Benford's Law: A Universal Law on Digit Patterns in Mathematics, Physics, Nature, and Economics

Narken Aimambet, Chingis Matayev, Alisher Urazbekov, Michael Santana (Graduate Assistant), A.J. Hildebrand (Faculty Mentor)

NAZARBAYEV UNIVERSITY ILLINOIS UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN

Illinois Summer Research Symposium (ISRS), July 23, 2015

Frank Benford

- Benford's Law is named after Frank Benford, an American physicist and electrical engineer at General Electric.
- F. Benford (1938) analyzed 20 data sets, such as areas of rivers, population census data, specific heat of chemical substances, atomic weights, house address data, mathematical sequences, and even randomly picked numbers from newspapers.
- He found that for nearly all data sets the distribution of the first digits is approximately given by $P(d) = \log_{10}\left(1 + \frac{1}{d}\right)$. This distribution is now refered to as **Benford's distribution**.

When Does Benford's Law apply?

- Data that spans several orders of magnitude
  e.g., city population, financial data, areas of world countries
  (but **not** credit card numbers, ID numbers).
- Sequences that grow at geometric rates
  e.g., Fibonacci numbers, Powers of two
  (but **not** primes or squares).
- Data that result from many mathematical operations
  e.g., accounting data, tax return data.
- Data that are modeled by random growth process
  e.g., stock prices.

Data Sets Satisfying Benford's Law

- **Geographical Data**
  - Populations of US Cities
  - Areas of world countries
  - Elevation of world cities
  - Masses of 1140 known planets
  - Time intervals between earthquakes
- **Financial Data**
  - Market Capitalization
  - Dow Jones Index
  - Numbers in US Tax Return
- **Mathematical/Physical Data**
  - Physical constants
  - Fibonacci sequence
  - Numerical algorithms
  - Random simulation of stock prices
- **Other Data Sets**
  - Number of Facebook/Twitter followers (Golbeck)
  - File sizes in Linux File system

Using Benford's Law to Detect Fraud

- **Greece Economic Data**. B. Rauch et al. (2011) analyzed financial data of European Union countries using Benford's law. They found that among these countries, Greece had the largest deviation from Benford's law, as measured by the chi-square test.
- **Johannesburg Stock Market**. A.D. Saville (2006) analyzed 34 companies listed on the Johannesburg Stock Exchange that were suspected or known to have committed accounting fraud. Benford's law correctly identified 30 of them.

Acknowledgements

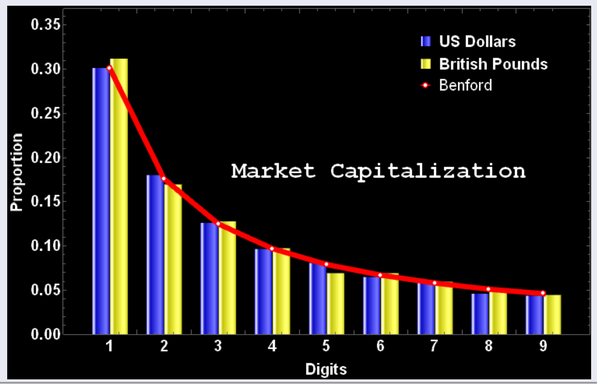
This poster was produced as part of the Illinois Summer Mathematics REU Program, with support from the Department of Mathematics at the University of Illinois and Nazarbayev University, Astana, Kazakhstan. The real world data was obtained from the Kossovsky Data Sets (2014) and Wolfram Data Sets. The mathematical data sets were generated using Wolfram *Mathematica*, Version 10.

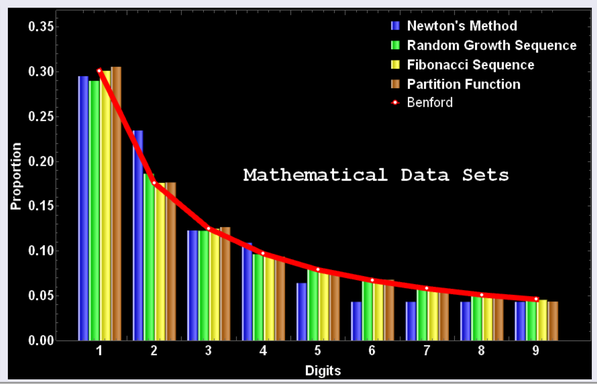
Benford's Distribution

First digit, d	1	2	3	4	5	6	7	8	9
Probability, $P(d)$	30.1%	17.6%	12.5%	9.7%	7.9%	6.7%	5.8%	5.1%	4.6%

Benford's Law in Action

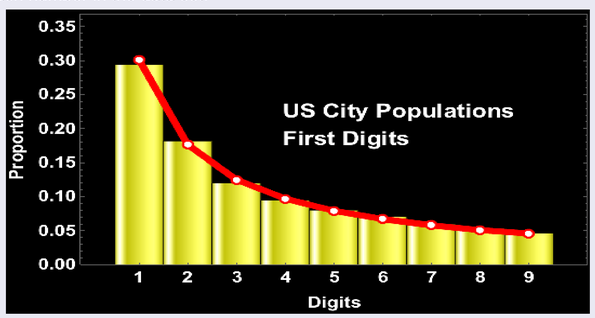
The graph on the left shows the distribution of first digits for the market capitalizations of 5486 international companies in both US Dollars and British Pounds, along with the Benford distribution (data from Kossovsky, 2014). The graphs on the right show the first digit distribution of several mathematical sequences. In both cases the distributions match the Benford distribution very closely.

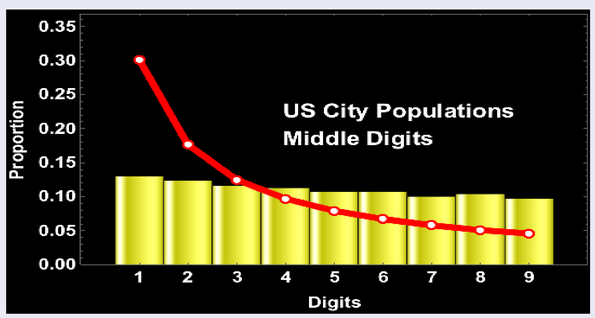


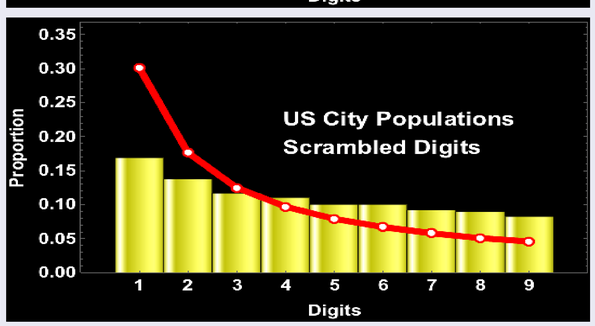


Significance of the First Digit

The graphs below illustrate the difference between the distribution of the first digits and the other digits. The first digits follow Benford's Law closely, while the middle and last digits are approximately uniformly distributed. The last graph shows the distribution of first digits after randomly reordering the digits of the each number in the data set.

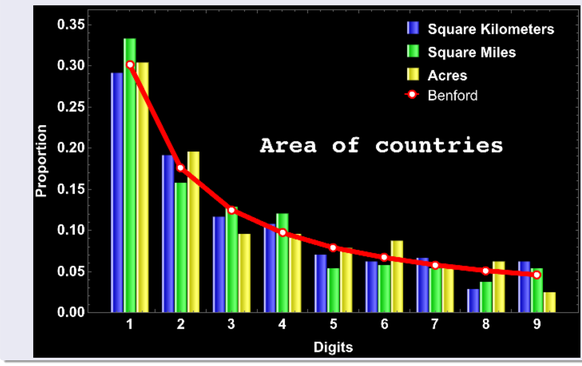






Universality of Benford

The first digit distribution of a Benford data set is independent of the units used. For example, in the graph below, changing square miles to square kilometers or acres does not affect the shape of the first digit distribution. The Benford distribution is the only distribution with this invariance of scale property (Hill, 1995).



Testing the Fit to Benford's Law

- **Mean Absolute Deviation (MAD) Test** (Cho & Gaines (2007)): $MAD = \frac{1}{9}\sum_{i=1}^{9} |p_i - o_i|$ where $(p_1, ..., p_9)$ is the Benford distribution and $(o_1, ..., o_9)$ is the observed distribution. For the Benford distribution, MAD = 0; for a uniform distribution MAD $\approx$ 0.059.
- **Ratio Test**: $R = o_1/o_9$ For a Benford distribution, $R = 6.5$; for a uniform distribution, $R = 1$.
- **Mantissa Test** (Nigrini (2011)): $M = \frac{1}{N}\left|\sum_{n=1}^{N} \exp(2\pi i \log_{10} x_i)\right|$ where $(x_1, ..., x_N)$ is the data set. For a Benford distribution, M = 0; for a uniform distribution, M $\approx$ 0.34.

Data Set	MAD Test	Ratio Test	Mantisa Test
US City Populations	0.0031	6.3473	0.0064
US City Populations (Scrambled)	0.0402	1.9095	0.2262
Market Capitalization	0.0020	6.8347	0.0109
Time between Earthquakes	0.0031	6.3473	0.0065
Fibonacci Sequence	0.0002	6.6031	0.0001
Random Growth Sequence	0.0051	7.1250	0.0281
Simulated Credit Card Numbers	0.0600	0.9621	0.3399

References

- F. Benford, *The law of anomalous numbers.* Proceeding of the American Philosophical Society (1938), 78(4), 551-572.
- A. Berger, T.P. Hill, *Newton's method obeys Benford's law.* American Mathematical Monthly (2007), 144(7), 588-601.
- W.K.T. Cho, B.J. Gaines, *Breaking the (Benford) law: Statistical fraud detection in campaign finance.* American Statistician(2007), 61(3), 218-223.
- P. Diaconis, *The Distribution of Leading Digits and Uniform Distribution Mod 1.* The Annals of Probability (1977), 5(1), 72-81.
- B. Rauch, M. Gottsche, *Fact and Fiction in EU-Governmental Economic Data.* German Economic Review(2011), 12(3), 243-255.
- T.P. Hill, *Base-Invariance Implies Benford's Law.* Proceedings of the American Mathematical Society (1995), 123(3), 887-895.
- A.E. Kossovsky, *Benford's Law: Theory, the General Law of Relative Quantities, and Forensic Fraud Detection Applications.* World Scientific Publishing Company (2014).
- A.E. Kossovsky, `http://www.forensicbenford.com/case-studies.html` Data Sets for the above book.
- M. Nigrini, *Forensic Analytics: Methods and Techniques for Forensic Accounting Investigations.* John Wiley & Sons, Hoboken, New Jersey (2011).
- A. Saville, *Using Benford's Law to detect data error and fraud: An Examination of Companies Listed on the Johannesburg Stock Exchange.* South African Journal of Economic and Management Science (2006), 9(3), 341-354.