

# Punk Rock Operations Research

\*\*\* operations research or die! \*\*\* a blog by Laura Albert McLay

February 21, 2012

## the first digits of many “random” numbers are not so random: on Benford’s law

By Laura Albert McLay

I asked for an example of a uniform random variable in a class last week. One of the students came up with a reasonable guess: the leading digits of city populations. The sizes of different cities seems more or less “random,” meaning that there should be no rhyme or reason in determining the first digits.

For those of you familiar with Benford’s Law, you will know that the leading digits of city populations do not follow a (discrete) uniform distribution between 1 and 9. Benford’s law explains the true distribution. It was discovered in the 19th century when number crunchers relied on logarithm tables. Astronomer Simon Newcomb noticed that the pages in a logarithm table book were not evenly worn: pages starting with a 1 were worn much more than pages starting with a 9. He published a paper on this phenomenon in the *American Journal of Mathematics*. Newcomb argued that the probability that the first digit  $d$  is:

$$P(\text{first digit} = d) = \log_{10} (1 + 1/d), d = 1, 2, \dots, 9.$$

These results suggest that a 1 is the first digit 30.1% of the time and that a 9 is the first digit 4.6% of the time (not  $1/9$  as one would expect if the numbers were uniformly distributed). Physicist Frank Benford discovered the same thing 57 years later and published his identical conclusions in the *Proceedings of the American Philosophical Society* after verifying that the distribution applied to many diverse data sets.

Some of the data sets that behave according to Benford’s law are: tax returns, stock market prices, baseball statistics, and numbers that appear on the front pages of newspapers.

Not all data sets follow Benford’s law—an example would be phone numbers.

What is the explanation for Benford’s law? First of all, it is impossible to “prove” Benford’s law, since it does not apply to every data set, but one can try to *explain* it. There have been many explanations over the years. Before 1995, the explanations did not result in valid probability distributions, i.e.,

$$p(1) + p(2) + p(3) + \dots + p(9) = 1.$$

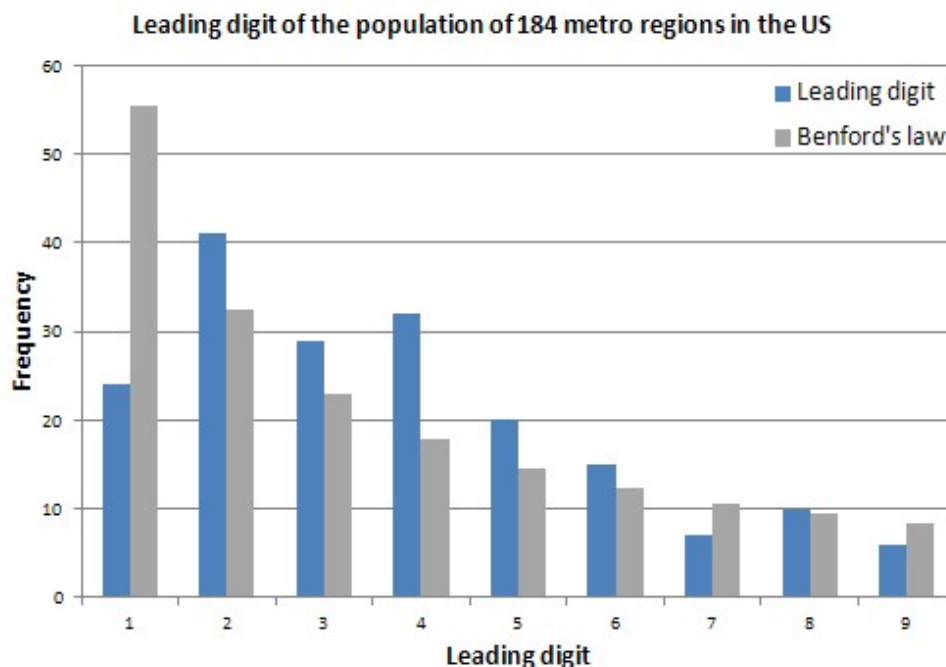
TP Hall gives a good explanation in his fantastic [Scientific American](http://www.americanscientist.org/issues/id.815.y.1998.no.4.content.true.page.1.css.print/issue.aspx)

(<http://www.americanscientist.org/issues/id.815.y.1998.no.4.content.true.page.1.css.print/issue.aspx>) article:

If distributions are selected at random (in any “unbiased” way) and random samples are taken from each of these distributions, then the significant-digit frequencies of the combined sample will converge to Benford’s distribution, even though the individual distributions selected may not closely follow the law.

For example, suppose you are collecting data from a newspaper, and the first article concerns lottery numbers (which are generally uniformly distributed), the second article concerns a particular population with a standard bell-curve distribution and the third is an update of the latest calculations of atomic weights. None of these distributions has significant-digit frequencies close to Benford’s law, but their average does, and sampling randomly from all three will yield digital frequencies close to Benford’s law.

Let’s come back to my example on population sizes. Below, I plot the frequency of leading digits for the 184 metro regions in the United States according to the 2010 census (blue) next to what one would expect according to Benford’s law (gray). It’s not a perfect match, but it’s clear that Benford’s law provides a better fit than a discrete uniform distribution.



([https://punkrockor.files.wordpress.com/2012/02/screenhunter\\_01-feb-16-13-46.jpg](https://punkrockor.files.wordpress.com/2012/02/screenhunter_01-feb-16-13-46.jpg))

A comparison of the leading digit of census data to Benford's law

This entry was posted on Tuesday, February 21st, 2012 at 8:51 am and posted in Uncategorized. You can follow any responses to this entry through the RSS 2.0 feed.

2 responses to “the first digits of many “random” numbers are not so random: on Benford’s law”

- [matforddavid](#)

February 23rd, 2012 at 3:35 pm

A less-publicised corollary of Benford’s Law is that the second digits (and the third, and ...) are not uniformly distributed either.

◦ *Jeffrey W. Herrmann*

February 29th, 2012 at 7:35 am

If you cannot access the article through the link, you may be able to find it using the following citation: Hill, T.P., “The first digit phenomenon.” *American Scientist*; Jul / Aug98, Vol. 86 Issue 4.

[Subscribe to RSS](#)

Blog at WordPress.com.