Importance of First Natural Number: Benford's Law

Sudipta Sinha*

Assistant Professor, Department of Mathematics, Burdwan Raj College, Burdwan, Purba Bardhaman, West Bengal, Email: sudipta.sinha08@gmail.com

Abstract:

The distribution of first digits of real-world observations would not be uniform, but instead follow a trend where measurements with lower first digit (1, 2,) occur more frequently than those with higher first digits (…,8,9). Digit 1 occurs with highest frequency and digit 9 occurs with lowest frequency. The frequency decays monotonically from 1 to 9.

Benford's law, also called the first-digit law, is an observation about the frequency distribution of leading digits in many real-life sets of numerical data which includes electricity bills, street addresses, stock prices, house prices, population numbers, death rates, lengths of rivers, physical and mathematical constants and processes described by power laws (which are very common in nature). The law states that in many naturally occurring collections of numbers, the leading significant digit is likely to be small. For example, in sets which obey the law, the number 1 appears as the most significant digit about 30% of the time, while 9 appears as the most significant digit less than 5% of the time. Benford's law also makes predictions about the distribution of second digits, third digits, digit combinations, and so on.

Keywords: First-digit law, Benford's law, frequency distribution.

§1. Introduction:

In 1881 when Newcomb was looking at the logarithm tables, suddenly he observed that the pages that started with 1 were much more worn and torn than others. He was in great wonder. In 1938, physicist Benford came with the same observation again. He searched in twenty different domains which range from surface area of rivers to physical constants and molecular weights of chemicals. In all of the observations, the number 1 appeared as the first digit in nearly 30 % of the cases. It has the highest frequency of appearance. 2, 3, 4,5,6,7 and 8 come with gradually lesser frequencies. The number 9 appears only in 4 to 5% of the cases. This non-uniform distribution of single digit natural numbers distribution is known as 'Law of first digit' or 'Benford's law of distribution'. One might have thought that $p(1) = p(2) = \ldots\ldots = p(9) = 0.11$

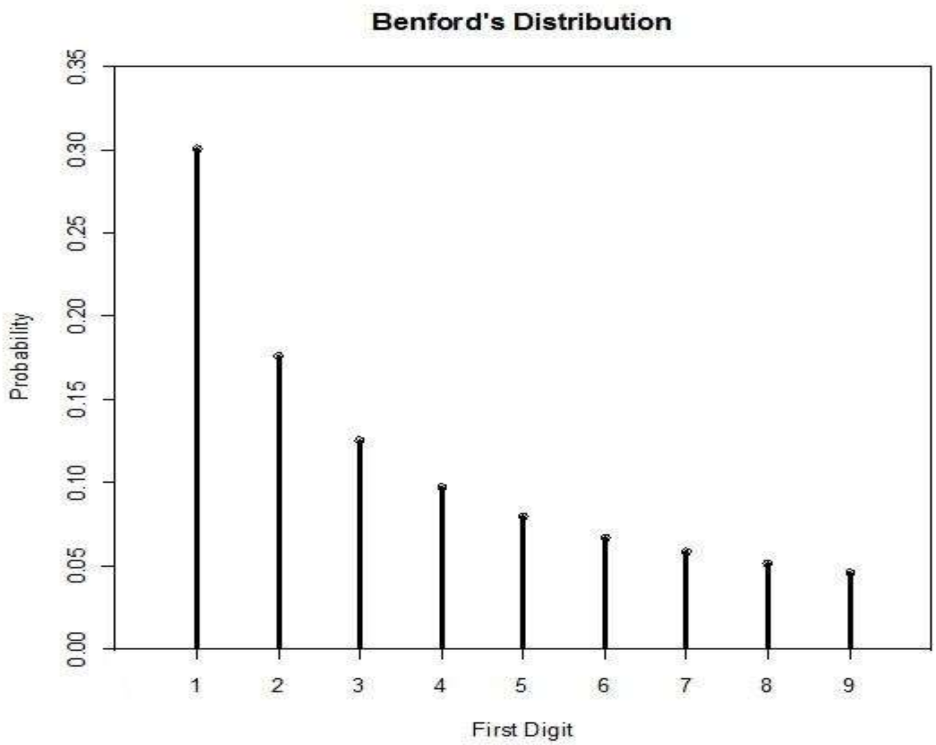
But in natural data, we surprisingly observe that

$p(1) = 0.30$; $p(4) = 0.10$; $p(7) = 0.06$

$p(2) = 0.18$; $p(5) = 0.08$; $p(8) = 0.05$

$p(3) = 0.12$; $p(6) = 0.07$; $p(9) = 0.045$

The idea will be clear if we look at the following figure



§2. Literature review

Recently Benford law has widely been used in music analysis. Isabel Barbancho et al (2015) applied Benford's law in music analysis. Kruger et al (2017) showed that the Benford's law is very effective in power law. They applied it as a power of one. Nirgini et al (2012) applied in fraud detection. Tota et al (2016) supports the Nirgini's statement and in favor of their statement they used this law in several cases of fraud detection. Berger et al (2011) developed the basics of Benford's law by probabilistic survey. Durtsch et al (2004) applied The Benford's law in fraud detection in data accounting. Newcomb (1881) described the frequency of digits in natural sectors. Scott et al ( ) revealed a new dimension by giving an empirical investigation of the law. Raimi (1976) studied in details the law of first digit.

In this article we have considered some of the natural sectors in the perspective of our country such as census data in the year 2011, length of the rivers of our country etc and analyzed the data of these sectors. We surprisingly observed that these data obey the Benford's law closely. We have organized our article in the following way: section 1 gives a brief introduction. Section-2 describes a literature review, section-3 provides a definition of Benford's law and in section-4 some applications along with the graphical representations are given. In section 5 &6 novelty and conclusion has been provided and a list of references is also given in section-7.

§3. Benford's Law

The formulas for the digit frequencies are shown next with D_1 representing the first digit, D_2 the second digit

$$\text{Prob}(D_1 = d_1) = \log_{10}(1+\frac{1}{d_1}) \quad \text{where } d_1 \in \{1,2.....9\}$$

$$\text{Prob}(D_1 = d_2) = \log_{10}(1+\frac{1}{d_2}) \quad \text{where } d_2 \in \{1,2.....9\}$$

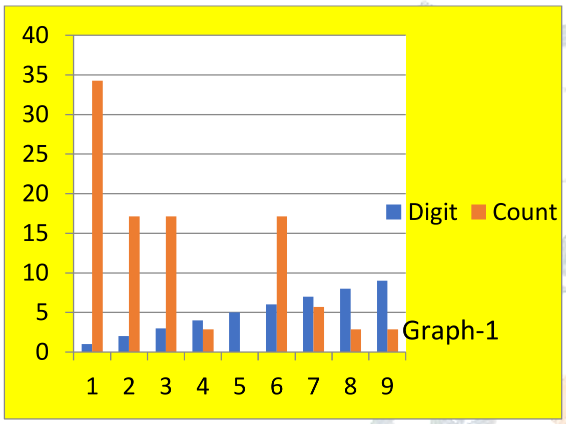
where Prob indicates the probability of observing the events in parentheses (Nigrini, 2012).

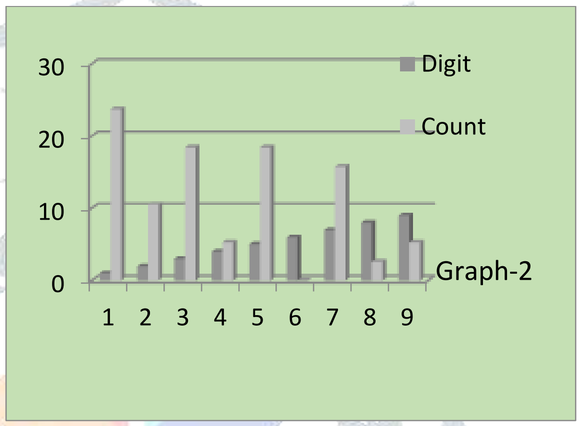
Using his formula, the probability that the first digit of a number is one is about 30% while the probability the first digit a nine is only 4.6 %.

The distribution law Benford is independent on the scale and base. As for example, if we collect the records of temperature of any area whatever may be the scale (i.e. Celsius, Fahrenheit or Kelvin) or measure the distance whatever may the unit (i.e. km, mile etc) all of those proves the truthfulness of the Benford's law. It is the great advantage of Benford's law. This law is well applicable in census data, rate of birth and date, length of rivers etc.

§4. Some applications:

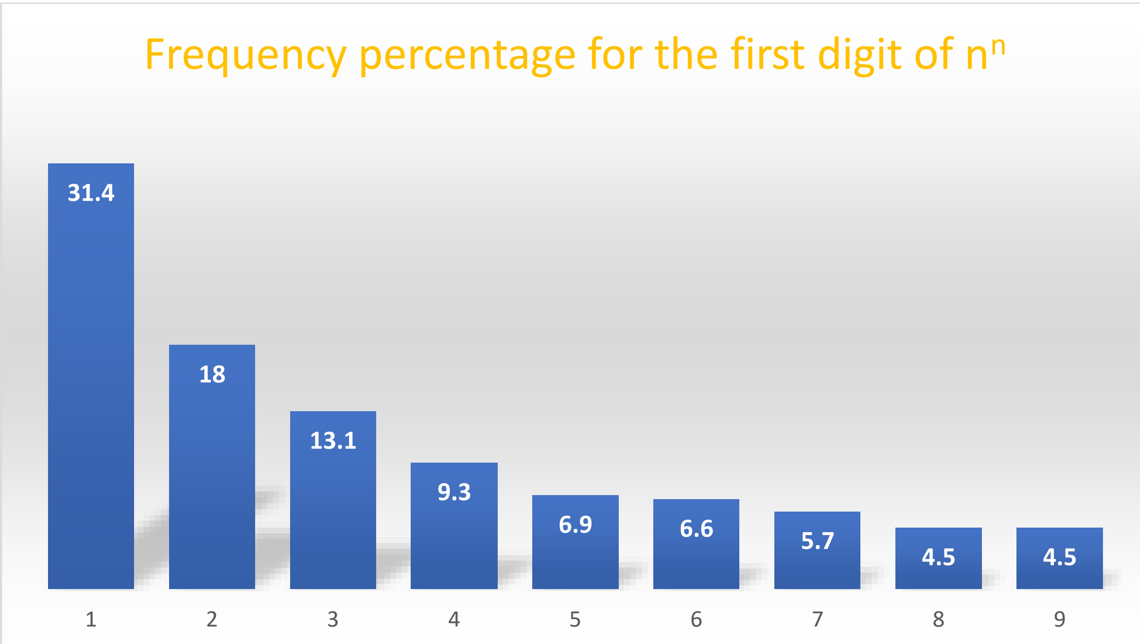
We consider the population of our country according to the census of 2011 and count the digits 1…9. We plot a graph based on it to study the digits '1' to '9' at the first position. With a great surprise **we** see that the graph astonishingly supports the Benford's law. The graph-1 represents the digits '1' to '9' at the first place according to the census of 2011 whereas the position of the same in case of the lengths of the different rivers of India is represented by the graph-2. We have considered here thirty-eight number of rivers in India.





But in everywhere the Benford's law is not applicable, i.e. there are some restrictions in the application of Benford's law. It is not effective for fewer number of data or for any artificial data created by men. As for example this distribution is not applicable for Cheque numbers, Challans, Pin code, Telephone numbers etc.In the early decade of nineteen eighty (after 1981) Nirgini enforced this law to detect the forgery. He applied the Benford's distribution to point out the forgery in election of the Ukraine Republic. At present it is widely used to determine duplicate payments, forgery in Income Tax, biased hypothesis, disparity in common ledger etc.We have seen that this distribution is very effective in case of power law. So, in the field of social network (Twitter, Facebook etc) this law is very effective.

The following graph supports our statement



Graph-3

Day-to-day examples of where it should apply are Electricity bills, Street addresses, Stock prices, Population numbers, Death rates, Lengths of rivers, Accounts payable invoice and payment values, General Ledger balances etc.

§ 5. Novelty and originality:

The application of Benford's law has a significant importance in counting digits at the first place. We have here shown that the census data of India, lengths of the rivers in our country (India) also follows the law of First digit (Benford,s law) which are the important observations. Moreover, we have graphically shown that this law is very close to the application of power function.

Benford's Law has now become a very active and attractive research area with numerous applications across several disciplines.

§ 6. Conclusion:

Benford law is very helpful to identify Duplicate payments (accounts payable), Fraudulent payments, Fraudulent expense claims, Tax return fraud, Biased estimation in General Ledger balances etc.Recently Benford's distribution is being applied in music analysis (whether it is original or not). Therefore, Benford's law is very relevant at the present situation.

§ 7. References:

1.Isabel Barbancho, Lorenzo J. Tard´on, Ana M. Barbancho, Mateu Sbert. (2015), "Benford's Law for Music Analysis", 16th International Society for Music Information Retrieval Conference.

2. P.S. Kruger & V.S.S. Yadavalli, (2017),'The power of one: Benford's law', South African Journal of Industrial Engineering, Vol 28(2), pp 1-13.

3. Arno Berger & Theodore P. Hill,' (2011) A basic theory of Benford's Law', Probability Surveys. Vol-8, pp. 1–126, DOI: 10.1214/11-PS175.

4. Ikbale Tota, Ada Aliaj, Jonida Lamcja, (2016),'The use of Benford's law as a tool for detecting fraud in accounting data', Interdisplinary Journal of Research and Development Vol -III (1), pp.73-77.

6.Cindy Durtschi, William Hillison & Carl Pacini,(2004), 'The effective use of Benford,s law to assist in detecting fraud in accounting data',Journal of Forensic Accounting,Vol-5,pp-17-34.

7. P. D. Scott & M. Fasli, 'Benford's Law: An Empirical Investigation and a Novel Explanation', CSM Technical Report 349, pp.1-21.

8. Nigrini, M. J. (1996). A Taxpayer Compliance Application of Benford's Law. Journal of the American Tax Association, Spring 1996 pp 72-91.

9. Newcomb, S. (1881). Note on the frequency of the use of the digits in natural numbers. Amer. Jour. Math. 4 pp 39-40.

10. Raimi, R. A. (1976). The First Digit Problem. American Mathematical Monthly. Vol.49, pp 521-538.