

# Benford's Law in the Classroom and the Courtroom

Larry G. Blaine

Department of Mathematics, Plymouth State University, Plymouth NH, USA

[lblaine@plymouth.edu](mailto:lblaine@plymouth.edu)

## Abstract

Benford's Law has to do with the peculiar fact that in many situations in which uniform randomness might reasonably be expected, the frequencies of first (leftmost) digits in numerical data sets display a clear pattern in which 1 is most common, and 9 least common. Long considered a mere curiosity, this "law" has in recent years been applied in surprising ways, including the examination of financial data. In this paper we attempt to show how it may be used as a means of leading students to explore the vast and incompletely mapped territory where mathematics meets the world.

## Introduction

A recent technical paper (Pimbley, 2014) begins with the words "What the world calls Benford's Law is a marvelous mélange of mathematics, data, philosophy, empiricism, theory, mysticism, and fraud." In this paper we try to show how we might help students understand this marvelous mélange, and perhaps learn some wider lessons about the uses of mathematics in understanding the world. It is not at all our intent to give an exposition of the mathematical foundations of Benford's Law (henceforth BL). Our aim is pedagogical, and thus we have attempted to keep examples simple and technicalities to an irreducible minimum. For those wishing to learn more about the theory and application of BL, a vast literature is available (Anonymous, 2015). We mention in particular the recent books (Negrini, 2012) and (Kossovsky, 2015). In this paper we assume only a knowledge of base 10 logarithms, though a little skill in elementary analysis will be helpful at one point.

Before even stating BL, let us look at an example. When the author had first heard of BL, but before he had studied it, he decided to use some data from the 2000 US census as a test. Michigan, where he was a student long ago, is one of 50 states in the US. It is divided into 83 districts called *counties*. Now, on the face of it, it would seem bizarre to suppose that the first digits of the populations of these counties should conform to any particular pattern at all. Furthermore, a data set of 83 numbers is rather small. Nevertheless, among first digits, 23 are 1's, 21 are 2's, 9 are 3's and 7 are 4's. Digits 5 to 9 each occur fewer than 7 times. Is this a coincidence? The reader is invited to examine some similar population data - the

counties of Great Britain, the 150 largest cities and towns of Bangladesh, anything of that sort. In nearly every case, there will be a strong nonuniformity in first digits, with the smaller digits predominating. Why? There truly is something here in need of explanation.

### What Does Benford's Law Say?

A large set of positive numbers conforms to BL if, when looking at only the first digits of these numbers, 1 occurs most frequently, 2 next most frequently, 3 third most frequently, etc. By *first digit* we mean leftmost nonzero digit, e.g. the first digit of 47000 is 4, and that of .00074 is 7. Precisely, BL is satisfied if among first digits, the proportion of d's is  $\log(d + 1) - \log(d)$ . Here  $d = 1, 2, 3, 4, 5, 6, 7, 8, \text{ or } 9$ , and  $\log$  refers to logarithm base 10. Rounded numerical values are:

First digit	d		1		2		3		4		5		6		7		8		9		
Proportion			.301		.176		.123		.097		.079		.067		.058		.051		.046		

From this we may get some idea of the sorts of data sets conforming to BL. (Any reader wishing to avoid technicalities is encouraged to go immediately to the next paragraph.) We may say that a data set with values from from 1 to 10 conforms exactly to BL if it has probability density function  $P(x) = 1/x$ . If the range of data values is larger, say 1 to 100, by some elementary calculations, we can show that a density function in reciprocal form will imply BL. By similar calculations, we may show that BL is satisfied whenever the mantissas (fractional parts) of the logarithms of the data values are uniformly distributed.

The above description of BL is precise, but alas not very helpful. After all, when faced with data from the world, we usually do not have any kind of specific probability distribution to rely on. We can however make the rough general statement: *If the data ranges over several orders of magnitude, and there is a strong tendency throughout the set for smaller values to be more common than larger ones, then it is likely that BL will be satisfied, at least approximately.*

Let us briefly revisit Michigan. There, as elsewhere on Earth, populations tend to be distributed very unevenly, with a few large concentrations (large cities, say), more medium concentrations (mid-sized towns and cities) and most commonly, small concentrations (villages or rural regions, say.) The counties of Michigan display this pattern nicely. The largest, Wayne County, contains the city of Detroit and had population about 2,100,000, while the smallest, Keweenaw County, contains nothing and had population about 2,300. The data thus ranges over about

three orders of magnitude. The first-digit results are in reasonable conformity with BL, given the relatively small size of the set.

### **An Important Case: Exponential Growth and Decay**

Numerous processes are characterized by an exponential law: the quantity under consideration increases or decreases by some fixed ratio whenever the variable, often time, increases by one unit. These processes include the calculation of interest continuously compounded, radioactive decay, certain population models, the chromatic scale of music theory, and very many others. To fix ideas, consider the simplest possible case: a population begins at 1 and increases by a factor of 10 over every time unit. (We do not care what the units are, of either time or population.) The population at time  $t$  is then  $P(t) = 10^t$ , and we suppose that our data consists of values measured at equally spaced small intervals of time- perhaps  $P(0)$ ,  $P(0.001)$ ,  $P(0.002)$ , etc. To further simplify, suppose at first that  $0 \leq t < 1$ , so that  $1 \leq P(t) < 10$ . Now, to ask what proportion of data values have first digit 1 is to ask how long the population remains between 1 and 2. To be precise,  $1 \leq P(t) < 2$  is equivalent to  $\log(1) \leq t < \log(2)$ , so the proportion we want is  $\log(2) - \log(1) \approx .301 - 0 = .301$ . In the same way we see that the proportion having first digit 2 is  $\log(3) - \log(2) \approx .176$ , etc.

This example not only illustrates BL, but in a sense embodies it. A student who understands it has grasped the essence. Calculations of various degrees of difficulty may be applied to demonstrate that for exponential growth/decay, BL holds over other time spans, for bases other than 10, etc. What is astonishing is that BL still holds for some processes that are exponential only in quite imprecise ways. Among many possible illustrations, a favorite is the following casino game, described in (Kossovsky, 2015), because it may easily be played in the classroom. The only equipment necessary is one die and a scientific calculator.

A single die is to be rolled many times, and a record kept of the cumulative products. (For example, for the five rolls 2, 5, 6, 1, 4, the products are 2, 10, 60, 60, 240.) Since the mean of the possible outcomes 1, 2, 3, 4, 5, 6 is 3.5, the products in some very vague sense should resemble the exponential process  $(3.5)^n$ , where  $n$  is the number of rolls. BL nevertheless prevails! The specific game involves 140 rolls. The player wins if the last product is at least  $10^{70}$  and the casino wins otherwise. The data set to be examined for conformity to BL consists of the 140 products. (In actual play, it is likely that neither the casino nor the player cares much about BL. The question of how probable it is that the player will win is interesting, but does not concern us here. Let us just say, not very probable at all.)

## Applications of Benford's Law

Now that we have seen a little of BL, we turn to the question of its application. For a long time BL stood as a striking but rather innocent mathematical oddity. In recent years, however, many uses have been proposed in fields as diverse as mathematical biology, digital image processing, and especially, finance. Indeed, the book (Nigrini, 2012) is largely devoted to financial matters, and serves as a general reference for this section. These applications have the following in common: If a set of data "ought to" conform to BL but does not in fact do so, then there is an anomaly to be explained. The anomaly may be a simple mistake, it may be caused by some social or physical factor (For example, cities may lose many of their inhabitants in times of war, or mass emigration, or plague, leading to distortion of population data), or it may be plain fraud. It is in this latter area that BL meets the law. We give no specifics, but the reader may find in (Nigrini, 2012) details of such lurid financial scandals as Enron, Madoff, and LIBOR. In these cases and many more, reported fraudulent financial data is displayed, and deviations from BL noted. Use of BL as an investigative tool is one thing, but its use as *direct* evidence in court is quite another. This is not common, but has occurred in both England and the US. In such cases, whether or not a data set should be expected to conform to BL is a legal and moral question, and not a purely statistical one.

## Historical Note

BL is named for physicist Frank Benford, who wrote about first-digit distributions in 1938. (Benford, 1938) He became interested in the subject when he noticed uneven wear on the pages of a table of logarithms. He was unaware, and so was everyone else, of Simon Newcomb's much earlier paper. (Newcomb, 1881)

## References

- Anonymous (2015): Benford Online Bibliography, *www.benfordonline.net*
- Benford, F. (1938): The Law of Anomalous Numbers, *Proceedings of the American Philosophical Society*, **78** (4): 551-572.
- Kossofsky, A. (2015): *Benford's Law*. Singapore: World Scientific.
- Newcomb, S. (1881): Note on the Frequency of Use of the Different Digits in Natural Numbers, *American Journal of Mathematics*, 4 (1): 39-40
- Nigrini, M. (2012): *Benford's Law: Applications for Forensic Accounting, Auditing, and Fraud Detection*. Hoboken: Wiley and Sons
- Pimbley, J. M. (2014): Benford's Law and the Risk of Financial Fraud, *Quant Perspectives*, *www.garp.org*