

Newcomb-Benford Law

STEVEN FINCH

August 17, 2011

The literature for Benford's law is quite large and growing [1]; we avoid interesting foundational issues [2, 3] and turn attention instead to a specific scenario [4, 5, 6, 7].

Let $\{a_n\}_{n=0}^\infty$ be an m^{th} order linear homogeneous recurrence. Consequently the sequence can be written as

$$a_n = p_1 q_1^n + P_2(n) q_2^n + P_3(n) q_3^n + \cdots + P_m(n) q_m^n$$

where q_1, q_2, \dots, q_m are associated eigenvalues; q_1 is the largest eigenvalue (in absolute value); p_1 is constant and P_2, P_3, \dots, P_m are polynomials. The sequence $\{a_n\}$ is called **random-enough** if q_1 is real, positive, not a rational power of 10, of multiplicity 1 (as an eigenvalue) and p_1 is positive. Famous integer examples include

$$a_n = 2^n \quad (\text{powers of 2}),$$

$$a_n = \frac{\varphi^n - (1 - \varphi)^n}{\sqrt{5}} \quad (\text{Fibonacci sequence}),$$

$$a_n = \varphi^n + (1 - \varphi)^n \quad (\text{Lucas sequence})$$

where φ is the Golden mean [8].

Consider the j^{th} leftmost decimal digit D_j of an integer a . If $j = 1$, then $1 \leq D_j(a) \leq 9$; if $j \geq 2$, then $0 \leq D_j(a) \leq 9$. Let $\{a_n\}_{n=0}^\infty$ be a random-enough sequence of positive integers. Benford's law states that [4]

$$\begin{aligned} \lim_{N \rightarrow \infty} \frac{1}{N} \# \{n \leq N : D_1(a_n) = d\} &= \log_{10} \left(1 + \frac{1}{d} \right) \\ &= \sum_{k=0}^0 \log_{10} \left(1 + \frac{1}{10^k + d} \right) \end{aligned}$$

for $1 \leq d \leq 9$. In words, the first digit of an arbitrary term a_n is not uniformly distributed over $\{1, 2, \dots, 9\}$, but instead favors small values:

$$\text{P} \{D_1 = 1\} = 0.30103\dots, \quad \text{P} \{D_1 = 2\} = 0.17609\dots, \quad \text{P} \{D_1 = 3\} = 0.12493\dots$$

⁰Copyright © 2011 by Steven R. Finch. All rights reserved.

and, of course, $P\{D_1 = 0\} = 0$.

Fix $j \geq 2$. A generalization of Benford's law states that [4]

$$\lim_{N \rightarrow \infty} \frac{1}{N} \# \{n \leq N : D_j(a_n) = d\} = \sum_{k=10^{j-2}}^{10^j-1} \log_{10} \left(1 + \frac{1}{10k+d} \right)$$

for $0 \leq d \leq 9$. The second digit of an arbitrary term a_n is not uniformly distributed over $\{0, 1, \dots, 9\}$:

$$P\{D_2 = 1\} = 0.11389\dots, \quad P\{D_2 = 2\} = 0.10882\dots, \quad P\{D_2 = 3\} = 0.10432\dots$$

and $P\{D_2 = 0\} = 0.11967\dots$; each of the probabilities are, however, closer to $1/10$ than before. The same is true for the third digit of an arbitrary term a_n :

$$P\{D_3 = 1\} = 0.10137\dots, \quad P\{D_3 = 2\} = 0.10097\dots, \quad P\{D_3 = 3\} = 0.10057\dots$$

and $P\{D_3 = 0\} = 0.10178\dots$. Such numerical results were first tabulated in [9, 10]. For simplicity, we henceforth refer to Benford's law and its generalization together ($j \geq 1$) as NBL.

Another way to illustrate the approach to uniformity (as $j \rightarrow \infty$) makes use of moments. It is straightforward to show that [11]

$$E(D_1) = 2 \log_{10}(2) - 4 \log_{10}(3) + 8 \log_{10}(5) - \log_{10}(7) = 3.4402369671\dots,$$

$$E(D_1^2) = 8 \log_{10}(2) - 50 \log_{10}(3) + 72 \log_{10}(5) - 13 \log_{10}(7),$$

$$\text{Var}(D_1) = E(D_1^2) - E(D_1)^2 = 6.0565126313\dots;$$

$$E(D_2) = 4.1873897069\dots, \quad \text{Var}(D_2) = 8.2537786232\dots;$$

$$E(D_3) = 4.4677656509\dots, \quad \text{Var}(D_3) = 8.2500943647\dots$$

The means approach $9/2$ and the variances approach $33/4$, as anticipated. We also have

$$\text{Cov}(D_1, D_2) = E(D_1 D_2) - E(D_1) E(D_2) = 14.8019478993\dots,$$

for example. Correlation coefficients are small but positive; the largest is

$$\rho(D_1, D_2) = \frac{\text{Cov}(D_1, D_2)}{\sqrt{\text{Var}(D_1)} \sqrt{\text{Var}(D_2)}} = 0.0560563403\dots$$

It is further known that the sequence $\{n!\}_{n=0}^{\infty}$ and triangular array $\{\binom{k}{\ell} : 0 \leq \ell \leq k, k \geq 1\}$ satisfy NBL [12]. The sequences $\{n^2\}_{n=0}^{\infty}$ and $\{n^3\}_{n=0}^{\infty}$ appear to offer

special challenges, since the limiting digital probabilities evidently do not exist [3]. We hope to report on this later.

First-digit phenomena were mentioned in [13] without elaboration. In the language of [14], $\{a_n\}_{n=0}^{\infty}$ satisfies NBL if and only if the fractional parts of $\log_{10}(a_n)$ are uniformly distributed in $[0, 1]$, proved by Diaconis [12]. Our discussion can be extended to non-integer variables X , where we agree that $D_1(1/2) = 5 = D_1(1/20)$ (the first significant decimal digit). For example,

$$P\{D_1(X) = 1\} = \frac{1}{9} < \log_{10}(2)$$

if X is Uniform(0, 1) and

$$P\{D_1(X) = 1\} = \sum_{k=-\infty}^{\infty} (\exp(-10^k) - \exp(-2 \cdot 10^k)) = 0.32965\dots > \log_{10}(2)$$

if X is Exponential(1). Thus NBL does not apply in either case [2].

REFERENCES

- [1] A. Berger and T. P. Hill, Benford Online Bibliography, <http://www.benfordonline.net>.
- [2] A. Berger and T. P. Hill, Benford's law strikes back: no simple explanation in sight for mathematical gem, *Math. Intelligencer* 33 (2011) 85–91; MR2774458.
- [3] K. A. Ross, Benford's law, a growth industry, *Amer. Math. Monthly* 118 (2011) 571–583.
- [4] S. Slijepčević, A note on initial digits of recurrence sequences, *Fibonacci Quart.* 36 (1998) 305–308; MR1640361.
- [5] L. C. Washington, Benford's law for Fibonacci and Lucas numbers, *Fibonacci Quart.* 19 (1981) 175–177; MR0614056 (82f:10009).
- [6] J. L. Brown and R. L. Duncan, Modulo one uniform distribution of the sequence of logarithms of certain recursive sequences, *Fibonacci Quart.* 8 (1970) 482–486; MR0360444 (50 #12894).
- [7] L. Kuipers and J. S. Shiue, Remark on a paper by Duncan and Brown on the sequence of logarithms of certain recursive sequences, *Fibonacci Quart.* 11 (1973) 292–294; MR0332699 (48 #11025).
- [8] S. R. Finch, The Golden mean, *Mathematical Constants*, Cambridge Univ. Press, 2003, pp. 5–11.

- [9] S. Newcomb, Note on the frequency of use of the different digits in natural numbers, *Amer. J. Math.* 4 (1881) 39–40.
- [10] F. Benford, The law of anomalous numbers, *Proc. Amer. Philos. Soc.* 78 (1938) 551–572.
- [11] Z. Li, L. Cong and H. Wang, Discussion on Benford’s law and its application, arXiv:math/0408057.
- [12] P. Diaconis, The distribution of leading digits and uniform distribution mod 1, *Annals of Probab.* 5 (1977) 72–81; MR0422186 (54 #10178).
- [13] S. R. Finch, Continued fraction transformation. III, unpublished note (2007).
- [14] S. R. Finch, Pisot-Vijayaraghavan-Salem constants, *Mathematical Constants*, Cambridge Univ. Press, 2003, pp. 192–199.