Benford's law

Benford's Law

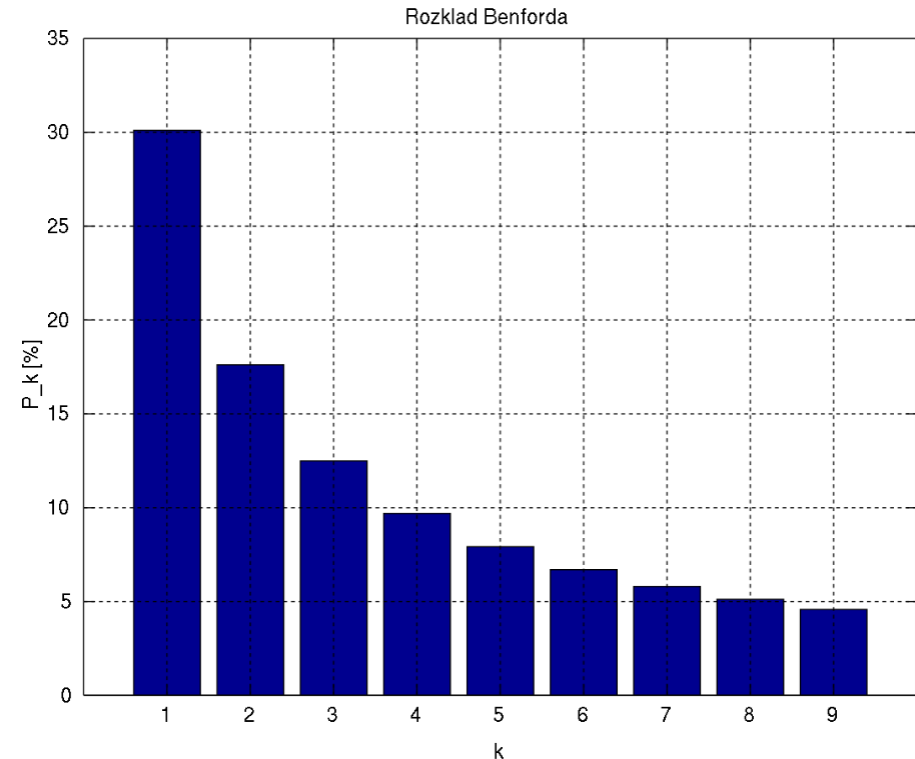In 1881, Simon Newcomb determined that the probability that a number has a first digit d is given by

$$P(d) = \log_{10}(d+1) - \log_{10}(d) = \log_{10}\left(\frac{d+1}{d}\right) = \log_{10}\left(1 + \frac{1}{d}\right)$$

In 1938, Frank Benford found that a large number of naturally-occurring datasets follow this pattern
- Surface areas of rivers, molecular weights, death rates, street addresses, and the numbers contained in an issue of Reader's Digest.

"Benford's Law" has been used to detect irregularities in
- Published scientific studies, fraudulent election data in Iran, suspicious macroeconomic data from Greece, and tax return misreporting

Source: https://en.wikipedia.org/wiki/Benford%27s_law
Source: https://www.coursera.org/learn/accounting-analytics?specialization=business-analytics

Benford's Law

Expected distribution of leading digits (Benford's law)

Leading digit	1	2	3	4	5	6	7	8	9
Frequency	30.1%	17.6%	12.5%	9.7%	7.9%	6.7%	5.8%	5.1%	4.6%

Amiram, Bozanic, and Rouen (2014) studied Benford's Law in financial statements

- Aggregating all financial statements by industries or by years for 2000-2011, the leading digits in 83.7% of firm's annual financial statements conform with Benford's distribution
- distribution of leading digits from 215 months of returns for the Fairfield Sentry Fund, a fund that invested only with Bernie Madoff

Leading digit	1	2	3	4	5	6	7	8	9
Frequency	39.6%	14.2%	10.4%	7.1%	7.5%	6.6%	6.1%	6.6%	1.9%

Use Kolmogorov – Smirnov test for similarity between two vectors.

Source: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2374093
Source: https://www.coursera.org/learn/accounting-analytics?specialization=business-analytics