$$P(\log_e x)\,\mathrm{d}(\log_e x) \propto 1{\cdot}\mathrm{d}(\log_e x) = x^{-1}\,\mathrm{d}x = P(x)\,\mathrm{d}x$$

- Power law distributions at work again...

Essential story

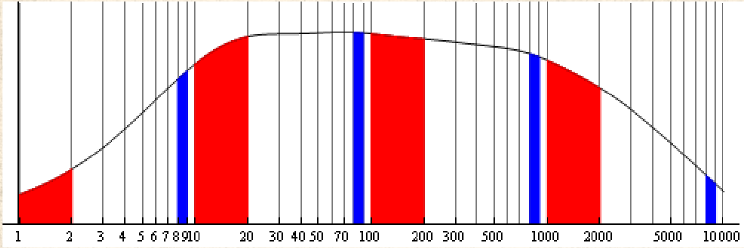
- $$P(\text{first digit} = d) \propto \log_b\left(1 + \frac{1}{d}\right)$$
$$= \log_b\left(\frac{d+1}{d}\right)$$
$$= \log_b(d+1) - \log_b(d)$$

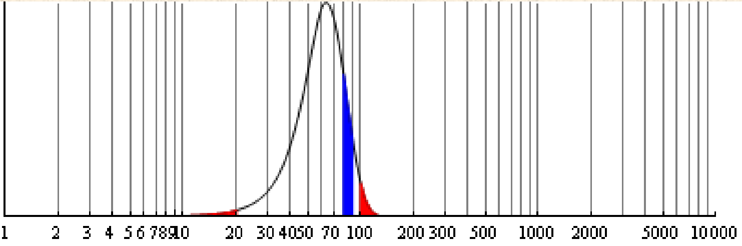
- Observe this distribution if numbers are distributed uniformly in log-space:

$$P(\log_e x)\,\mathrm{d}(\log_e x) \propto 1 \cdot \mathrm{d}(\log_e x) = x^{-1}\,\mathrm{d}x = P(x)\,\mathrm{d}x$$

- Power law distributions at work again...
- Extreme case of $\gamma \simeq 1$.

Benford’s law





Taken from here.

"Citations to articles citing Benford's law: A Benford analysis"
Tariq Ahmad Mir,
Preprint available at http://arxiv.org/abs/1602.01205, 2016. [2]

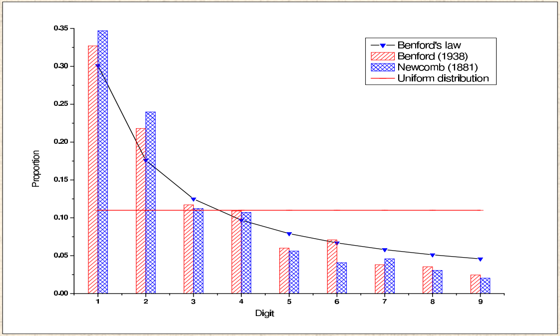


Fig. 1: The observed proportions of first digits of citations received by the articles citing FB and SN on September 30, 2012. For comparison the proportions expected from BL and uniform distributions are also shown.

On counting and logarithms:

- Earlier: Listen to Radiolab's "Numbers.".
- Now: Benford's Law.

References I

[1] T. P. Hill.
The first-digit phenomenon.
American Scientist, 86:358–, 1998.

[2] T. A. Mir.
Citations to articles citing Benford's law: A Benford analysis, 2016.
Preprint available at http://arxiv.org/abs/1602.01205. pdf

[3] S. Newcomb.
Note on the frequency of use of the different digits in natural numbers.
American Journal of Mathematics, 4:39–40, 1881.
pdf