A Weak First Digit Law for
a Class of Sequences

M. A. Nyblom

School of Mathematics and Geospatial Sciences
RMIT University, Melbourne, Victoria 3001, Australia

Abstract

A non-ergodic approach is employed to establish Benford’s law for the leading digit \( d = 1 \) for the sequence of powers of two. For the sequence of powers \( \{\alpha^n\}, 1 < \alpha \leq 10/9 \), this method is extended to obtain a weak first digit law by establishing Benford like lower and upper bounds to the asymptotic relative frequency of terms with a given leading digit.

Mathematics Subject Classification: 11A99, 11K36

Keywords: Frequency distribution, Benford’s law

1 Introduction

When examining numerical data which is derived from large naturally occurring data sets such as population, lengths of rivers or stock prices, often (but not always), a curious phenomena known as Bendford’s law is observed. This phenomenological law concerns the frequency distribution of the leading digits of numerical data, which typically spans many orders of magnitude. Such a set of numerical data will satisfy Benford’s law if the proportion of data values having a leading digit of \( d \in \{1, 2, \ldots, 9\} \) is approximately \( \log_{10}(1 + \frac{1}{d}) \). Thus with numerical data sets which obey this law, the number 1 would appear
as a leading digit with a frequency of approximately 30%, while larger digits such as 9 would appear as a leading digit with a frequency of approximately 4.5%. Benford’s law was first empirically observed in 1881 by the American astronomer Simon Newcomb [2], but remained largely forgotten until 1938, when Frank Benford [1], a physicist, re-tested the phenomenon against many sources of data as well as on some infinite integer sequences. He found in both case, that the frequency distribution of the leading digit d exhibited a close adherence to the logarithmic rule.

For a sequence of positive reals \( \{a_n\} \), if one defines

\[
N_d(n) = \#\{0 \leq k \leq n : \lfloor a_k \rfloor \text{ has leading digit } d\},
\]

then \( \{a_n\} \) will satisfy Benford’s law if \( \lim_{n \to \infty} \frac{N_d(n)}{n+1} = \log_{10}(1 + \frac{1}{d}) \). To establish that an integer sequence such as \( \{2^n\} \) satisfies Benford’s law, one can appeal to an Ergodic property of the irrational numbers proved by Weyl [3]. Namely, if \( x \) is a positive irrational number, \( n \) a positive integer and \( \lfloor nx \rfloor \) the integer part of \( nx \), then the sequence of fractional part of \( nx \), given by \( c(n) = nx - \lfloor nx \rfloor \), is homogeneously distributed over the interval \([0, 1]\). That is, given any arbitrary reals \( 0 \leq a < b \leq 1 \) then

\[
\lim_{n \to \infty} \frac{\#\{0 \leq k \leq n : c(k) \in (a, b)\}}{n + 1} = b - a.
\]

Now if \( 2^n \) has a leading digit of \( d \), then \( d \cdot 10^k < 2^n < (d + 1) \cdot 10^k \) for some integer \( k \). Upon taking logarithms one finds \( \log_{10}(d) + k < n \log_{10}(2) < \log_{10}(d + 1) + k \), and as \( \log_{10}(d + 1) \in [0, 1] \), one deduces \( k = \lfloor n \log_{10}(2) \rfloor \), and \( \log_{10}(d) < c(n) < \log_{10}(d + 1) \), where \( c(n) = n \log_{10}(2) - \lfloor n \log_{10}(2) \rfloor \). Thus when \( a_k = 2^k \) with \( a = \log_{10}(d) \) and \( b = \log_{10}(d + 1) \), we have \( N_d(n) = \#\{0 \leq k \leq n : c(k) \in (a, b)\} \), from which the required frequency distribution follows via (1), as \( \log_{10}(2) \) is irrational. In this article an entirely elementary proof will first be presented, which establishes the logarithmic rule for the frequency distribution of terms in the sequence \( \{2^n\} \) having leading digit \( d = 1 \). This is possible, as one can show for \( a_n = 2^n \) that \( N_1(n) = \lfloor n \log_{10}(2) \rfloor + 1 \), via an argument based on counting the number of integers in an interval. By applying a variation to this argument, one can then obtain a weak first digit law for the sequence \( a_n = \alpha^n \), for a fixed \( 1 < \alpha \leq \frac{10}{9} \), where in particular it is shown that for all \( n \)

\[
L_n \leq \frac{N_d(n)}{n+1} \leq U_n,
\]

with \( L_n \sim \log_{10}(1 + \frac{1}{d}) - \log_{10}(\alpha) \) and \( U_n \sim \log_{10}(1 + \frac{1}{d}) + \log_{10}(\alpha) \) as \( n \to \infty \). Although this result is only suggestive of the logarithmic rule of Benford’s law, the advantage of the approach taken here rests on the non-ergodic nature of the argument employed.
2 First digit law for powers of two

We begin with a technical lemma concerning the counting of integers in an interval of the form \([a, b]\). In what follows, define for a real number \(x\), the floor and ceiling of \(x\) as \([x] = \max\{n \in \mathbb{Z} : x \geq n\}\) and \(\lceil x \rceil = \min\{n \in \mathbb{Z} : n \geq x\}\) respectively. Furthermore, we shall make use of the fact that \([x + n] = [x] + n\) for all \(n \in \mathbb{Z}\) and \(x \in \mathbb{R}\).

**Lemma 2.1** Suppose \(0 \leq a < b\) with \(b - a = L\), then the number of integers contained in the interval \([a, b]\) is either \([L]\) or \([L] + 1\).

**Proof:** Let \(I(a, b)\) denote the number of integers contained in the interval \([a, b]\). We note that for two integers \(0 \leq m \leq n\), the number of integers in the interval \([m, n]\) is equal to \(n - m + 1\).

**Case 1:** \(b \notin \mathbb{N}\).

In this instance \(I(a, b) = [b] - [a] + 1\) since \([a], [b] \subseteq [a, b]\). As \([b] \leq b\) and \([a] \geq a\) one finds \(I(a, b) \leq b - a + 1\), but as \([b - a + 1]\) is the largest integer less than or equal to \(b - a + 1\), we deduce that \(I(a, b) \leq [b - a + 1] = [b - a] + 1 = \lceil L \rceil + 1\). Now as \([a] < a + 1\), then \(b - [a] > b - a - 1 \geq [b - a - 1]\) and as \([b - [a]] = [b] - [a]\) is the largest integer less than or equal to \(b - [a]\), we further deduce \([b] - [a] \geq [b - a - 1] = [b - a] - 1\), that is \(I(a, b) \geq [L]\). Hence \(I(a, b)\) is either equal to \([L]\) or \([L] + 1\).

**Case 2:** \(b \in \mathbb{N}\).

If \(a \in \mathbb{N}\), then \(I(a, b) = (b - 1) - a + 1 = b - a = [L]\), since \([a, b - 1] \subseteq [a, b]\). Alternatively if \(a \notin \mathbb{N}\), then \(I(a, b) = (b - 1) - ([a] + 1) + 1 = b - [a] - 1\), since \(([a], b - 1) \subseteq [a, b]\) and \([a] = [a] + 1\). But as \([a] < a < [a] + 1\) we find \(b - [a] > b - a > b - [a] - 1\), that is \(I(a, b) + 1 > b - a > I(a, b)\) and so \(I(a, b) = [b - a] = [L]\).

**Theorem 2.1** The frequency distribution of the terms in the sequence \(\{2^n\}\) having leading digit \(d = 1\), is \(\log_{10}(2)\).

**Proof:** Recall \(N_1(n) = \#\{0 \leq m \leq n : 2^m\) has leading digit \(1\}\). If \(2^m\) has \(k\) digits and leading digit \(1\), then \(10^{k-1} \leq 2^m < 2 \cdot 10^{k-1}\). Upon taking logarithm to base ten yields \(k - 1 \leq m \log_{10}(2) < \log_{10}(2) + k - 1\).

Now consider the continuous linear function \(f(x) = \log_{10}(2)x\), and suppose the interval of values on which \(f(x)\) satisfies (2) is \([a, b]\). As \(f(b) - f(a) = \log_{10}(2)\), we deduce that the interval \([a, b]\) must have unit length. Thus from Lemma 2.1, the interval \([a, b]\) can contain either one or two positive integers. To
show there is exactly one such integer in \([a, b)\), suppose there are two integers \(m_1 < m_2\) such that both \(2^{m_1}\) and \(2^{m_2}\) have \(k\) digits with leading digit 1. As \(m_2 \geq m_1 + 1\), observe \(2^{m_2} - 2^{m_1} \geq 2^{m_1}(2^1 - 1) = 2^{m_1}\) and so \(2^{m_2} - 2^{m_1}\) will have at least \(k\) digits, but as both \(2^{m_2}\) and \(2^{m_1}\) have \(k\) digits with a leading digit 1, the difference \(2^{m_2} - 2^{m_1}\) can have at most \(k - 1\) digits, a contradiction. Hence there is exactly one power of two having \(k\) digits with leading digit 1. Consequently \(N_1(n)\) must be equal to the positive integer \(k\) such that \(2^n \in [10^{k-1}, 2 \cdot 10^{k-1}) \cup [2 \cdot 10^{k-1}, 10^k)\) or equivalently \(k - 1 = \lfloor n \log_{10}(2) \rfloor\) and so \(N_1(n) = \lfloor n \log_{10}(2) \rfloor + 1\). Finally as \(x - 1 < [x] \leq x\) one finds that

\[
\frac{n \log_{10}(2)}{n + 1} < \frac{N_1(n)}{n + 1} < \frac{n \log_{10}(2) + 1}{n + 1},
\]

from which the desired limiting value of \(\log_{10}(2)\) follows upon taking \(n \to \infty\).

3 A weak first digit law for \(\alpha^n\)

In Section 2 we were able to determine the correct frequency distribution for the powers of two having a leading digit of \(d = 1\), since an explicit formula for \(N_1(n)\) could be derived. By varying the argument used to prove Theorem 2.1, we can obtain upper and lower estimates of \(N_d(n)\), for the sequence \(\{\alpha^n\}\), where \(1 < \alpha \leq \frac{10}{9}\). These estimates can then be used to obtain upper and lower bounds on the frequency distribution for the terms of the sequence \(\{\alpha^n\}\) having leading digit \(d\), which are suggestive of the logarithmic rule predicted via Benford’s law.

**Theorem 3.1** Given the sequence \(\{\alpha^n\}\), where \(1 < \alpha \leq \frac{10}{9}\), then for all \(n\)

\[
L_n \leq \frac{N_d(n)}{n + 1} \leq U_n ,
\]

with \(L_n \sim \log_{10}(1 + \frac{1}{d}) - \log_{10}(\alpha)\) and \(U_n \sim \log_{10}(1 + \frac{1}{d}) + \log_{10}(\alpha)\) as \(n \to \infty\).

**Proof:** Recall \(N_d(n) = \#\{0 \leq m \leq n : |\alpha^m|\) has leading digit \(d\}\}, moreover from assumption observe \(\log_{10}(\alpha) \leq \log_{10}(\frac{10}{9}) \leq \log_{10}(\frac{d+1}{d})\). If \(|\alpha^m|\) has \(k\) digits and leading digit \(d\), then \(d \cdot 10^{k-1} \leq \alpha^m < (d + 1) \cdot 10^{k-1}\). Upon taking logarithm to base ten yields

\[
\log_{10}(d) + (k - 1) \leq m \log_{10}(\alpha) < \log_{10}(d + 1) + (k - 1) .
\] (3)

Now consider the continuous linear function \(f(x) = \log_{10}(\alpha)x\), and suppose the interval of values on which \(f(x)\) satisfies (3) is \([a, b)\). As \(f(b) - f(a) = \)
log_{10}(\frac{d+1}{d})$, we deduce that the interval $[a, b)$ must have length $L = \log_{10}(\frac{d+1}{d}) / \log_{10}(\alpha) \geq 1$. Thus from Lemma 2.1, the interval $[a, b)$ contains either $[L]$ or $[L] + 1$ positive integers. Consequently, the number of terms of the sequence $\{\alpha^m\}$ for which $|\alpha^n|$ has $k$ digits and leading digit $d$, is either $[L]$ or $[L] + 1$, moreover these upper and lower bounds are independent of $k$. If $|\alpha^n|$ contains $s$ digits, then $10^{s-1} \leq \alpha^n < 10^s$, from which one deduces that $s - 1 = \lfloor n \log_{10}(\alpha) \rfloor$, and so $\{\alpha^0, \alpha^1, \ldots, \alpha^n\} \subseteq \bigcup_{k=1}^{\lfloor n \log_{10}(\alpha) \rfloor + 1} [10^{k-1}, 10^k)$. Now as $\alpha^n \in [10^{\lfloor n \log_{10}(\alpha) \rfloor}, d \cdot 10^{\lfloor n \log_{10}(\alpha) \rfloor}] \cup [d \cdot 10^{\lfloor n \log_{10}(\alpha) \rfloor}, (d+1) \cdot 10^{\lfloor n \log_{10}(\alpha) \rfloor}] \cup [(d+1) \cdot 10^{\lfloor n \log_{10}(\alpha) \rfloor}, 10^{\lfloor n \log_{10}(\alpha) \rfloor} + 1)$, the leading digit of $|\alpha^n|$ may or may not be $d$, and so the number of sequence elements in $[10^{\lfloor n \log_{10}(\alpha) \rfloor}, 10^{\lfloor n \log_{10}(\alpha) \rfloor} + 1)$ having a leading digit $d$, can only be bounded between $0$ and $[L] + 1$. While from above, the number of sequence elements having a leading digit of $d$ in $[10^{k-1}, 10^k)$, for $k = 1, \ldots, \lfloor n \log_{10}(\alpha) \rfloor$, must be bounded between $[L]$ and $[L] + 1$. Thus from definition, we have for all $n \geq 0$

$$\frac{\lfloor L\rfloor \lfloor n \log_{10}(\alpha) \rfloor}{n+1} \leq \frac{N_d(n)}{n+1} \leq \frac{\lfloor L\rfloor \lfloor n \log_{10}(\alpha) \rfloor + 1}{n+1}. \quad (4)$$

Finally, using again the inequality $x - 1 < \lfloor x \rfloor \leq x$, one finds that the upper and lower bounds of (4) are respectively bounded above and below by

$$U_n = \left( \frac{\log_{10}(\frac{d+1}{d})}{\log_{10}(\alpha)} + 1 \right) \frac{n \log_{10}(\alpha) + 1}{n+1} \quad \text{and} \quad L_n = \left( \frac{\log_{10}(\frac{d+1}{d})}{\log_{10}(\alpha)} - 1 \right) \frac{n \log_{10}(\alpha) - 1}{n+1},$$

with $L_n \sim \log_{10}(1 + \frac{1}{d}) - \log_{10}(\alpha)$ and $U_n \sim \log_{10}(1 + \frac{1}{d}) + \log_{10}(\alpha)$ as $n \to \infty$.

To conclude, we note the result of Theorem 3.1 can be strengthened for the terms of the sequence $\{(\frac{10}{9})^n\}$, having a leading digit of $9$ as follows. If we set $\alpha = \frac{10}{9}$ and $d = 9$, then the length of the interval $[a, b)$ on which the linear function $f(x)$ satisfies (3) is $L = 1$, and so there can only be one or two terms of the sequence $\{(\frac{10}{9})^n\}$ having $k$ digits with a leading digit of $9$. To show there is exactly one, assume there are two integers $m_1 < m_2$ such that both $\lfloor (\frac{10}{9})^{m_1} \rfloor$ and $\lfloor (\frac{10}{9})^{m_2} \rfloor$ have $k$ digits with a leading digit $9$. As $m_2 \geq m_1 + 1$, observe $(\frac{10}{9})^{m_2} - (\frac{10}{9})^{m_1} \geq (\frac{10}{9})^{m_1+1} - (\frac{10}{9})^{m_1} = (\frac{10}{9})^{m_1}(\frac{1}{9})$, but $\lfloor (\frac{10}{9})^{m_1}(\frac{1}{9}) \rfloor$ has $k$ digits with a leading digit of $1$. However as both $(\frac{10}{9})^{m_2}$ and $(\frac{10}{9})^{m_1}$ have $k$ digits with a leading digit $9$, the difference $\lfloor (\frac{10}{9})^{m_2} - (\frac{10}{9})^{m_1} \rfloor$ can have at most $k - 1$ digits, a contradiction. Now $\{(\frac{10}{9})^0, (\frac{10}{9})^1, \ldots, (\frac{10}{9})^n\} \subseteq \bigcup_{k=1}^{\lfloor n \log_{10}(\frac{10}{9}) \rfloor + 1} [10^{k-1}, 10^k)$, and as the number of sequence elements having a leading digit of $9$ in $[10^{k-1}, 10^k)$, for $k = 1, \ldots, \lfloor n \log_{10}(\frac{10}{9}) \rfloor$, is precisely one, we can conclude as $(\frac{10}{9})^n$ may or may not have a leading digit of $9$, that

$$\frac{n \log_{10}(\frac{10}{9}) - 1}{n+1} \leq \frac{\lfloor n \log_{10}(\frac{10}{9}) \rfloor}{n+1} \leq \frac{N_9(n)}{n+1} \leq \frac{\lfloor n \log_{10}(\frac{10}{9}) \rfloor + 1}{n+1} \leq \frac{n \log_{10}(\frac{10}{9}) + 1}{n+1}.$$
from which the desired limiting value of $\log_{10}\left(\frac{10}{9}\right)$ follows upon taking $n \to \infty$.

**References**


Received: June 7, 2016; Published: July 25, 2016