INVESTIGATION ON A CLASS OF INFINITE NUMBER SEQUENCES WHICH SATISFY BENFORD'S LAW

DESHU CHEN

(Under the instruction of Lei Hou)
High School Affiliated to Shanghai Jiao Tong University,
42 West Yin Gao Road, Shanghai 200439
chende97@mail.qq.com

ABSTRACT. For the sequence of integers 1,2,4,..., 2^{n-1}, ..., by computation, one finds that the frequency (over front n terms) that a random decimal begins with the digit 1 is nearly $\lg 3/2$, while the frequency that a random decimal begins with the digit 9 approaches to $\lg 10/9$. In fact, such kind of phenomenon about infinite series is usually referred to as the so-called "Benford's law". As early as 1881, Simon Newcomb first observed this phenomenon. The first one who studied this phenomenon might be Frank Benford, a physician, who published his paper in 1938. In 1971 and 1978, Wlodarski and Brady respectively published their papers which stated that Fibonacci and Lucas numbers satisfy Benford's law from statistics. In 1981, Lawrence Washington strictly proved the above mentioned observation. The aim of this paper is to improve Washington's theorem. As a direct application, we find out more new number sequences that satisfy Benford's law.

1. Introduction

If one considers the number sequence 1, 2, ..., 2^{n-1}, ... and count the frequency that a random decimal begins with the digit d ($1 \le d \le 9$), say $f(d,n)$, the following table can be obtained:

Table 1

$F(d)$	$d=1$	$d=2$	$d=3$	$d=4$	$d=5$	$d=6$	$d=7$	$d=8$	$d=9$
$n=10^2$	0.300	0.170	0.130	0.100	0.070	0.070	0.060	0.050	0.050
$n=10^3$	0.301	0.176	0.125	0.097	0.079	0.069	0.056	0.052	0.045
$n=10^4$	0.301	0.176	0.125	0.097	0.079	0.067	0.058	0.051	0.046

One knows from Table 1 that the frequency $f(n,d)$ displays its own regularities as d varies from 1 to 9. In fact, it is proven that $f(n,1) \mapsto P(1) = \lg\ 2 \approx 0.301$ and that $f(9,n) \mapsto P(9) = \lg \frac{10}{9} \approx 0.046$ as $n \mapsto +\infty$. Thus, quite contrary to the naive expectation, $f(n,1) \neq f(n,9)$.

As early as 1881, Simon Newcomb [3] observed the above phenomenon. More systematical study might initiate with Frank Benford, a physician, who published his paper [1] on this topic in 1938. Such kind of distribution rules of infinite number series is usually referred to as "Benford's Law" to recognize the contribution of Benford.

Definition 1.1. *For an infinite number sequence $\{a_n\}$, as $n \mapsto \infty$, if the frequency $f(n,d)$ $(1 \le d \le 9)$ that the random decimal begins with digit d has the limit*

$$P(d) = \lg \frac{d+1}{d} = \lg(1 + \frac{1}{d}),$$

we say that $\{a_n\}$ satisfies Benford's law.

Write $(1+\sqrt{5})^n = L_n + F_n\sqrt{5}$ where $\{L_n\}$ and $\{F_n\}$ are usually called Lucas sequence and Fibonacci sequence, respectively.

In this paper, we are interested in the following aspect of Benford's law:

- In 1971 and 1978, Wlodarski [6] and Brady [2] respectively published their papers and pointed out that both Fibonacci sequence and Lucas sequence satisfy Benford's law from statistics.
- In 1981, Washington [5] applied Weyl's theorem to prove that Fibonacci numbers and Lucus numbers satisfy Benford's law.

Hence it is interesting to find out more new number sequences which can be proven to satisfy Benford's law. The aim of this note is to improve the method of Washington [5] and to generalize his result. As a direct application of our theorem, we are able to find out some new examples (see Example 2.5, 2.6 and Corollary 2.7) which satisfy Benford's law. In the last section, we have written a short program to compute some other number sequences, which do not satisfy the conditions of our theorem, and to investigate statistically if they satisfy Benford's law.

2. The main result and the proof

The main result of this paper is the following theorem:

Theorem 2.1. *Let $\{x_n\}$ be a sequence of positive real numbers. Assume that*

(1). *there are complex numbers c_i, a_i $(i = 1, \dots, k)$ such that*

$$x_n = \sum_{i=1}^{k} c_i a_i^n$$

holds for any $n > 0$;

(2). *there exists an integer $s > 0$ such that $c_s \neq 0$, $\lg |a_s|$ is not a rational number and that $|a_s| > |a_j|$, $\forall\ j \neq s$.*

Then the number sequence $\{x_n\}$ satisfies Benford's law.

Proof. First of all we may write

$$x_n = c_s a_s^n \big(1 + \sum_{j\neq s} \frac{c_j}{c_s} \cdot \frac{a_j^n}{a_s^n}\big) = c_s a_s^n (1 + r_n),$$

where $r_n = \sum_{j\neq s} \frac{c_j}{c_s} \cdot \frac{a_j^n}{a_s^n}$ and $|r_n| \mapsto 0$ as $n \mapsto \infty$.

Lemma 2.2. *Both a_s and c_s must be real numbers. Hence r_n is a real number for any $n > 0$.*

Proof. If a_s is not real, we may always find some very large integer n_0 so that $c_s a_s^{n_0+1}$ has non-zero argument θ which does not depend on n. Since $r_n \mapsto 0$, x_{n_0+1} is in a very small neighborhood of $c_s a_s^{n_0+1}$. Thus x_{n_0+1} can not be a real number, a contradiction. So a_s must be real. Similarly c_s must be real too. □

Step 1. The real sequence $y_n = \alpha^n(1 + r_n)$.

We first study the special number sequence $\{y_n\}$ where $\alpha > 0$, $\lg \alpha$ is irrational and $r_n \mapsto 0$ as $n \mapsto \infty$. Denote by $\langle \lg y_n \rangle$ the fractional part of $\lg y_n$. We are going to prove that the fractional sequence is equidistributed mod 1, namely, the probability that $\langle \lg y_n \rangle$ is in the given interval $[u, v) \subset [0, 1]$ is equal to $v - u$.

We recall Weyl's criterion here.

Lemma 2.3. *([4, P.112]) A sequence of real numbers $\{\xi_n\}$ in $[0,1)$ is equidistributed if and only if, for all integers $l \neq 0$, one has*

$$\lim_{N\mapsto\infty} \frac{1}{N}\sum_{n=1}^{N} e^{2\pi i l \xi_n} = 0. \tag{1}$$

As a slightly special case, since $\lg\alpha$ is irrational, we have

$$\lim_{n\mapsto\infty} \frac{1}{n}\sum_{j=1}^{n} e^{2\pi i l \lg \alpha^j} = \lim_{n\mapsto\infty} \frac{1}{n}\cdot \frac{e^{2\pi i l \lg \alpha} - e^{2\pi i l \lg \alpha^{n+1}}}{1-e^{2\pi i l \lg\alpha}} = 0. \tag{2}$$

Hence the sequence $\{\langle n\lg\alpha\rangle\}$ is equidistributed mod 1. By applying Relation (2), we have

$$\begin{aligned}
& \lim_{n\mapsto\infty} \frac{1}{n}\sum_{j=1}^{n} e^{2\pi i l \lg y_j} \\
= & \lim_{n\mapsto\infty} \frac{1}{n}\sum_{j=1}^{n} e^{2\pi i l (j\lg\alpha + \lg(1+r_j))} \\
= & \lim_{n\mapsto\infty} \frac{1}{n}\sum_{j=1}^{n} e^{2\pi i l \lg(1+r_j)}\cdot e^{2\pi i l j\lg\alpha} - \lim_{n\mapsto\infty} \frac{1}{n}\sum_{j=1}^{n} e^{2\pi i l \lg\alpha^j} \\
= & \lim_{n\mapsto\infty} \frac{1}{n}\sum_{j=1}^{n} u_j v_j = 0, \qquad (3)
\end{aligned}$$

where $u_j = e^{2\pi i l \lg(1+r_j)} - 1 \mapsto 0$ as $j\mapsto\infty$, $v_j = e^{2\pi i l j \lg\alpha}$ satisfies $\{v_n\}$ being bounded. The last equation is due to Lemma 2.4 which is a basic property of limits.

Lemma 2.4. *For real number sequences $\{u_n\}$ and $\{v_n\}$, if $\lim_{n\mapsto\infty} u_n = 0$ and $\{v_n\}$ is bounded, then $\lim_{n\mapsto\infty}\frac{1}{n}\sum_{j=1}^{n} u_j v_j = 0$.*

Step 2. The statement for $\{x_n\}$.

Set $\alpha = a_s$ and $c = c_s$. Then $x_n = cy_n$ for any $n > 0$. Since $\alpha > 0$ and $\lg\alpha$ is irrational, the number sequence $\{\langle\lg x_n\rangle\} = \{\langle\lg c + \lg y_n\rangle\}$ is equidistributed by virtue of Step 1. Assume that m is the integral part of $\lg x_n$. Then we have seen that the probability that $\{\lg x_n - m\}$ is in the interval $[\lg d, \lg(d+1))$ is equal to $\lg(d+1)-\lg d$, which means that the probability that x_n begins with the digit d is equal to $\lg(d+1)-\lg d$ for $1\le d\le 9$. Thus $P(d) = \lg(d+1) - \lg d$ and $\{x_n\}$ satisfies Benford's law. □

As direct applications of Theorem 2.1, we provide some new examples.

Example 2.5. Let $(a+b\sqrt{2})^n = A_n + B_n\sqrt{2}$ where a and b are positive rational numbers. Clearly we have

$$\begin{aligned}
A_n &= \frac{1}{2}\big((a+b\sqrt{2})^n + (a-b\sqrt{2})^n\big); \\
B_n &= \frac{\sqrt{2}}{4}\big((a+b\sqrt{2})^n + (a-b\sqrt{2})^n\big).
\end{aligned}$$

By Theorem 2.1, one knows that both $\{A_n\}$ and $\{B_n\}$ satisfy Benford's law.

Example 2.6. Let $(a+b\sqrt[3]{2})^n = C_n + D_n\sqrt[3]{2} + E_n\sqrt[3]{4}$ where a and b are positive rational numbers. Let $\omega = -\frac{1}{2} + \frac{\sqrt{3}}{2}i$. One has the following relations:

$$\begin{cases} C_n & + & \sqrt[3]{2}D_n & + & \sqrt[3]{4}E_n & = & (a+b\sqrt[3]{2})^n \\ C_n & + & \sqrt[3]{2}\omega D_n & + & \sqrt[3]{4}\omega^2 E_n & = & (a+b\sqrt[3]{2}\omega)^n \\ C_n & + & \sqrt[3]{2}\omega^2 D_n & + & \sqrt[3]{4}\omega^4 E_n & = & (a+b\sqrt[3]{2}\omega^2)^n. \end{cases}$$

The above equations in terms of C_n, D_n and E_n has the coefficient determinant a "Vander Monde Determinant" which is equal to

$$\Delta = \begin{vmatrix} 1 & \sqrt[3]{2} & \sqrt[3]{4} \\ 1 & \sqrt[3]{2}\omega & \sqrt[3]{4}\omega^2 \\ 1 & \sqrt[3]{2}\omega^2 & \sqrt[3]{4}\omega^4 \end{vmatrix} = 2(\omega-1)(\omega^2-1)(\omega^2-\omega).$$

Thus C_n, D_n and E_n all have their formulae solution. For example, one has

$$C_n = \frac{1}{3}\big((a+b\sqrt[3]{2})^n + (a+b\sqrt[3]{2}\omega)^n + (a+b\sqrt[3]{2}\omega^2)^n\big).$$

Notice that $(\omega-1)(\omega^2-1) = 3$, $\alpha = a + b\sqrt[3]{2}$ and that $\lg\alpha$ is irrational. Theorem 2.1 implies that $\{C_n\}$ satisfies Benford's law. So do $\{D_n\}$, $\{E_n\}$ for similar reasons.

Inspired by above examples, we have the following corollary.

Corollary 2.7. *Let t and m be positive integers. Assume that the minimal degree of rational polynomials $f(x)$ with $f(\sqrt[m]{t}) = 0$ is m. Let a and b be positive rational numbers. Write*

$$(a+b\sqrt[m]{t})^n = A_n^{(0)} + A_n^{(1)}\sqrt[m]{t} + \cdots + A_n^{(m-1)}\sqrt[m]{t^{m-1}}.$$

Then each of $\{A_n^{(0)}\}$, $\{A_n^{(1)}\}$, ..., $\{A_n^{(m-1)}\}$ satisfies Benford's law.

Proof. Let $\omega = e^{\frac{2\pi i}{m}}$. By binomial formula, we have the following linear equations in terms of $A_n^{(j)}$:

$$\begin{cases} A_n^{(0)} & + & A_n^{(1)}\sqrt[m]{t} & + & \cdots & + & A_n^{(m-1)}\sqrt[m]{t^{m-1}} & = & (a+b\sqrt[m]{t})^n \\ A_n^{(0)} & + & A_n^{(1)}\sqrt[m]{t}\omega & + & \cdots & + & A_n^{(m-1)}\sqrt[m]{t^{m-1}}\omega^{m-1} & = & (a+b\sqrt[m]{t}\omega)^n \\ \cdots & & & & & & & & \\ A_n^{(0)} & + & A_n^{(1)}\sqrt[m]{t}\omega^{m-1} & + & \cdots & + & A_n^{(m-1)}\sqrt[m]{t^{m-1}}\omega^{(m-1)^2} & = & (a+b\sqrt[m]{t}\omega^{m-1})^n \end{cases}$$

For each $j = 0, \ldots, m-1$, multiply the first equation by ω^{m-1}, the second by ω^{m-1-j}, ..., the k-th equation by $\omega^{m-1-(k-1)j}$, the last equation by $\omega^{m-1-(m-1)j}$ and add up all together to obtain:

$$A_n^{(j)} = \frac{1}{m\sqrt[m]{t^j}}\big((a+b\sqrt[m]{t})^n + \omega^{-j}(a+b\sqrt[m]{t}\omega)^n + \cdots + \omega^{-(m-1)j}(a+b\sqrt[m]{t}\omega^{m-1})^n\big).$$

Note that $|a+b\sqrt[m]{t}| > |a+b\sqrt[m]{t}\omega^j|$, $\forall j = 1, \cdots, m-1$. Note also that $\lg(a+b\sqrt[m]{t})$ is irrational. Otherwise, $a + b\sqrt[m]{t} = 10^{u/v}$ for positive integers u, v, which implies $(a + b\sqrt[m]{t})^v = 10^u$ and then $\sqrt[m]{t}$ is a root of a rational polynomial $g(x)$ of degree $< m$, a contradiction. Thus Theorem 2.1 implies that $\{A_n^{(j)}\}$ satisfies Benford's law for any $j \geq 0$. □

Remark 2.8. It sounds very interesting to consider possible generalization to Corollary 2.7 by replacing $\sqrt[m]{t}$ with any real root of any rational polynomial. But that certainly goes beyond the scope of knowledge of the author.

3. Investigation of some other number sequences

In this section the author wrote at the beginning the following simple program to count the statistics of several common number sequences to see if they are subject to Benford's law. Here is our program in Visual Basic:

```
Dim n As Integer
n = Val(InputBox("n="))
Dim x(2000) As Double
Dim f(2000) As Double
Dim t(2000) As Double
For j = 1 To n
        x(j) = Log(2 ^ j) / Log(10)
Next j
For d = 1 To 9
        t(d) = 0
        For j = 1 To n
            m = Int(x(j))
            u = x(j) - m
            If u >= (Log(d) / Log(10)) And u < (Log(d + 1) / Log(10)) Then
                t(d) = t(d) + 1
            End If
            f(d) = t(d) / n
        Next j
        Print f(d)
Next d
```

In the process of our calculation, if $n > 1500$, the author's PC stops to work. Our guider suggested us to try to use Matlab. So we spent a lot of time to learn the way in applying Matlab, a very effective software. Here we provide our small program in Matlab, which helped us to calculate up to the level $n = 50000$ successfully.

```
   %Benford Law
function F=F(d)
F=[];
N=[];
% input n
n=1000;
for j=1:n
x(j)=log10(j^ 100*sin(j));
end
for d=1:9
N(d)=0;
% fractional part
for j=1:n
m=floor(x(j));
u=x(j)-m;
%Statistics
if u>= log10(d) & u< log10(d+1)
N(d)=N(d)+1;
```

```
end
F(d)=N(d)/n;
end
end
```

1. The sequence $x_n = n^\mu$.
We set $n = 1000$ and let μ grows larger. Then we obtain the following distributing frequency $F(d)$:

Table 2.

$F(d)$	$d=1$	$d=2$	$d=3$	$d=4$	$d=5$	$d=6$	$d=7$	$d=8$	$d=9$
$\mu=2$	0.194	0.146	0.123	0.111	0.097	0.091	0.084	0.078	0.076
$\mu=5$	0.256	0.165	0.129	0.100	0.088	0.077	0.070	0.058	0.057
$\mu=10$	0.278	0.173	0.125	0.100	0.082	0.073	0.062	0.054	0.053
$\mu=100$	0.303	0.178	0.125	0.095	0.079	0.065	0.060	0.051	0.044

When μ grows larger, Table 2 seems to hint that $\{n^\mu\}$ is closer to Benford's law.

2. The sequence $x_n = \tan n$.

Table 3.

$F(d)$	$d=1$	$d=2$	$d=3$	$d=4$	$d=5$	$d=6$	$d=7$	$d=8$	$d=9$
$n=10^3$	0.306	0.170	0.114	0.099	0.078	0.064	0.069	0.049	0.051
$n=10^4$	0.305	0.169	0.115	0.099	0.072	0.067	0.068	0.054	0.050
$n=5*10^4$	0.309	0.169	0.118	0.094	0.079	0.069	0.060	0.054	0.049
$n=10^5$	0.309	0.169	0.119	0.094	0.080	0.068	0.060	0.058	0.048

The sequence $\{\tan(n)\}$ shows some symptom of Benford's law.

3. The sequence $x_n = n^\mu * \sin(n)$.
We set $n = 1000$ and let μ varies.

Table 4.

$F(d)$	$d=1$	$d=2$	$d=3$	$d=4$	$d=5$	$d=6$	$d=7$	$d=8$	$d=9$
$\mu=2$	0.248	0.181	0.136	0.099	0.096	0.085	0.065	0.055	0.076
$\mu=5$	0.285	0.177	0.144	0.100	0.077	0.066	0.060	0.047	0.044
$\mu=10$	0.276	0.197	0.123	0.097	0.078	0.065	0.063	0.056	0.045
$\mu=100$	0.305	0.170	0.136	0.082	0.085	0.056	0.071	0.050	0.045

One can investigate many more. The difficulty is to prove the statement rigorously. The author keeps his interests on this problem.

Acknowledgment. The author would like to thank Mr. Lei Hou for his guidance and instruction. He also appreciates constant encouragement and support from his English tutor Ms. Ying Shen.

References

[1] F. Benford, *The law of anomalous numbers.* Proc. Am. Philos. Soc. **78** (1938), 551–572.
[2] W. G. Brady, *More on Benford's law*, The Fibonacci Quarterly **16** (1978), 51–52.
[3] S. Newcomb, *Note on the frequency of use of the different digits in natural numbers.* American Journal of Mathematics, **4** (1881), 39–40.
[4] E. M. Stein, R. Shakarchi, *Fourier Analysis*Preceton University Press2006 (For sale only in PRC)
[5] L. C. Washington, *Benford's Law for Fibonacci and Lucas Numbers*, The Fibonacci Quarterly, **19** (1981), 175–177.

[6] J. Wlodarski, *Fibonacci and Lucas numbers tend to obey Benford's law.* The Fibonacci Quarterly **9** (1971), 87–88.