Benford's Law and Accelerated Growth

MATHEMATICAL CURIOSITIES
By Mark Levi

Benford's law, discovered by Simon Newcomb [2], is an empirical observation that in many sets of numbers arising from real-life data, the leading digit of a number is more likely to be 1 than 2, which in turn is more likely than 3, and so on. Figure 1 shows a count of leading digits from the 107 NASDAQ-100 prices,[1] sampled around noon on June 14, 2017. Although the monotonicity is clearly violated, there is a bias in favor of low leading digits.

The following example illustrates a mathematically rigorous deterministic (as opposed to probabilistic) counterpart of Benford's law. Consider a geometric sequence, for instance

$$1, 2, 4, 8, 16, 32, 64, 128, 256, \ldots,$$

and extract from it the sequence of leading digits:

$$1, 2, 4, 8, 1, 3, 6, 1, 2, \ldots.$$

[1] http://www.cnbc.com/nasdaq-100/

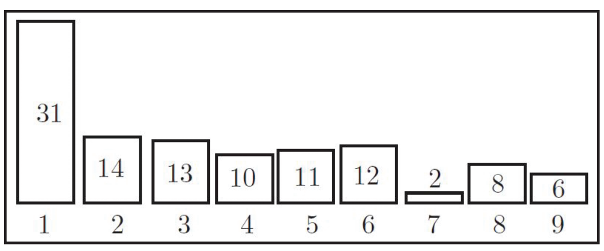

Figure 1. *A snapshot of the distribution of leading digits in NASDAQ-100 stock prices.*

It turns out that the frequency p_k of digit k is well defined and given by

$$p_k = \lg(k+1) - \lg k, \tag{1}$$

see [1]. In particular, the frequency decreases with k:

$$p_1 = \lg\frac{2}{1} > \lg\frac{3}{2} > \ldots > \lg\frac{10}{9} = p_9. \tag{2}$$

In light of this example, if the price of a stock undergoes an exponential-like growth (in a loose analogy with the geometrical sequence), then the bias illustrated in Figure 1 may not be that surprising.

What is behind Benford's frequency bias (2) for the geometric series? A one-word answer is "acceleration." To see why, consider a continuous counterpart of 2^n — say, the exponential function e^t, visualizing the point $x = e^t \in \mathbb{R}$ as a particle moving with time along the x-axis.

Figure 2 shows the x-axis cut into segments $[10^j, 10^{j+1})$ and stacked on top of each other, all of them scaled (linearly) to the same length. This cutting and scaling allows us to see the leading digit of e^t at a glance. Now the reason for Benford's law becomes clear: *since e^t accelerates, it passes the higher-digit segments faster, and thus is less likely to be found there.*

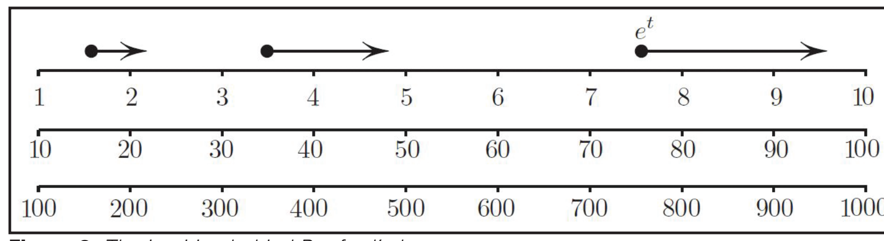

Figure 2. *The intuition behind Benford's law.*

Figure 2 makes it easy to compute the probability, i.e., the proportion of time, of observing the leading digit k. To that end, we find the time spent having the leading digit k while traversing the jth row in Figure 2:

$$\ln[(k+1)10^j] - \ln[k10^j] = \ln\frac{k+1}{k}.$$

We then divide it by the time of traversal $\ln[10 \cdot 10^j] - \ln 10^j = \ln 10$; both times are independent of j, and thus the proportion of time spent with the leading digit k over time $[0, T]$ approaches

$$\frac{\ln\frac{k+1}{k}}{\ln 10} = \lg\frac{k+1}{k},$$

as $T \to \infty$, the same as the discrete result (1).

The figures in this article were provided by the author.

References

[1] Arnold, V.I. (1983). *Geometrical Methods in the Theory of Ordinary Differential Equations*. New York, NY: Springer-Verlag.

[2] Newcomb, S. (1881). Note on the frequency of use of the different digits in natural numbers. *American Journal of Mathematics, 4*(1), 39-40.

Mark Levi (levi@math.psu.edu) is a professor of mathematics at the Pennsylvania State University.