Benford’s Law

Michael Walker, July 25, 2011

Benford’s Law, simply states that the first digit in a set of data occurs at a predictable probability.

I.e. the leading digit *d* (*d* ∈ {1, …, 9}) occurs with probality

$$P(d) = \log_b(d+1) - \log_b(d) = \log_b\left(1 + \frac{1}{d}\right).$$

Where b=10 in decimal (base 10) math. This may be counterintuitive since

P(1 being the first significant digit) =0 .301

Is much more likely than

P(9 being the first significant digit)=0.046.

This law is difficult to prove because it is a statement about data found in empirical sets (or tables). However the logarithmic property of “scale invariance” (like multiplying the table by scalar) gives support to the intuitive idea of “Base Invariance”. I.e. it stands to reason that if the law holds in decimal notation, it should hold in bi-nary and tri-nary math.

The occurrence of the ‘first-digit phenomenon’ is frequent but not absolute. There are recent theories that the Union of all data sets approaches Benford’s. There are further generalizations of Benford’s law to the second digit, third digit, and so on.

References:

The First Digit Phenomenon: A century-old observation about an unexpected pattern in many numerical tables applies to the stock market, census statistics and accounting data
Author(s): T. P. Hill
Source: American Scientist, Vol. 86, No. 4 (JULY-AUGUST 1998), pp. 358-363

The Significant-Digit Phenomenon
Author(s): Theodore P. Hill
Source: The American Mathematical Monthly, Vol. 102, No. 4 (Apr., 1995), pp. 322-327