SAMUEL NEAMAN INSTITUTE
FOR NATIONAL POLICY RESEARCH

Benford's Law

A survey of its varied applications

By Prof. (emer.) Shlomo Maital

With Ella Barzani

July 2022

Abstract

Benford's Law, or Newcomb-Benford Law, states that in many data sets, the leading digits (1's, 2's, etc.) will be much more frequent – 30% will be 1's, for instance. This law has been widely used in many applications, including detection of fraud and data manipulation. This paper surveys some of the more interesting papers, and offers an explanation of the Law.

The Search for Truth in Numbers

In history, people have sought truth (and especially, insights into the future) in a wide variety of strange ways – patterns in entrails, book passages, herring fat, chickens, flight of birds, and many others. (See Box: The Strange Search for Truth).

Box: The Strange Search for Truth

Through history, truth (especially regarding the future) has been sought (and, allegedly, found) in a variety of very strange places. Here are ten. (Source: Lovejoy[1]).

"1. **Hieromancy (Divination Using Entrails)**. Beginning in Mesopotamia and then in classical Greece and beyond, animals were sacrificed in divinatory rituals and their internal organs (notably the liver) were inspected for omens. 2. **Ornithomancy (Divination Using Birds)**. Interpreting the behavior of birds is one of the oldest forms of divination, and was a common part of Greek religious life. 3. **Pyro-osteomancy (Bone Oracles)**. In ancient China, bones were used to tell the future. During the Neolithic period, the Shang dynasty, and beyond specialists would inscribe questions on animal (often cattle) shoulder-blades and tortoise shells, then chisel pits into them and insert heated points. 4. **Bibliomancy (Divination Using Books)**. The practice of asking a question, opening a book at random, and interpreting the first passage your eyes (or fingers) hit upon as an answer was once widespread among the Greeks and Romans, the Muslim world, medieval Europe, and elsewhere. 5. **Alectryomancy (Divination Using Chickens)**. Chickens weren't just a handy food source in the ancient world - they could also predict the future. Various divinatory methods were employed in which chickens were offered a choice of grain in a particular location or direction, which corresponded to an answer to the subject in question (the parties in a battle, say, or the direction from which a future husband might come). 6. **Tyromancy (Divination Using Cheese)**. The use of cheese as a divination tool was known in the ancient world and the Middle Ages, although the details aren't very well-recorded. Some say the shapes of the holes in the cheeses were thought to hold meaning - a heart shape could

> indicate love, and certain holes could be read as initials. According to occultopedia.com, young women in the countryside would predict future husbands by writing the names of suitors on pieces of cheese. 7. **Ceromancy (Divination Using Melted Wax)**. A drawing of women practicing divination with lead or wax. A drawing of women practicing divination with lead or wax. The swirling shapes made by pouring melted wax into water were used as a divination tool in both ancient and medieval Europe. 8. **Cledonomancy (Divination by Words Overheard)**. For the ancient Greeks and Romans, chance utterings weren't always just that. The art of cledonomancy, or divination from overhead words, could be practiced either inside or outside of a specific ritual. In De Divinatione. 9. **Ring Oracles and "Under-the-Bowl Songs"**. In Russia, divination was once a popular pastime for the days just after the New Year, known as the strashyne ("fearful") days, when evil spirits were said to be particularly active. According to W. F. Ryan's The Bathhouse at Midnight: An Historical Survey of Magic and Divination in Russia, divination performed between midnight and 3 a.m. on these days was especially effective. One popular practice involved "under-the-bowl songs," in which rings and other personal objects were placed in a bowl and special divinatory songs sung, with each verse corresponding to a particular fortune (poverty, spinsterhood, etc.). 10. **Herring Fat and Membranes**. In mid-19th-century Belfast, according to Oxford's Dictionary of Superstitions, women predicted the character of their future husbands using the slimier parts of a herring."

Since 1881, and well before, scholars have sought truth by microscopic examination of numbers themselves. Can patterns in numbers reveal truth? Can patterns in data themselves reveal underlying truth or falsehood?

In 1881, Simon Newcomb, a mathematician, published an article[2] with a strange observation. He noted that the early pages of logarithm books, used at that time to carry out logarithmic calculations, were far more worn than the later pages. This led him to state the law that, in any list of numbers taken from an arbitrary set of data, more numbers will tend to begin with "1" than with any other digit. As we will see, the clue to the law lies in the nature of logarithms.

Later, in 1938, physicist Frank Benford[3] extended Newcomb's finding. He noted that "It has been observed that the first pages of a table of common logarithms show more wear than do the last pages, indicating that more used numbers begin with the digit 1 than with the digit 9. A compilation of some 20,000 first digits taken from widely divergent sources shows that there is a logarithmic distribution of first digits when the numbers are composed of four or more digits. An analysis of the numbers from different sources shows that the numbers taken from unrelated subjects, such as a group of newspaper items, show a similar pattern."

Since then, the Newcomb-Benford Law, or Benford's Law, stating that in many data sets the leading digit is likely to be small and follow a pattern (See Figure 1), has been widely confirmed.

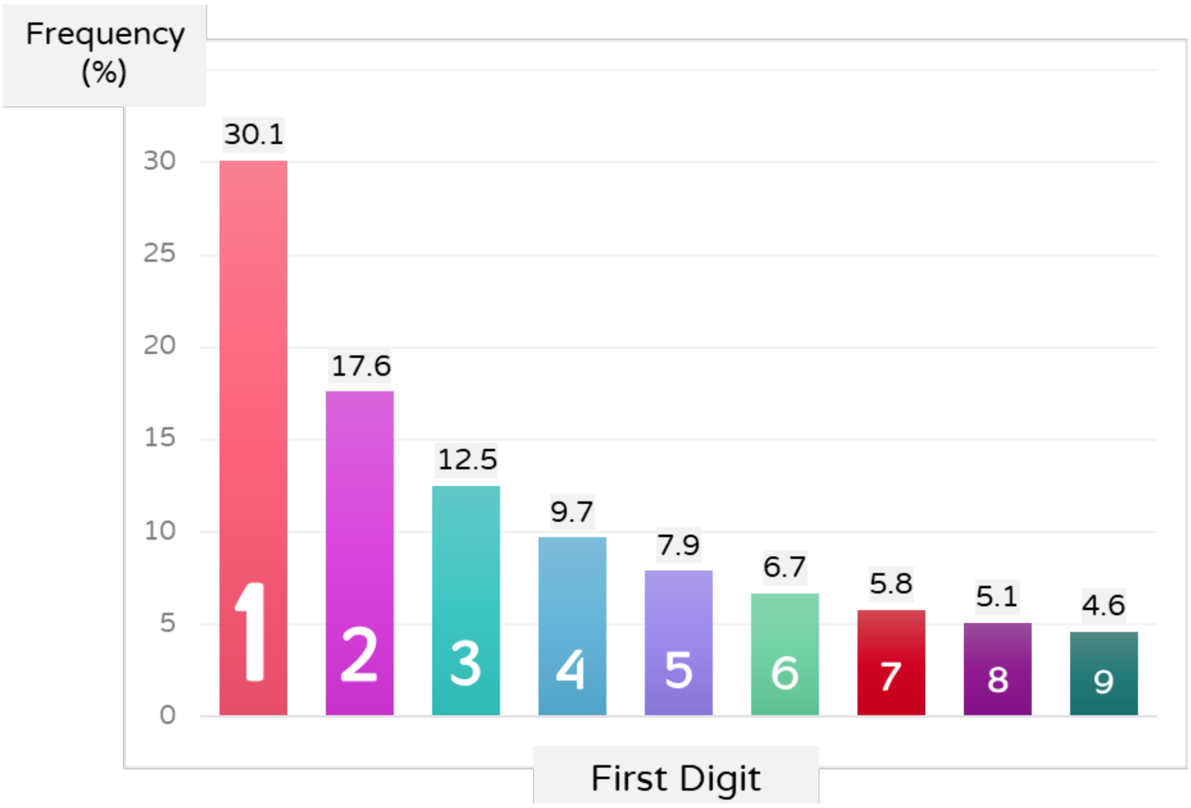

Figure 1. Frequency of digits, 1 to 9, according to Benford's Law[4]

Hill[5] concludes that Benford's Law has been employed "in different empirical contexts… (such as) computer design, mathematical modelling, and detection of accounting fraud". Miller[6] offers a comprehensive review of Benford's Law studies, published in 2015.

This survey provides an overview of a selection of research on Benford's Law, focused on applications related to fraud, error, and data manipulation. But first – an effort at explanation. Why should 1's, 2's and 3's be more frequent in data?

Benford's Law: A mathematical explanation

Gonsalves[7] offers a clear rationale for Benford's Law, through an example.

Suppose you have a large number of students each roll a six-sided die and record the result. Overall, if the die is fair, each number 1 through 6 will appear roughly the same number of times.

Now, have each student roll the die 100 times each and add up the numbers. The sum will average 350 (midway between 1 and 6) and the sums in all will have a bell-shaped (normal) distribution.

Finally, have each student roll the die 15 times each, and multiply the 15 results together. The average will be about 76 billion – and the result will be a log-normal distribution (that is: the logarithm of the result will have a bell-shaped distribution). Why?

Multiplying independent random variables (as the roll of a die) will tend toward a log-normal distribution. The logarithm of a given number x is the exponent to which another fixed number, say 10, is raised. It is the nature of multiplicative data, that expressing the number as the exponent of 10 greatly favors 1's and 2's. Tens and hundreds in data are given by 1's and 2's, in logarithmic form. Hence, Benford's Law. And generally, in data, there are more tens and hundreds than billions and trillions.

Detecting Fraud & Manipulation

If, then, the nature of data is such that it can be expected to follow Benford's Law, (e.g., multiplicative in nature), then perhaps Benford's Law can be used to test whether the data are true or manipulated. Arrange the data in logarithmic form, and examine whether the distribution is indeed log-normal, conforming to the Benford conjecture. This approach has been used to study the veracity of a wide variety of data.

Azevedo et al.[8] uses Benford's Law to test for fraud in Brazil's Bolsa Familia welfare program. They studied some 13 million records. They conclude that "according to the methodology adopted in the scope of this work, aggregated data from January 2018 conforms to Benford's Law". (p. 9).

Barabesi et al.[9] use goodness-of-fit methods (Pearson, or variants of it) to examine observed digit counts and to see if they fit Benford's Law. They also develop new tests beyond goodness-of-fit. They observe importantly that detection of outliers, and conformance to Benford's Law, each requires its own unique approach. They apply their novel methods of analysis of international trade data.

Bradshaw et al.[10] use Benford's Law to develop a method for fraud detection in big data analysis, focused on detecting sophisticated computational fraud. They make extensive use of machine learning as well. Using gene-copy number data as their inputs, and with digit frequency as their inputs, their method is shown capable of detecting fraud, with accuracy between 82% and 100%.

Feifei Li et al.[11] motivate the importance of data verification by noting that data volume worldwide is doubling every two years. "The era of big data has arrived," they note. They apply Benford's Law to the PRC People's Republic of China Central Bank M2 data and find it conforms. They observe that it is not sufficient to analyze the first digit only of data; further digits are needed. Reinforcing Benford's Law with other detection methods has proved highly effective.

Cerri J.[12] use datasets from Romanian and Soviet commercial whale catches in the Pacific Ocean to spot falsification. In both, they find, the first and second digits taken together deviate strongly from Benford's distribution. Revealing falsification in ecological research is "crucial for evidence-based conservation", they note.

Detecting Fraudulent Financial Data

Trust in financial data is crucial for well-functioning capital markets.

Nigrini[13] has written an entire volume, guiding financial auditors in use of Benford's Law to detect manipulation and deception. "The potential fraudster seeks to hide what is being attempted and has the time, resources and a significant incentive to do so. By contrast, the IT auditor… has comparatively limited time and resources. "His book offers a potentially powerful statistical tool to uncover financial fraud. Nigrini shows that routine use of Benford's Law "can be considered a basic IT auditing best practice."

Cella at al.[14] examine the link between 'transparency' (openness) of municipality expenditure data and veracity. They examine data from two municipalities in the Brazilian state of Goias, one with low transparency, one with high. They conclude that "the municipality with greater transparency has greater conformity with Benford's Law".

Sugiartu[15] in doctoral research uses Benford's Law to test digital frequency in financial data, and concludes that "Benford's Law can serve as an indicator tool in detecting the possibility of wrong data in financial statements". Barabesi et al.[16] develop Benford's Law tools for detecting fraudulent customs data in international trade. Similarly, Caarioli et al.[17] also analyze customs declarations in the European Union and "open the door to the development of modified goodness-of-fit procedures with wide applicability".

Voting Fraud

An Axios-Momentive poll finds majority of Americans do not believe Joe Biden legitimately won the 2020 presidential election despite massive evidence to the contrary. Can Benford's Law help reveal, or validate, election data?

Groharing et al.[18] analyze county-level voting data from Pennsylvania. That state is controversial, because of Trump's baseless contention that there was extensive fraud. Using Bedford's Law, the authors conclude that "Biden's county vote data in Pennsylvania is not suspicious".

Chinese data

In the face of unprecedented rapid economic growth in China, it has been widely contended that the data are manipulated and inaccurate. Huang et al.[19] use Benford's Law to examine Chinese macro data, specifically the first digit law, to test the quality of data from the Chinese Industrial Census. "We find no conclusive evidence of data manipulation by state-owned enterprises, whereas private firms tend to under-report performance", they find.

Power Systems vs. Hackers

Globally, power plans and electric power grids have been subject to hacking. Russian hackers carried out the first major cyberattack on a nation's electricity grid in late 2015, taking down part of the Ukrainian national grid for six hours. Reportedly, U.S. officials have been deeply concerned for years about a potential Russian hack of the US electric grid.

Milano et al.[20] suggest creatively that Benford's Law can be used to predict and forestall electric power plant and power grid hacking. How? Use of Benford's Law can detect malicious data introduced by hackers in the supervisory control and data acquisition system of a transmission network. They note that "tests based on power system models…show promising results".

Benford and Bach: Applications to Music

Nelson et al.[21] apply Benford to music. They examine musical note frequencies of the 88 keys of the piano and finds that they are Benford distributed (log normal). They extended the analysis beyond the 88 keys, to the range of notes from low (16.35 Hz) to high (7992 Hz) notes. They find that classical music (a random sample of songs from the Romantic period) adhere to the Benford distribution, while "modern music (a random sampling of songs from the 2000s) do not".

Benford's Law and COVID

During the global COVID-19 pandemic, there has been widespread controversy over the accuracy of COVID data across countries.

Morillas-Jurado et al.[22] provide a "management tool" to monitor death registration data, in the context of the pandemic. They rightly observe that it is crucial "to have objective and homogeneous criteria at the national level to guide health managers in the correct recording and evaluation of the magnitude of the pandemic". They propose Benford's Law as an auditing tool to monitor the reliability of the data.

Balashov et al.[23] study data on COVID-19 morbidity “from autocratic regimes” using Benford’s Law, as a first screening for “potential data manipulation” and conclude that “data from autocratic regimes and less-developed countries should be treated with more caution”.

Summary and Conclusion

It is truly fascinating that an 1881 observation by a mathematician, that the initial pages of a book of logarithms were more heavily thumbed than the latter pages, has evolved into a key observation, or law, that now finds use in checking veracity of data.

Fake news has become rampant. And fake data may not be far behind. As public skepticism grows regarding facts and data, Benford’s Law may find increasing use to restore public trust in numbers.

References

[1] Lovejoy, B. 10 Historical Divination Methods for Predicting the Future. https://www.mentalfloss.com/article/585258/historical-divination-methods-predict-future

[2] Newcomb, S. (1881). Note on the frequency of use of the different digits in natural numbers. American Journal of mathematics, 4(1), 39-40.

[3] Benford, F. (1938). The law of anomalous numbers. Proceedings of the American philosophical society, 551-572.

[4] Benford's Law. Wikipedia: www.wikipedia/Benford's Law

[5] Hill, T. P. (1995). A statistical derivation of the significant-digit law. Statistical science, 354-363.

[6] Miller, S. J. (Ed.). (2015). Benford's law. Princeton University Press.

[7] Gonsalves, R. Benford's Law – A Simple Explanation. www.towarddatascience.com/benford's-law-a-simple-explanation

[8] da Silva Azevedo, C., Gonçalves, R. F., Gava, V. L., & de Mesquita Spinola, M. (2021). A Benford's law based method for fraud detection using R Library. MethodsX, 8, 101575.

[9] Barabesi, L., Cerasa, A., Cerioli, A., & Perrotta, D. (2021). On characterizations and tests of Benford's law. Journal of the American Statistical Association, 1-17.

[10] Bradshaw, M. S., & Payne, S. H. (2021). Detecting fabrication in large-scale molecular omics data. PloS one, 16(11), e0260395.

[11] Li, F., Han, S., Zhang, H., Ding, J., Zhang, J., & Wu, J. (2019, February). Application of Benford's law in Data Analysis. In Journal of Physics: Conference Series (Vol. 1168, No. 3, p. 032133). IOP Publishing.

[12] Cerri, J. (2018). A fish rots from the head down: how to use the leading digits of ecological data to detect their falsification. bioRxiv, 368951.

[13] Nigrini, M. (2006). Digital analysis using benford's law: Tests and statistics for auditors. Global Audit Publications: Vancouver BC

[14] Cella, R. S., & Zanolla, E. (2018). Benford's Law and transparency: an analysis of municipal expenditure. BBR. Brazilian Business Review, 15, 331-347.

[15] Sugiarto, T., Budiman, A. I., & Rosini, I. (2016, December). The First Digits Analysis Until the Fifth Benford Law in Financial Statement. In International Conference on Ethics in Governance (ICONEG 2016) (pp. 1-4). Atlantis Press.

[16] Barabesi, L., Cerasa, A., Cerioli, A., & Perrotta, D. (2018). Goodness-of-fit testing for the Newcomb-Benford law with application to the detection of customs fraud. Journal of Business & Economic Statistics, 36(2), 346-358.

[17] Cerioli, A., Barabesi, L., Cerasa, A., Menegatti, M., & Perrotta, D. (2019). Newcomb–Benford law and the detection of frauds in international trade. Proceedings of the National Academy of Sciences, 116(1), 106-115.

[18] Groharing, B., & McCune, D. (2022). Benford's law and county-level votes in US presidential elections. CHANCE, 35(2), 4-10.

[19] Huang, Y., Niu, Z., & Yang, C. (2020). Testing firm-level data quality in China against Benford's Law. Economics Letters, 192, 109182.

[20] Milano, F., & Gomez-Exposito, A. (2020). Detection of cyber-attacks of power systems through Benford's law. IEEE Transactions on Smart Grid, 12(3), 2741-2744.

[21] Nelson, S. P., Wickman, B., Null, J., & Gazin, E. (2022). Benford's Law and Music Note Frequencies. In International Conference on Mathematics and Computation in Music (pp. 383-389). Springer, Cham.

[22] Morillas-Jurado, F. G., Caballer-Tarazona, M., & Caballer-Tarazona, V. (2021). Applying Benford's Law to Monitor Death Registration Data: A Management Tool for the COVID-19 Pandemic. Mathematics, 10(1), 46.

[23] Balashov, V. S., Yan, Y., & Zhu, X. (2021). Using the Newcomb–Benford law to study the association between a country's COVID-19 reporting accuracy and its development. Scientific reports, 11(1), 1-11.