# Benford's Law

DOWNLOAD *Mathematica* Notebook

A phenomenological law also called the first digit law, first digit phenomenon, or leading digit phenomenon. Benford's law states that in listings, tables of statistics, etc., the digit 1 tends to occur with probability $\sim 30\%$, much greater than the expected 11.1% (i.e., one digit out of 9). Benford's law can be observed, for instance, by examining tables of logarithms and noting that the first pages are much more worn and smudged than later pages (Newcomb 1881). While Benford's law unquestionably applies to many situations in the real world, a satisfactory explanation has been given only recently through the work of Hill (1998).

Benford's law was used by the character Charlie Eppes as an analogy to help solve a series of high burglaries in the Season 2 "The Running Man" episode (2006) of the television crime drama *NUMB3RS*.

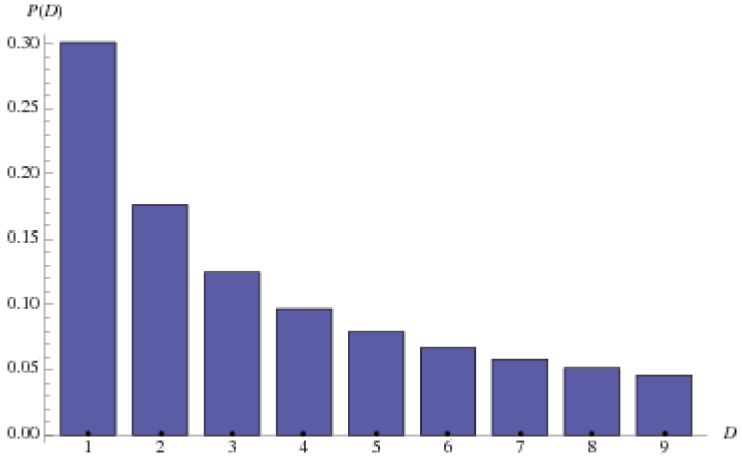
Benford's law applies to data that are *not* dimensionless, so the numerical values of the data depend on the units. If there exists a universal probability distribution $P(x)$ over such numbers, then it must be invariant under a change of scale, so

$$P(k x) = f(k) P(x). \qquad (1)$$

If $\int P(x)\, dx = 1$, then $\int P(k x)\, dx = 1/k$, and normalization implies $f(k) = 1/k$. Differentiating with respect to $k$ and setting $k = 1$ gives

$$x P'(x) = -P(x), \qquad (2)$$

having solution $P(x) = 1/x$. Although this is not a proper probability distribution (since it diverges), both the laws of physics and human convention impose cutoffs. For example, randomly selected street addresses obey something close to Benford's law.




If many powers of 10 lie between the cutoffs, then the probability that the first (decimal) digit is $D$ is given by a logarithmic distribution

$$P_D = \frac{\int_D^{D+1} P(x)\,dx}{\int_1^{10} P(x)\,dx} = \log_{10}\left(1 + \frac{1}{D}\right) \quad (3)$$

for $D = 1$, ..., 9, illustrated above and tabulated below.

| $D$ | $P_D$ | $D$ | $P_D$ |
|---|---|---|---|
| 1 | 0.30103 | 6 | 0.0669468 |
| 2 | 0.176091 | 7 | 0.0579919 |
| 3 | 0.124939 | 8 | 0.0511525 |
| 4 | 0.09691 | 9 | 0.0457575 |
| 5 | 0.0791812 | | |

However, Benford's law applies not only to scale-invariant data, but also to numbers chosen from a variety of different sources. Explaining this fact requires a more rigorous investigation of central limit-like theorems for the mantissas of random variables under multiplication. As the number of variables increases, the density function approaches that of the above logarithmic distribution. Hill (1998) rigorously demonstrated that the "distribution of distributions" given by random samples taken from a variety of different distributions is, in fact, Benford's law (Matthews).

One striking example of Benford's law is given by the 54 million real constants in Plouffe's "Inverse Symbolic Calculator" database, 30% of which begin with the digit 1. Taking data from several disparate sources, the table below shows the distribution of first digits as compiled by Benford (1938) in his original paper.

| col. | title | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | samples |
|---|---|---|---|---|---|---|---|---|---|---|---|
| A | Rivers, Area | 31.0 | 16.4 | 10.7 | 11.3 | 7.2 | 8.6 | 5.5 | 4.2 | 5.1 | 335 |
| B | Population | 33.9 | 20.4 | 14.2 | 8.1 | 7.2 | 6.2 | 4.1 | 3.7 | 2.2 | 3259 |
| C | Constants | 41.3 | 14.4 | 4.8 | 8.6 | 10.6 | 5.8 | 1.0 | 2.9 | 10.6 | 104 |
| D | Newspapers | 30.0 | 18.0 | 12.0 | 10.0 | 8.0 | 6.0 | 6.0 | 5.0 | 5.0 | 100 |
| E | Specific Heat | 24.0 | 18.4 | 16.2 | 14.6 | 10.6 | 4.1 | 3.2 | 4.8 | 4.1 | 1389 |
| F | Pressure | 29.6 | 18.3 | 12.8 | 9.8 | 8.3 | 6.4 | 5.7 | 4.4 | 4.7 | 703 |
| G | H.P. Lost | 30.0 | 18.4 | 11.9 | 10.8 | 8.1 | 7.0 | 5.1 | 5.1 | 3.6 | 690 |
| H | Mol. Wgt. | 26.7 | 25.2 | 15.4 | 10.8 | 6.7 | 5.1 | 4.1 | 2.8 | 3.2 | 1800 |
| I | Drainage | 27.1 | 23.9 | 13.8 | 12.6 | 8.2 | 5.0 | 5.0 | 2.5 | 1.9 | 159 |
| J | Atomic Wgt. | 47.2 | 18.7 | 5.5 | 4.4 | 6.6 | 4.4 | 3.3 | 4.4 | 5.5 | 91 |
| K | $n^{-1}$, $\sqrt{n}$ | 25.7 | 20.3 | 9.7 | 6.8 | 6.6 | 6.8 | 7.2 | 8.0 | 8.9 | 5000 |

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| L | Design | 26.8 | 14.8 | 14.3 | 7.5 | 8.3 | 8.4 | 7.0 | 7.3 | 5.6 | 560 |
| M | Reader's Digest | 33.4 | 18.5 | 12.4 | 7.5 | 7.1 | 6.5 | 5.5 | 4.9 | 4.2 | 308 |
| N | Cost Data | 32.4 | 18.8 | 10.1 | 10.1 | 9.8 | 5.5 | 4.7 | 5.5 | 3.1 | 741 |
| O | X-Ray Volts | 27.9 | 17.5 | 14.4 | 9.0 | 8.1 | 7.4 | 5.1 | 5.8 | 4.8 | 707 |
| P | Am. League | 32.7 | 17.6 | 12.6 | 9.8 | 7.4 | 6.4 | 4.9 | 5.6 | 3.0 | 1458 |
| Q | Blackbody | 31.0 | 17.3 | 14.1 | 8.7 | 6.6 | 7.0 | 5.2 | 4.7 | 5.4 | 1165 |
| R | Addresses | 28.9 | 19.2 | 12.6 | 8.8 | 8.5 | 6.4 | 5.6 | 5.0 | 5.0 | 342 |
| S | $n^1$, $n^2 \cdots n!$ | 25.3 | 16.0 | 12.0 | 10.0 | 8.5 | 8.8 | 6.8 | 7.1 | 5.5 | 900 |
| T | Death Rate | 27.0 | 18.6 | 15.7 | 9.4 | 6.7 | 6.5 | 7.2 | 4.8 | 4.1 | 418 |
| | Average | 30.6 | 18.5 | 12.4 | 9.4 | 8.0 | 6.4 | 5.1 | 4.9 | 4.7 | 1011 |
| | Probable Error | $\pm 0.8$ | $\pm 0.4$ | $\pm 0.4$ | $\pm 0.3$ | $\pm 0.2$ | $\pm 0.2$ | $\pm 0.2$ | $\pm 0.3$ | | |

The following table gives the distribution of the first digit of the mantissa following Benford's Law using a number of different methods.

| method | Sloane | sequence |
|---|---|---|
| Sainte-Lague | A055439 | 1, 2, 3, 1, 4, 5, 6, 1, 2, 7, 8, 9, ... |
| d'Hondt | A055440 | 1, 2, 1, 3, 1, 4, 2, 5, 1, 6, 3, 1, ... |

| | | |
|---|---|---|
| largest remainder, Hare quotas | A055441 | 1, 2, 3, 4, 1, 5, 6, 7, 1, 2, 8, 1, ... |
| largest remainder, Droop quotas | A055442 | 1, 2, 3, 1, 4, 5, 6, 1, 2, 7, 8, 1, ... |

**SEE ALSO:** Logarithmic Distribution

**REFERENCES:**


Barlow, J. L. and Bareiss, E. H. "On Roundoff Error Distributions in Floating Point and Logarithmic Arithmetic." *Computing* **34**, 325-347, 1985.

Benford, F. "The Law of Anomalous Numbers." *Proc. Amer. Phil. Soc.* **78**, 551-572, 1938.

Bogomolny, A. "Benford's Law and Zipf's Law." http://www.cut-the-knot.org/do_you_know/zipfLaw.shtml.

Boyle, J. "An Application of Fourier Series to the Most Significant Digit Problem." *Amer. Math. Monthly* **101**, 879-886, 1994.

Flehinger, B. J. "On the Probability that a Random Integer Has Initial Digit $A$." *Amer. Math. Monthly* **73**, 1056-1061, 1966.

Franel, J. *Naturforschende Gesellschaft, Vierteljahrsschrift (Zürich)* **62**, 286-295, 1917.

Havil, J. "Benford's Law." §14.2 in *Gamma: Exploring Euler's Constant.* Princeton, NJ: Princeton University Press, pp. 145-155, 2003.

Hill, T. P. "Base-Invariance Implies Benford's Law." *Proc. Amer. Math. Soc.* **12**, 887-895, 1995a.

Hill, T. P. "The Significant-Digit Phenomenon." *Amer. Math. Monthly* **102**, 322-327, 1995b.

Hill, T. P. "A Statistical Derivation of the Significant-Digit Law." *Stat. Sci.* **10**, 354-363, 1995c.

Hill, T. P. "The First Digit Phenomenon." *Amer. Sci.* **86**, 358-363, 1998.

Knuth, D. E. "The Fraction Parts." §4.2.4B in *The Art of Computer Programming, Vol. 2:*

*Seminumerical Algorithms, 3rd ed.* Reading, MA: Addison-Wesley, pp. 254-262, 1998.

Ley, E. "On the Peculiar Distribution of the U.S. Stock Indices Digits." *Amer. Stat.* **50**, 311-313, 1996.

Livio, M. *The Golden Ratio: The Story of Phi, the World's Most Astonishing Number.* New York: Broadway Books, pp. 232-236, 2002.

Matthews, R. "The Power of One." http://www.fortunecity.com/emachines/e11/86/one.html.

Newcomb, S. "Note on the Frequency of the Use of Digits in Natural Numbers." *Amer. J. Math.* **4**, 39-40, 1881.

Nigrini, M. J. *The Detection of Income Tax Evasion Through an Analysis of Digital Frequencies.* Ph.D. thesis. Cincinnati, OH: University of Cincinnati, 1992.

Nigrini, M. "A Taxpayer Compliance Application of Benford's Law." *J. Amer. Tax. Assoc.* **18**, 72-91, 1996.

Nigrini, M. "I've Got Your Number." *J. Accountancy* **187**, pp. 79-83, May 1999. http://www.aicpa.org/pubs/jofa/may1999/nigrini.htm.

Nigrini, M. *Digital Analysis Using Benford's Law: Tests Statistics for Auditors.* Vancouver, Canada: Global Audit Publications, 2000.

Plouffe, S. "Graph of the Number of Entries in Plouffe's Inverter." http://www.lacim.uqam.ca/~plouffe/statistics.html.

Raimi, R. A. "The Peculiar Distribution of First Digits." *Sci. Amer.* **221**, 109-119, Dec. 1969.

Raimi, R. A. "On the Distribution of First Significant Digits." *Amer. Math. Monthly* **76**, 342-348, 1969.

Raimi, R. A. "The First Digit Phenomenon." *Amer. Math. Monthly* **83**, 521-538, 1976.

Schatte, P. "Zur Verteilung der Mantisse in der Gleitkommadarstellung einer Zufallsgröße." *Z. Angew. Math. Mech.* **53**, 553-565, 1973.

Schatte, P. "On Mantissa Distributions in Computing and Benford's Law." *J. Inform. Process. Cybernet.* **24**, 443-455, 1988.

Sloane, N. J. A. Sequences A055439, A055440, A055441, and A055442 in "The On-Line Encyclopedia of Integer Sequences."