2012 Hawaii University International Conferences
Education, Math & Engineering Technology
July 31st to August 2nd
Waikiki Beach Marriot Resort & Spa
Honolulu, Hawaii

Mathematica Simulations of Benford Distributed Random Variables

C. Rasinariua, A. Khosravania
aColumbia College Chicago, Department of Science and Mathematics
600 S. Michigan Ave., Chicago, IL 60605, U.S.A

Mathematica Simulations of Benford Distributed Random Variables

C. Rasinariu[a,*], A. Khosravani[a]

[a] *Columbia College Chicago, Department of Science and Mathematics*
600 S. Michigan Ave., Chicago, IL 60605, U.S.A

Abstract

A random variable is Benford distributed if the occurrence frequency of its most significant d digit is $P_d = \log(1 + 1/d)$. Many empirical data sets obey this law with various degrees of accuracy. It is known that certain symmetries of the probability density function are sufficient conditions for exact Benford conformance. In this paper we identified new conditions for exact Benford distributions, and constructed novel examples of such random variables. The Mathematica simulations strongly support our theoretical results.

Keywords: Benford's law, first digit law, Mathematica simulations

Benford (1938) noticed that the pages of logarithm tables were more worn out for smaller digits such as 1 and 2 than for larger ones, and explicitly gave the formula for the probability of a number having the first digit d

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

He gathered substantial empirical evidence for formula (1) by collecting 20,229 numbers from diverse datasets, such as the area of the river beds, atomic weights of elements, etc., and generalized equation (1) to a formula for the probability of occurrence of an arbitrary sequence of digits $d_1, d_2, \cdots d_q$. In particular, the probability that a second-place digit d_2 is following a first-place digit d_1 is given by

$$P_{d_1 d_2} = \log_{10}\left(1 + \frac{1}{d_1 d_2}\right) \Big/ \log_{10}\left(1 + \frac{1}{d_1}\right) \ . \tag{2}$$

Thus the probability for the second-place digit 7 to follow the first-place digit 3 is

$$P_{37} = \log_{10}\left(1 + \frac{1}{37}\right) \Big/ \log_{10}\left(1 + \frac{1}{3}\right) = 0.0927 \ . \tag{3}$$

Some data from Benford's tables are rather poor examples of the logarithmic distribution (1), while others, are a very good fit. It is the union of all data

*Corresponding author: crasinariu@colum.edu

March 26, 2012

collected by Benford that follow closely enough the logarithmic distribution, Raimi (1976). These empirical observations led naturally to the question on finding random variables that respect *exactly* Benford's law.

Adhikari and Sarkar (1968) showed that if X is a uniformly distributed random variable, then X^n asymptotically approaches Benford's law when $n \to \infty$. Furthermore, they proved Benford conformance for the product of n uniformly distributed random variables $X_1, X_2, \dots, X_n$, when $n \to \infty$.

Hill (1995) proved that base invariance implies Benford distribution. In addition he showed that only Benford distributions are scale invariant. Base invariance means that for any base b, the probability of having d as the first digit is

$$P_b(d) = \log_b \left(1 + \frac{1}{d}\right) \ , \quad d = 1, 2, \dots, b-1 \ . \tag{4}$$

Scale invariance indicates that the probability of occurrence of the leftmost digits remains unchanged under a scalar multiplication. For example, suppose that the prices of goods are Benford distributed. Then, this is true, regardless of the currency in which those prices are converted.

Leemis, Schmeiser, and Evans (2000) provided examples of non-uniform random variable distributions X such that 10^X satisfies Benford's law exactly. Expanding on this work Balanzario and Sanchez-Ortiz (1991) derived sufficient conditions for a random variable X such that b^X satisfies Benford's law. They found a collection of symmetry requirements on X that will produce Benford distributed b^X random variables, thus enabling the construction of infinitely many examples of such distributions.

Although not universal, Benford's law has surprisingly wide applications in statistics, economics, engineering, and physics, as reflected by the large number of papers published in these areas. As an exhaustive reference list is beyond the scope of this paper, the interested reader should consult BOB (2011) and the references therein.

The scope of this paper is twofold. We theoretically constructed new examples of exact Benford random variables, and using Mathematica simulations we experimented with concrete numerical models. We generated random variables and determined the conditions that would make these variables follow Benford's law. Starting with a random variable X with the probability density function $g(x)$ we built a new random variable $Y = 10^X$ whose probability density function is $f(y) = g(\log y)/(y \ln 10)$. Then we generated the discrete approximation of Y using a large number of points. We extracted the first digit of each number in this set, and built the histogram of the corresponding probabilities. Finally we compared the resulting histogram with the probabilities of each digit as given by Benford's law. The Benford conformance of these random variables is quite remarkable.

One of the new examples of exactly distributed Benford variables is $Y = 10^X$ where the probability density function of X is the triangle $(0, k, 2k)$ with k a positive integer. Its cumulative distribution function is given by

$$F_Y(y) = \begin{cases} B_1^k(y) & 1 \le y < 10^k \\ 1 - B_2^k(y) & 10^k \le y < 10^{2k} \\ 1 & y \ge 10^{2k} \end{cases} ,$$

where $B_1^k(y) = (\log y)^2/2k^2$ and $B_2^k(y) = (\log y - 2k)^2/2k^2$.

This is a direct generalization of the known case for the triangle $(0, 1, 2)$. We first prove theoretically that the new class of variables is Benford distributed by proving that

$$P(Y = d) \equiv \sum_{i=-\infty}^{\infty} \left[F_Y(10^i(d+1)) - F_Y(10^i d)\right] = \log_{10}\left(1 + \frac{1}{d}\right) , \quad (5)$$

and provide Mathematica generated simulations that illustrate this phenomenon. In the following example we illustrate the case $k = 1$. We generated a discrete approximation of Y using 100 000 data points. Using formula (5) we have calculated the theoretical probability $P(Y = d)$ and compared it with the experimental probability and the exact Benford distribution. The results are recorded in Table 1. For a Benford distributed random variable such as triangle $(0, 1, 2)$ the theoretical probabilities match the column with the Benford probabilities. Note the accuracy of the experimental Mathematica generated numbers.

Digit	Experimental Probability	Theoretical Probability	Benford Probability
1	0.29995	0.30103	0.30103
2	0.17759	0.176091	0.176091
3	0.12727	0.124939	0.124939
4	0.09513	0.09691	0.09691
5	0.07673	0.0791812	0.0791812
6	0.06866	0.0669468	0.0669468
7	0.05797	0.0579919	0.0579919
8	0.05080	0.0511525	0.0511525
9	0.04590	0.0457575	0.0457575

Table 1: Probabilities for triangle $(0, 1, 2)$

References:

F. Benford, "The Law of Anomalous Numbers," Proceedings of the American Philosophical Society, Vol. 78, No. 4, pp. 551-572, 1938.

A. K. Adhikari, B. P. Sarkar, "Distribution of most significant digit in certain functions whose arguments are random variables," The Indian Journal of Statistics, Series B, Vol. 30, No. 1/2, pp. 47-58, 1968.

R. A. Raimi, "The First Digit Problem," The American Mathematical Monthly, Vol. 83, No. 7, pp. 521-538, 1976.

T. P. Hill, "Base Invariance Implies Benford's Law," Proceedings of the American Mathematical Society, Vol. 123, No. 3, pp 887-895, 1995.

T. P. Hill, "The significant-digit phenomenon," The American Mathematical Monthly, Vol. 102, No. 4, pp. 322-327, 1995.

L. M. Leemis, B. W. Schmeiser, and C. Evans, "Survival Distributions Satisfying Benford's Law," American Statistician Vol. 54 No.4, pp. 236-241, 2000

E. P. Balanzario, J. Sanchez-Ortiz, "Sufficient conditions for Benford's law," Statistics and Probability Letters, 2010, doi:101016/j.spl.2010.07.014

Benford Online Bibliography, `http://www.benfordonline.net`