Abdelwahid Benslimane
wahid.benslimane@gmail.com

Benford’s law
(also known as Newcomb-Benford’s law)

$$P(\text{"the digit at the first position of a number is d"}) = \log_{10}(1 + \frac{1}{d})$$

Benford's law is concerned with numbers from everyday life or nature, such as salary, turnover, number of followers on a social network, atomic weights etc., but not with purely random data (e.g. lottery draws).

Benford’s law measures the likelihood that a digit between 1 and 9 is the first of a number, i.e. the one on the most left-hand side.

This law does not consider 0 and takes into account the first digit other than 0 after the decimal point for float numbers strictly lower than 1. For example, for the number 7541, the first digit is 7, and for 0.0059 the first digit considered is 5.

Benford's law indicates that in a general context, and therefore whatever the domain under consideration, the probabilities of encountering the different digits at the first position of the numbers are respectively:

P(1) = 30.1%, P(2) = 17.6%, P(3) = 12.5%, P(4) = 9.7%, P(5) = 7.9%,

P(6) = 6.7%, P(7) = 5.8%, P(8) = 5.1%, P(9) = 4.6%.

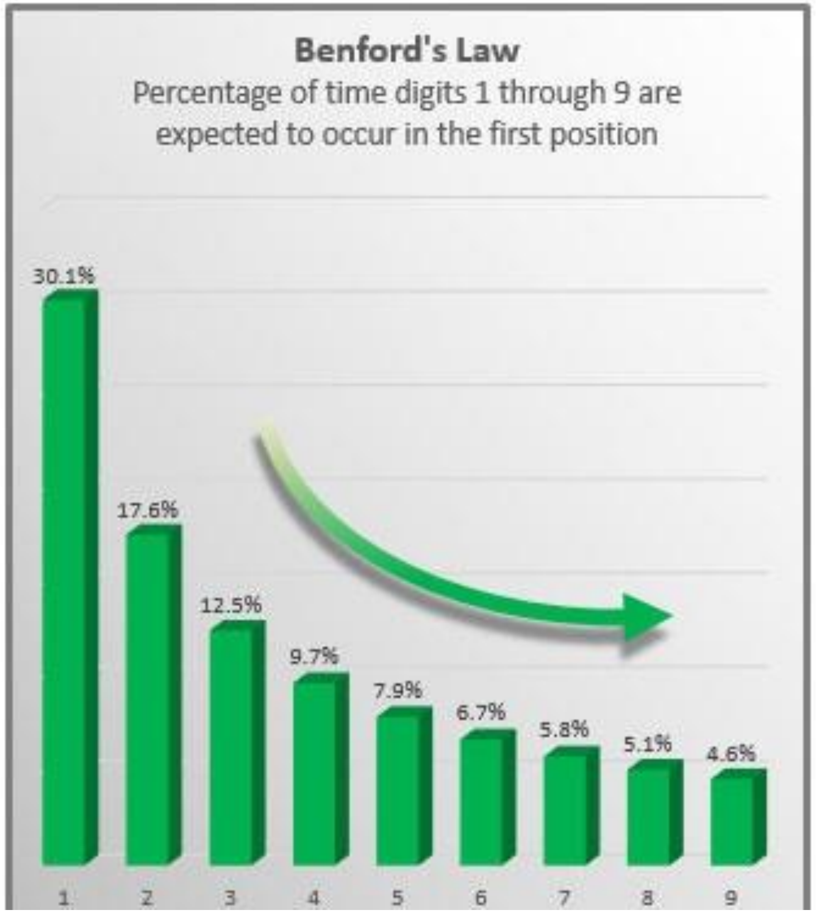


Image from https://www.journalofaccountancy.com/issues/2017/apr/excel-and-benfords-law-to-detect-fraud.html

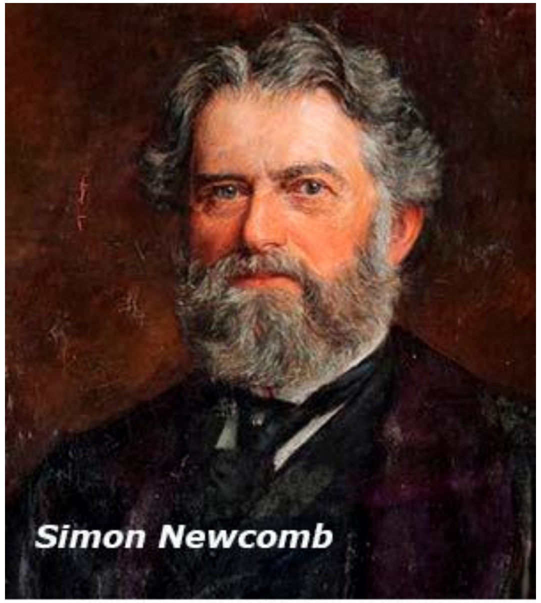
The mathematical formulation of this law is:

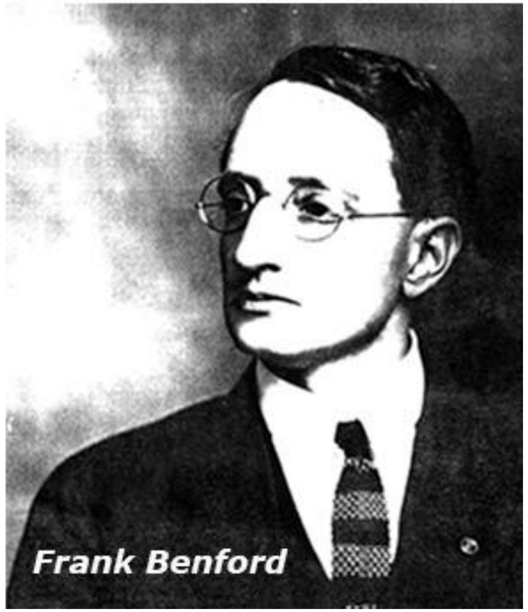
$$P(\text{"the digit at the first position of a number is d"}) = \log_{10}(1 + \tfrac{1}{d})$$
$$d \in \{1, 2, 3, 4, 5, 6, 7, 8, 9\}$$

A little bit of history…

Although it is called Benford's Law, this law was first formulated by Canadian astronomer Simon Newcomb in 1881 who observed that the first pages of logarithm tables (those giving the logarithms of numbers beginning with 1) were more damaged than the following ones.

His paper on this subject was ignored until 57 years later when the American physicist Frank Benford made the same observation about the deterioration of the pages of numerical tables.





Why then are there more 1s than 9s?

Since its discovery, Benford's law has been the subject of a large number of publications. The Benford Online Bibliography site (https://www.benfordonline.net) lists an impressive number of them. The mathematical articles essentially seek to answer two questions:

- What general conditions can explain the appearance of Benford's law?
- Why do most empirical data approximately verify this law?

Abdelwahid Benslimane wahid.benslimane@gmail.com

Mathematics without borders

Let us assume that there is indeed a mathematical law that governs the numbers encountered in everyday life. The least we can expect is that this law is the same everywhere in the world!

It is therefore also expected to be the same in countries where distances are expressed in the metric system and in countries using Anglo-Saxon measurements, and the same whether we are talking about Dollar, Euro, Yen or any other currency.

In other words, this law should not depend on the units in which the quantities are expressed.

It can be shown that this property of invariance by change of unit characterises Benford's law:

https://web.williams.edu/Mathematics/sjmiller/public_html/BrownClasses/197/benford/Pinkham_FirstDigit.pdf

Metric Units	Customary Units
1 centimeter	0.394 inch
1 meter	3.281 feet or 1.093 yards
1 kilometer	0.621 mile
1 gram	0.035 ounce
1 kilogram	2.205 pounds
1 milliliter	0.034 fluid ounce
1 liter	1.057 quart or 0.264 gallon

Abdelwahid Benslimane wahid.benslimane@gmail.com

Fraud detection

In the 1990s, the American economist Mark Nigrini suggested the use of tests based on Benford's Law for fraud detection.

Experience shows that authentic data should follow Benford's Law. On the other hand, certain numbers tend to be over-represented in falsified data.

For example, an accountant writing fictitious checks may intentionally keep the check amounts below the company's authorization threshold. If that threshold is €500, an analysis of the amounts of those checks may show that the number 4 appears at the top of the list more frequently than would be expected under Benford's Law.

In practice, more refined tests involving the distribution of the first two significant digits are used. These tests have been used to detect falsifications in company accounts and have since been used in various fields.