ON NON-LINEAR RECURSIVE SEQUENCES AND BENFORD'S LAW

PAUL K. ROMANO AND HARRY MCLAUGHLIN

ABSTRACT. A large class of deterministic sequences are known to obey Benford's law. Recall that a sequence $\{x_n\}$ obeys Benford's law if and only if $\log_{10}|x_n| \pmod 1$ is uniformly distributed. It is proved herein that a particular class of sequences defined by multiplicative recursions obey Benford's law. This includes the three-term multiplicative Fibonacci sequence defined by $x_n = x_{n-1} \cdot x_{n-2}$.

1. INTRODUCTION

The frequency of occurrence of significant digits in numbers occurring in empirical data and deterministic sequences is a well-established topic. While there is no general agreement on the definition of Benford's law, it is most often associated with the distribution of the significant digits of elements in a set of data. If a set has the property that the mantissae of its elements are *nearly* uniformly distributed on the unit interval, then the set is said to obey Benford's law. As a direct result of this, the probability that a randomly selected number from a set of data obeying Benford's law has first significant digit n is $\log_{10}\frac{n+1}{n}$. For example, the probability that the first significant digit is 1 is 0.301, the probability that the first significant digit is 2 is 0.176, etc. Benford-like laws take the form $\mathbb{P}_k(n) > \mathbb{P}_k(n+1)$ where $\mathbb{P}_k(n)$ is the probability that the kth digit is n.

2. BENFORD'S LAW IN SEQUENCES

The distribution of significant digits given by Benford's law occurs not only in empirical data, but in many deterministic sequences as well. The sequences $n!$, a^n, and the Fibonacci numbers all obey Benford's law. A sequence of numbers $\{a_n\}$ that is bounded above and below is said to be uniformly distributed if the probability of finding a_n in any subinterval of its range is proportional to the length of the subinterval. This can be formalized by a limit. Let the counting function $C_{[a,b)}(N)$ be defined as the number of terms of $\{a_n\}$, $1 \le n \le N$, for which $a_n \in [a,b)$. Thus, $\{a_n\}$ is uniformly distributed if for each subinterval $[a,b)$

$$\lim_{N\to\infty} \frac{C_{[a,b)}(N)}{N} = b - a \tag{2.1}$$

where it is assumed that $[a,b)$ is a subinterval of a fixed interval of unity length. It is known that a sequence $\{a_n\}_{n=1}^{\infty}$ obeys Benford's law if and only if the sequence $\log_{10}|a_n|$ is uniformly distributed mod 1. As a result of this, the study of Benford's law is intimately related with the theory of uniform distribution modulo 1.

Research supported in part by the Office of Undergraduate Education at Rensselaer Polytechnic Institute through an Undergraduate Research Program.

Let $\{a_n\}$ be an infinite sequence of real numbers. Let $b_n = \log_{10}|a_n| \pmod 1$. The sequence $\{a_n\}$ is called a strong Benford sequence if for each subinterval $[a, b) \subset [0, 1)$,

$$\lim_{k\to\infty} \frac{1}{k} \sum_{n=1}^{k} \beta(n) = b - a, \tag{2.2}$$

where

$$\beta(n) = \begin{cases} 1 & \text{if } b_n \in [a, b) \\ 0 & \text{if } b_n \notin [a, b). \end{cases} \tag{2.3}$$

Fibonacci numbers are a good example of a strong Benford sequence. It is clear that if $\{a_n\}$ is a strong Benford sequence, then $\{b_n\}$ is uniformly distributed on the unit interval.

3. Multiplicative Recursions

Fibonacci numbers represent a special type of sequence in that they are defined recursively. Many authors have found that almost all sequences defined by linear recursions obey Benford's law. We are inclined to ask ourselves whether linear recursions are the only type of recursive sequence that obeys Benford's law. It is not difficult to confirm empirically that many multiplicative recursions obey Benford's law as well. In light of this, we present the following theorem.

Theorem 3.1. *The recursive sequence $x_n = \prod_{i=1}^{k} x_{n-i}$ for $n > k$ is a strong Benford sequence for all $k > 1$ and for almost all choices of $x_1, \ldots, x_k$ where $x_i > 0$.*

Proof. We seek to show that the sequence $\{\log_{10} x_n\}$, $n = 1, 2, \ldots$, is uniformly distributed mod 1. Note that each term of x_n can be written as a product of powers of x_1 through x_k.

$$x_n = (x_1)^{F_n^{(1)}} (x_2)^{F_n^{(2)}} \cdots (x_k)^{F_n^{(k)}} = \prod_{i=1}^{k} (x_i)^{F_n^{(i)}} \tag{3.1}$$

where each sequence $F_n^{(i)}$ is a Fibonacci k-step number which satisfies

$$F_n^{(i)} = F_{n-1}^{(i)} + F_{n-2}^{(i)} + \cdots + F_{n-k}^{(i)} = \sum_{j=1}^{k} F_{n-j}^{(i)}. \tag{3.2}$$

For each i, the initial values of $F_1^{(i)}, \ldots, F_k^{(i)}$ will be different. With equation (3.2), we can now write the logarithm of x_n as

$$\log_{10} x_n = \log_{10} \prod_{i=1}^{k} (x_i)^{F_n^{(i)}} = \sum_{i=1}^{k} F_n^{(i)} \log_{10} x_i. \tag{3.3}$$

It is well-known that the Fibonacci numbers can be written as a continuous function. This is also true of the Fibonacci k-step numbers. Flores has shown [1] that an exact formula for the Fibonacci k-step numbers can be given in terms of the k roots of the characteristic polynomial $P_k(x) = x^k - x^{k-1} - \cdots - 1 = 0$.

$$F_n^{(i)} = \alpha_1^{(i)} r_1^n + \alpha_2^{(i)} r_2^n + \cdots + \alpha_k^{(i)} r_k^n = \sum_{j=1}^{k} \alpha_j^{(i)} r_j^n \tag{3.4}$$

where $r_1, \ldots, r_k$ are the roots of $P_k(x)$. We will order the roots such that $|r_1| > |r_2| > \cdots > |r_k| > 0$. Hence, it follows that

$$\begin{aligned}\log_{10} x_n &= \sum_{i=1}^{k} \left(\sum_{j=1}^{k} \alpha_j^{(i)} r_j^n \right) \log_{10} x_i \\ &= \sum_{j=1}^{k} \left(\sum_{i=1}^{k} \alpha_j^{(i)} \log_{10} x_i \right) r_j^n \\ &= \sum_{j=1}^{k} \beta_j r_j^n.\end{aligned} \tag{3.5}$$

It can be demonstrated that the characteristic polynomial $P_k(x)$ has only one root for which $|r_j| > 1$ and all other roots lie inside the unit circle in the complex plane [4].

Theorem 1.2 in Kuipers and Niederreiter [3] states that if the sequence $\{x_n\}$, $n = 1, 2, \ldots$, is uniformly distributed mod 1, and if $\{y_n\}$ is a sequence with the property $\lim_{n\to\infty}(x_n - y_n) = c$, a real constant, then $\{y_n\}$ is uniformly distributed mod 1 as well. Therefore, it suffices to show that $\{\beta_1 r_1^n\}$ is uniformly distributed mod 1 in order to establish $\{\log_{10} x_n\}$ as being uniformly distributed mod 1 since

$$\lim_{n\to\infty} \beta_j r_j^n = 0 \quad \text{for } r_j \neq r_1. \tag{3.6}$$

Now, let $\lambda \in \mathbb{R}^+$ and $\theta > 1$. Koksma [2] showed that the sequence $\{\lambda\theta^n\}$ for $n = 1, 2, \ldots$ is uniformly distributed mod 1 unless θ belongs to a λ-dependent exceptional set S of measure zero. Thus, $\{\beta_1 r_1^n\}$ is uniformly distributed mod 1 for nearly all values of β_j. Since β_j is determined by the initial values of the sequence x_n, it immediately follows that $\log_{10} x_n$ is uniformly distributed mod 1 for nearly all values of $x_1, \ldots, x_k$. $\square$

With the theorem proved, it is worthwhile to make some remarks regarding the exceptional values of θ. The λ-dependent exceptional values of θ are known as Pisot-Vijayaraghavan numbers. It has been proven that for $\lambda = 1$, all quadratic irrationals whose conjugates lie within the unit circle belong to the exceptional set S. By the rational zeros theorem, one can show that all the roots of the characteristic polynomial $P_k(x)$ are irrational. Since all the conjugates of r_1 lie inside the unit circle, this implies that r_1 is a Pisot-Vijarayaghavan number. Therefore, if we have $\beta_1 = 1$, the sequence x_n will not obey Benford's law. However, it is easy to demonstrate empirically that nearly any choice of β_1 and hence nearly any choice of $x_1, \ldots, x_k$ will yield a sequence that obeys Benford's law.

It is instructive to look at a particular example of a multiplicative sequence that does obey Benford's law. Let us look at the sequence $x_n = x_{n-1} \cdot x_{n-2}$ with initial values $x_1 = 2$ and $x_2 = 3$. Notice that each term of x_n can be written as a product of powers of x_1 and x_2 as follows:

$$\begin{aligned}x_3 &= x_2 x_1 \\ x_4 &= x_3 x_2 = (x_2 x_1) x_2 = x_1 (x_2)^2 \\ x_5 &= x_4 x_3 = (x_1 (x_2)^2)(x_2 x_1) = (x_1)^2 (x_2)^3 \\ x_6 &= x_5 x_4 = ((x_1)^2 (x_2)^3)(x_1 (x_2)^2) = (x_1)^3 (x_2)^5 \\ x_7 &= x_6 x_5 = ((x_1)^3 (x_2)^5)((x_1)^2 (x_2)^3) = (x_1)^5 (x_2)^8\end{aligned} \tag{3.7}$$

or, in general

$$x_n = (x_1)^{F_{n-2}}(x_2)^{F_{n-1}}$$

where F_n is the nth Fibonacci number. Since $F_n = (\phi^n - \Phi^n)/\sqrt{5}$ where

$$\phi = \frac{1+\sqrt{5}}{2} \quad \text{and} \quad \Phi = \frac{1-\sqrt{5}}{2} \tag{3.8}$$

it follows that

$$\begin{aligned}
\log_{10} x_n &= \log_{10}(x_1)^{F_{n-2}}(x_2)^{F_{n-1}} \\
&= \left(\frac{\phi^{n-2} - \Phi^{n-2}}{\sqrt{5}}\right)\log_{10} x_1 + \left(\frac{\phi^{n-1} - \Phi^{n-1}}{\sqrt{5}}\right)\log_{10} x_2 \\
&= \left(\frac{\log_{10} x_1}{\phi^2\sqrt{5}} + \frac{\log_{10} x_2}{\phi\sqrt{5}}\right)\phi^n - \left(\frac{\log_{10} x_1}{\Phi^2\sqrt{5}} + \frac{\log_{10} x_2}{\Phi\sqrt{5}}\right)\Phi^n \\
&= \beta_1\phi^n - \beta_2\Phi^n.
\end{aligned} \tag{3.9}$$

For $x_1 = 2$ and $x_2 = 3$, we have that $\beta_1 \approxeq 0.1833$, and thus we expect x_n to obey Benford's law. Figure 1 shows the frequency of first digits for the first million terms of x_n.

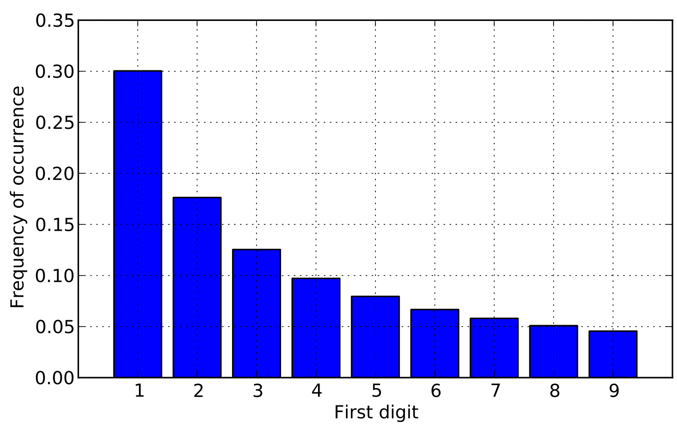

FIGURE 1. Frequency of first digits for the first million terms of $x_n = x_{n-1} \cdot x_{n-2}$ with $x_1 = 2$ and $x_2 = 3$.

We can also determine a pair of exceptional values for x_1 and x_2 such that $\beta_1 = 1$. One such example would be the choices $x_1 = 1$ and $x_2 = 10^{\phi\sqrt{5}} \approx 4149.87$. The sequence x_n with this pair of initial values will not obey Benford's law. One can observe that the sequence $\log_{10}|x_n|$ (mod 1) has zero and unity as its only limit points.

Lastly, we note that there is nothing intrinsic to this derivation that relies on using base 10. Without loss of generality, we can conclude that the sequence $x_n = \prod_{i=1}^{k} x_{n-i}$ for $n > k$ is a strong b-Benford sequence for all $k > 1$ and for nearly all choices of $x_1, \ldots, x_k$ where b is the base.

REFERENCES

[1] I. Flores, *Direct calculation of k-generalized Fibonacci numbers*, The Fibonacci Quarterly, **5** (1967), 259–266.

[2] J. F. Koksma, *Ein mengentheoretischer Satz über die Gleichverteilung modulo Eins*, Compositio Mathematica, **2** (1935), 250–258.

[3] L. Kuipers and H. Niederreiter, *Uniform Distribution of Sequences*, Wiley-Interscience Publications, New York, 1974.
[4] E. P. Miles Jr., *Generalized Fibonacci numbers and associated matrices*, The American Mathematical Monthly, **67** (1960), 745–752.

MSC2010: 11K31, 11B37

Department of Nuclear Science and Engineering, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Building 24-607, Cambridge, MA 02139
E-mail address: `paul.romano@alum.mit.edu`

Department of Mathematical Sciences, Rensselaer Polytechnic Institute, 110 Eighth Street, Troy, NY 12180
E-mail address: `mclauh@rpi.edu`