THE BISECTION METHOD OBEYS BENFORD'S LAW

by

Jesse Campbell

1. Introduction

Benford's Law is a statistical observation that the leading digits in many real-life data sets is likely to be small. This observation was first recorded by Simon Newcomb in 1881, who noticed that in books which contained logarithmic tables, "the first pages wear out faster than the last ones". [**New81**] Later in 1938, the phenomenon was re-discovered by Frank Benford, who compiled over 20 tables containing over 20,000 data points supporting the law. [**Ben38**]

Digit	Probability $P(d)$	Relative Size of $P(d)$
1	0.301	
2	0.176	
3	0.125	
4	0.097	
5	0.079	
6	0.067	
7	0.058	
8	0.051	
9	0.046	

FIGURE 1. Visualization of probability of each leading digit occuring according to Benford's Law.

Benford's Law is particularly applicable to the field of numerical analysis as arithmetic operations performed by computers transform distributions towards Benford's Law. That being said, it is relatively taken for granted within the field, or as Richard Hamming wrote in 1970, [**Ham70**]

> Since floating point numbers are the basis of most of numerical analysis one may well ask why this obvious and experimentally well-verified distribution (Benford's Law) is so often ignored. It is because it appears to contradict the usually accepted model of the number system in which numbers correspond to points on a homogeneous straight line? Not only are the floating point numbers not uniformly spaced in a computer, but [Benford's Law] shows that even in intervals in which the numbers are equally spaced they are not equally likely to occur.

Likewise, in 1968, Donald Knuth remarked, [**Knu68**]

> [I]n order to analyze the average behavior of floating-point arithmetic algorithms, we need some statistical information that allows us to determine how often various cases arise... [If, for example, the] leading digits tend to be small [, that] makes the most obvious techniques of "average error" estimation for floating-point calculations invalid. The relative error due to rounding is usually... more than expected.

That being said, the unequal distribution of numbers is largely overlooked in the study of numerical analysis, both in its prevalence and applicability.

In this note, we will review the application of Benford's Law to numerical methods through the very well-known root-finding problem. In specific, we will analyze the approximations generated by the Bisection method for conformity with Benford's Law.

2. Main Result

Suppose we are given a continuous function f on the interval $[a, b]$ with root $f(p) = 0$ and p is contained within $[a, b]$. Let $p_n = (b + a)/2$. Suppose $p \neq p_n$, if $p \in [a, p_n]$, set $b = p_n$, otherwise, set $a = p_n$ and repeat this procedure.

This method is known as the Bisection method and it produces a series of approximations p_n which will converge to a root p for almost any continuous function.

It is relatively simple to show that the approximations generated by the Bisection method follow Benford's Law, or to be precise,

Theorem 2.1. — *If p is a root of $f : [a, b] \to \mathbb{R}$ and $p \in [a, b]$, $(p_{n+1} - p_n)$ is Benford.*

Proof: A well-known result in the theory of Benford's Law is that a sequence (x_n) of real numbers in Benford if and only if

$$\langle \log_{10} |x_n| \rangle \text{is uniformly distributed modulo 1 } (\textit{u.d. mod 1})$$

where $\langle x \rangle$ denotes the decimal part of x in base-10 (i.e. $\langle \pi \rangle = 0.14159...$).

Furthermore, for the series of approximations from the Bisection method, p_n, it is not hard to see that,

$$|p_{n+1} - p_n| = (b-a)2^{-(n+1)}$$
$$\implies \log_{10} |p_{n+1} - p_n| = \log_{10}((b-a)2^{-(n+1)}) = \log_{10}(b-a) - (n+1)\log_{10}(2)$$

Let $d_n = \log_{10} |p_{n+1} - p_n|$, then,

$$\lim_{n\to\infty} (d_{n+1} - d_n) = -\log_{10}(2)$$

A known result in the theory of uniform distribution of sequences states that, [**KN74**]

Lemma 2.2. — *If* $\lim_{n\to\infty}(d_{n+1} - d_n) = \theta$ *for some irrational* θ*, then* d_n *is u.d. mod 1.*

Since $-\log_{10}(2)$ is irrational, we have shown that the sequence (d_n) is *u.d. modulo 1.* However, this implies by the definition of *u.d. modulo 1* that the sequence $\langle d_n \rangle = \langle \log_{10} |p_{n+1} - p_n| \rangle$ is *u.d. modulo 1.*

Therefore the sequence $(p_{n+1} - p_n)$ is Benford. □

3. Consequence for Roundoff Error

On the note of Knuth's observation, we can obtain a rough idea of the magnitude of difference between expected roundoff error when Benford's law is and isn't assumed through the following method.

Let X denote absolute roundoff error and Y the fractional part of the approximation at the time of stopping such that it is between $[1/10, 1)$. Then, the expected relative error is given by $E(X) \times E(1/Y)$. If the approximation Y is assumed to be uniformly distributed, then,

$$E(X) \times E(1/Y) = E(X) \times \int_{0.1}^{1} \frac{10}{9}\frac{1}{t} dt \approx 2.558 \times E(X)$$

However if Y is Benford, then the true average relative error is,

$$E(X) \times E(1/Y) = \int_{0.1}^{1} \frac{1}{t\ln(10)} \frac{1}{t} dt \approx 3.909 \times E(X)$$

Therefore we underestimate the relative error of the algorithm by $2.558/3.909 \approx 34.5\%$ when we naïvely ignore the fact that Y is Benford.

Here we utilized the "Benford Distribution" or "Reciprocal Distribution". We can find this distribution by first realizing that a Benford's Law can be restated as:

$$\mathbb{P}(S(X) \leq t) = \log_{10}(t)$$

Where $0.1 \leq S(X) < 1$ is the significand or mantissa of X and $\mathbb{P}$ is a probability function. Furthermore,

$$log_{10}(t) = \frac{\ln(t)}{\ln(10)} = F_{S(X)}(t)$$

Where $F_{S(X)}(t)$ is a cumulative probability density function. Therefore, the probability density function $f_{S(X)}(t)$ is given by,

$$f_{S(X)}(t) = \frac{d}{dt} F_{S(X)}(t) = \frac{1}{t\ln(10)}$$

.

References

[Ben38] F. Benford – "The law of anamalous numbers", *Proceedings of the American Philosophical Society* **78** (1938), p. 551–572.

[Ham70] R. W. Hamming – "On the distribution of numbers", *The Bell System Technical Journal* **49** (1970), p. 1609–1625.

[KN74] L. Kupiers & H. Niederreiter – *Uniform distribution of sequences*, John Wiley & Sons, 1974.

[Knu68] D. Knuth – *The art of computer programming*, Addison-Wesley, 1968.

[New81] S. Newcomb – "Note on the frequency of use of the different digits in natural numbers", *American Journal of Mathematics* **4** (1881), p. 39–40.

October 7th, 2023

Jesse Campbell • *E-mail :* `jesse.campbell@duke.edu`