A Benford Walk Down Wall Street

Xixi Edelsbrunner, Karen Huan, Blake Mackall, Jasmine Powell, Madeleine Weinstein

Advisor: Steven J. Miller

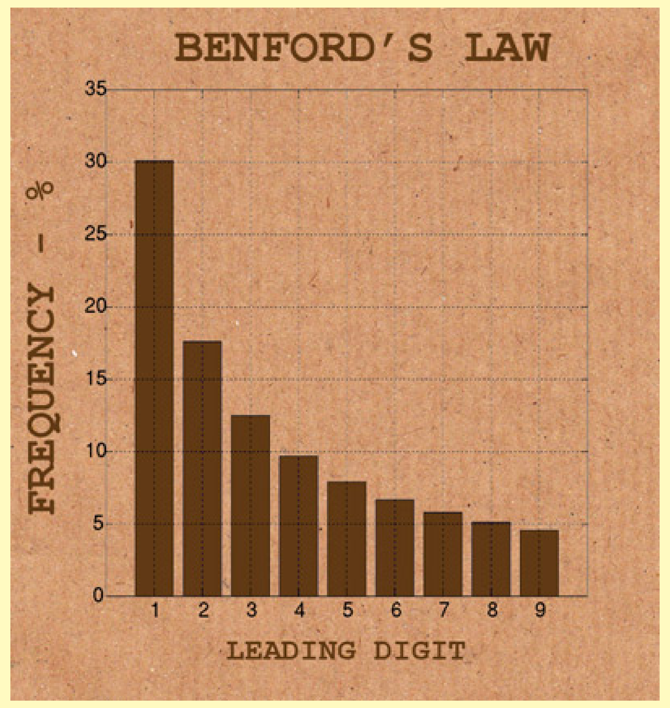
SMALL 2014, Williams College

Intro and Definitions

Benford's Law

In many naturally arising data sets, the frequency of first digits follows a distribution known as Benford's Law. In such sets, a first digit of 1 occurs about 30% of the time, whereas a first digit of 9 occurs only 4.5% of the time. Some examples of systems following Benford's Law are:

- Fibonacci numbers
- Lengths of rivers
- Death rates
- Mathematical constants



Formal definition

Weak: A system is Benford base B if the probability of seeing a first digit d is $\log_B \frac{d}{d+1}$.

Strong: A system is Benford base B if the probability of seeing a significand at most s is $\log_B s$.

Background

Speculative Stock Prices

Burton Gordon Malkiel's random walk hypothesis of markets predicts that stock market prices follow a random walk and are thus distributed according to a Gaussian. In *The Variation of Certain Speculative Prices*, Benoit Mandelbrot observed that fluctuations in price data were too variable to be modeled with a normal distribution and instead are better modeled by stable distributions, including the Cauchy and Lévy distributions. Many of these distributions have infinite variance, which Mandelbrot believed the stock prices exhibited.

First digit frequencies from Cauchy and Lévy distributions are almost Benford.

The Lévy Distribution

The Lévy P.D.F. centered at 0 is given by

$$p(x;\mu,c) = \sqrt{\frac{c}{2\pi}}\,\frac{e^{c/(2x)}}{x^{3/2}}$$

where c is a scale parameter. Using methods described in "Approach," we found that the fluctuation from Benford (which we measure as the maximum difference between the derivative and 1) for a given B is

$$E_{B,c} \le 2\sqrt{2}\,\frac{e^{-\pi^2/\log B}}{1-e^{-\pi^2/\log B}}$$

, which for $B = 10$ is ≈ 0.039449143.

Approach

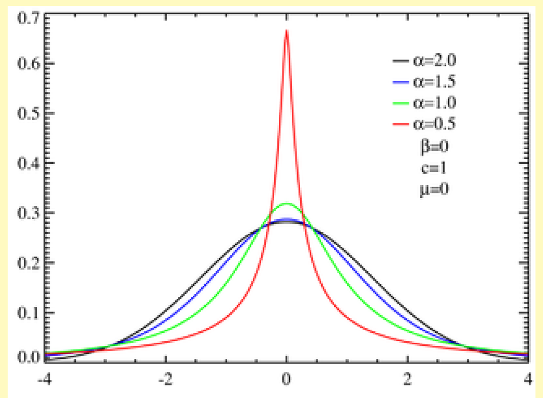
Let $p(x)$ be a probability density function for a random variable X. An equivalent definition of X being Benford base B is if its logarithm base B is equidistributed modulo 1, or if

$$\sum_{n=-\infty}^{\infty} \int_n^{n+b} p(B^y) B^y \log B \, dy = b.$$

Taking the derivative with respect to b of both sides, we looked at the derivative of the logarithm base B modulo 1, and we quantified how far away it is from 1 (If the derivative is 1, we know that the logarithm base B modulo 1 is equidistributed, and the system is Benford). For the Cauchy and Lévy distributions, we were able to take the Fourier transform of the P.D.F. and use Poisson's Summation to evaluate the infinite sum and quantify how close it is to 1. Poisson's Summation states that, if $\hat{f}(y)$ is the Fourier transform of $f(x)$, then

$$\sum_{n=-\infty}^{\infty} \hat{f}(n) = \sum_{k=-\infty}^{\infty} f(k).$$

These forms of the derivative decay very rapidly as n increases. And, as the original function is a probability density function, the Fourier transform evaluated at $n = 0$ is equal to 1. Thus, this form is better suited to quantify and bound how close the derivative is to 1.



The Cauchy Distribution

The Standard Cauchy Probability Density function is given by

$$p(x) = \frac{1}{\pi(1+x^2)}.$$

We consider the more general function

$$f_{a,r}(x) = \frac{1}{1+\left(\frac{x}{a}\right)^r},\ 0 < a < \infty,\ 1 < r < \infty$$

which we can turn into a probability density function with the right constant. By transforming this P.D.F. and using Possoin summation, we find that the derivative is very close to 1, where the maximum fluctuation given a base B and fixing a and r is

$$E_{B,a,r} = \sin\left(\frac{\pi}{r}\right) \sum_{n\in\mathbb{Z},n\neq 0} \csc\left(\frac{\pi}{r}(1-2\pi i n)\right)$$

In the special case of the Cauchy Distribution base 10, ($a = 1, r = 2, B = 10$), we find that the fluctuation from 1 is less than 0.05578. So the Cauchy Distribution is very close to Benford! Furthermore, we see that in the limit as r approaches 1, our P.D.F. becomes arbitrily close to Benford.

Digit	Observed Cauchy %	Predicted Benford %
1	30.908	30.103
2	17.130	17.609
3	11.819	12.494
4	9.340	9.691
5	7.924	7.918
6	6.857	6.695
7	6.065	5.799
8	5.252	5.115
9	4.705	4.576

Acknowledgements

This research was conducted as part of the 2014 SMALL REU program at Williams College and was partially supported funded by NSF grant DMS1347804, Williams College and the Clare Boothe Luce Program of the Henry Luce Foundation. We also thank Mark Nigrini for the problem idea and topic discussions, as well as our colleagues at the SMALL 2014 REU for many helpful conversations.