Acta Scientific Applied Physics

Volume 4 Issue 1 January 2024

Review Article

Are Those Data Real? Benford's Law

Douglas Winslow Cooper*

Douglas Winslow Cooper, PhD, LLC, Walden, NY, USA

***Corresponding Author:** Douglas Winslow Cooper, Douglas Winslow Cooper, PhD, LLC, Walden, NY, USA.

Received: January 02, 2024

Published: May 15, 2024

Abstract

Viewing an extensive data set, you wonder whether to trust it. A simple method for checking a theory's or experiment's results as a first step in its evaluation is usually less costly than attempting replication.

Keywords: Benford's Law; Zipf's Law

After scanning the data set for obvious problems, try using Benford's Law [1] of First Digits, discussed below, as an additional check: comparison of the individual first digit frequencies against Benford's Law.

If we write the data in the format d.ef x 10^n, using scientific notation, the "significand"= d.ef in the data in many situations will follow "Benford's Law." The first digit, d, can be expected to follow not a uniform probability distribution, having f(d) = 1/9, but the Benford (1938) [1] distribution,

f(d) = log[(d+1)/d] -------- [1]

This has a cumulative distribution for the first digit:

F(0<d<s) = log s --------- [2]

The logarithm is to the base 10. For example, the probability of d=1 for the first digit is f(1)=log(2) = 0.301, rather than f = 1/9 = 0.111 one would expect if the first digit was equally probable for d=1 to d=9. The largest first digit is much less probable than the first, f(9) = log(10/9) = 0.046 per Benford's Law. The relationship has been widely used (Miller, 2015) [2] in forensic accounting for situations where the data cover orders of magnitude. For example, it arises naturally in cases of geometric (compound) growth or decay and in some power law relationships (e.g., Zipf's Law).

Compound growth, a constant rate of fractional change per time, is quite familiar in economics, where Benford's Law has seen the most application. General Electric applied physicist Frank Benford (1938) [1] presented a detailed exposition and multiple demonstrations, following a pattern first described by the astronomer Simon Newcomb (1881) [4].

Compound growth or decay

In the physical sciences, situations of compound growth or decay are modeled with the relationships

dy/dt = k y ------- [3]

y = a exp(k t) ------ [4]

The derivative dy/dt resembles a velocity, indicating the rate at which y goes through the significand's digits for a particular order of magnitude.

For a constant k, log y is equally probable in equal time intervals, dt. A graph of y versus t on semi-log coordinates is a straight line. Population growth, radioactive decay, Fibonacci numbers, and some chemical reaction kinetics show such behavior over orders of magnitude. Doyle's article in Miller (2015) [2] gives many physical science examples.

Linear growth

Contrast compound growth with linear, additive growth,

y = b + ct ---------- [5]

dy/dt = c ---------- [6]

Changes in y are equally probable in each time interval, dt. A graph of y versus t on linear coordinates is a straight line, with y spending the same residence duration in each time interval. We expect the first digits, d, from 1 to 9 to be equally probable, f = 1/9.

The difference in linear versus compound growth is analogous to the difference in the random additive changes that give rise to Gaussian (normal) distributions versus the random multiplicative changes that produce lognormal distributions. Miller (2015) [2] notes that data from lognormal distributions with large standard deviations follow Benford closely and that we should expect this behavior more in broad distributions rather than narrow ones.

Power-law behavior

For the general power-law relationship

y = m t^n ------- [7]

log y = log m + n log t ---------- [8]

dy/dt = (m n) t^(n-1) -------- [9]

The variable y will spend less, the same, or more time in interval dt for the values of the exponent, n>1, n=1, n<1. Thus, the probability of being measured with such values will follow a similar pattern.

Products of Benford random variables (with some restrictions) are themselves Benford, but sums of Benford variables are not (Miller, 2015) [2].

Benford's law in the natural sciences

The article with this title, by David Hoyle in Miller (2015) [2], is particularly relevant for scientists and engineers.

Power-law probabilities are "scale-invariant," with the general form for the probability density of x indicating p(x) is proportional to x^-a. This includes the familiar Zipf's Law, with a =1, which gives rise to the Benford distribution.

Miller (2015) [2] discusses the conditions for the solutions of the partial differential equation for heat transfer to be Benford or not. This reference has an extensive bibliography.

Is this data set Benford?

Benford's Law uses the frequency distribution of the first digits and compares it to what is expected from the Benford distribution, perhaps graphically or using a chi-square distribution.

Miller (2015) [2] notes that the sums of the significands of each of the leading digits approach being equal: the sum of the 1.xx values will approximate the sum of the 2.xx values, etc., a "sum-invariance" behavior whose discovery he credits to Nigrini (1992). This may turn out to be better in some situations.

The article by Nigrini in Miller (2015) [2] presents three elements that data must have to be possibly Benford:

- Real things counted or measured.
- No built-in maximum or minimum value (except for zero).
- The number should not represent names or labels rather than magnitudes.

In Conclusion

Doyle in Miller (2015) [2] wrote the following concerning Benford's Law [1] distribution and scientific studies:

- Benford's Law gives the natural FSD distribution for many scientific data sets.
- Benford's Law can be used to assess data quality and consistency.
- Benford's Law can be used to check for fraudulent scientific data.
- Benford's Law can be used to create appropriate simulated data sets.

Benford's Law is receiving increased attention as a method for prescreening data for possible errors or falsification (Murtaugh, 2023) [3].

Bibliography

1. Benford F. "The Law of Anomalous Numbers". *Proceedings of the American Philosophical Society* 78 (1938): 551-572.

2. Miller S. J. "Benford's Law: Theory and Application". Princeton University Press, Princeton, NJ (2015).

3. Murtaugh J. "This Unexpected Pattern of Numbers Is Everywhere". *Scientific American* 329.5 (2023): 82.

4. Newcomb S. "Note on the Frequency of the Use of the Different Digits of Natural Numbers". *American Journal of Mathematics* 4 (1881): 551-552.