Ann. Inst. Statist. Math.
36 (1984), Part A, 337-352

ON BENFORD'S LAW

KENJI NAGASAKA

(Received Sep. 28, 1983; revised Dec. 27, 1983)

Summary

It seems empirically that the first digits of random numbers do not occur with equal frequency, but that the earlier digits appear more often than the latters. This peculiality was at first noticed by F. Benford, hence this phenomenon is called Benford's law.

In this note, we fix the set of all positive integers as a model population and we sample random integers from this population according to a certain sampling procedure. For polynomial sampling procedures, we prove that random sampled integers do not necessarily obey Benford's law but their Banach limit does. We also prove Benford's law for geometrical sampling procedures and for linear recurrence sampling procedures.

1. Introduction and notations

It has long been known empirically that in most statistical tables the proportion of numbers with the first significant digit equal to or less than k ($k=1, 2, \cdots, 9$) is approximately $\log_{10}(k+1)$. This phenomenon suggested us that the first significant digit from numerical tables did not occur with equal frequency, but that the earlier digits appeared more often than the latters.

After making many counts from a large body of physical data, Farmer's Almanac, Census Reports, Chemical Rubber Handbook, etc., F. Benford [2] first noticed this peculiarity, hence this logarithmic law for the first significant digits is called Benford's law. This law does not hold, of course, in every numerical table.

Several authors, such as W. H. Furry and H. Hurwitz [8], S. A. Goudsmit and Furry [9], R. S. Pinkham [13] and R. A. Raimi [14] have sought the explanation of this phenomenon by assuming that all physical constants are selected from a population with some underlying distribution and have shown that certain assumptions about this distribution lead to the logarithmic law.

To B. J. Flehinger [7] it occurred that the smallest population which contains the set of significant figures of all possible physical constants, past, present and future, must be the set of all positive integers. The explanation for Benford's law should, therefore, lie in the properties of the set of integers as represented in a radix number system.

As far as considering the distribution of the first significant digits it is quite natural to the author also that we restrict ourselves to the set of all positive integers. But this restriction does not lead automatically the smallestness of the population containing all possible physical constants.

In this paper we fix the population, as a model, the set of all positive integers which would contain significant numbers of all possible numerical constants. From this population, that is the set of all positive integers, we sample random integers according to a certain sampling procedure. What kind of sampling procedures produce the set of random intergers which obey Benford's law, it is our problem treated in this paper.

A sampling procedure C is identified to the set of sampled integers and c_n is the nth smallest sampled integer of C. The number of sampled integers c_n, $1 \leqq n \leqq N$, with the first digit less than or equal to k $(k=1, 2, \cdots, 9)$ is denoted by $A_N(C; k)$ and we define

$$P_N(C; k) = P_N^1(C; k) = \frac{A_N(C; k)}{N} .$$

Flehinger considered successive cumulative averages (Hölder sums) of $P_N(N; k)$:

$$P_N^l(N; k) = \frac{1}{N} \sum_{M=1}^{N} P_N^{l-1}(N; k) , \qquad l=2, 3, \cdots ,$$

for $C=N$ and proved that

$$\lim_{l\to\infty} \lim_{N\to\infty} P_N^l(N; k) = \log_{10}(k+1) .$$

The limit of successive cumulative averages is called Banach limit. R. L. Adler and A. G. Konheim [1] proved that the Banach limit is finitely additive measure on the set of all positive integers and assigns zero measure on every finite subset of positive integers. Hence it is quite natural to consider the Banach limit of $P_N(N; k)$. Roughly speaking, the probability that a random integer has initial digit less than or equal to k seems to be $\log_{10}(k+1)$.

Linear recurrence sampling procedures have been considered with special reference to Fibonacci numbers. R. L. Duncan [6] proved that the sequence $\{\log F_n\}_{n=1,2,\ldots}$ is uniformly distributed mod 1, where F_n is the nth Fibonacci numbers, which signifies that Benford's law holds for Fibonacci numbers. L. Kuipers [10] gave another proof of Duncan's result.

J. L. Brown, Jr. and R. L. Duncan [4] and L. Kuipers and J-S. Shiue [12] extended the results in [6] and [10] and proved that the sequence $\{\log V_n\}_{n=1,2,\ldots}$ is uniformly distributed mod 1, where V_n satisfies a linear recurrence formula with some restrictions.

J. Wlodarski [17] and W. Brady [3] observed empirically the distribution of the first digits of Fibonacci numbers and L. C. Washington [16] reproved Benford's law for Fibonacci and Lucas numbers.

R. A. Raimi [15] reviewed the first digit problem making clearer the hypotheses and results with a nearly complete bibliography. P. Diaconis [5] also tested Benford's law for a large class of arithmetic sequences.

In this paper we shall deal with various sampling procedures by examining for the resulting sampled integers whether Benford's law holds or not. In the next section we shall consider linear sampling procedures and prove that their Banach limit obeys Benford's law. Geometrical sampling procedures will be considered in the third section and Benford's law will be shown to hold for geometrical sampling procedures. We shall consider linear recurrence sampling formulae in the fourth section and prove Benford's law except some special cases. In the last section we shall consider polynomial sampling procedures which are exceptions of linear recurrence sampling procedures. We shall prove that any polynomial sampling procedure does not necessarily obey Benford's law but their Banach limit does so.

Throughout in the following we shall write

$$\log x = \log_{10} x \quad \text{and} \quad \ln x = \log_e x .$$

2. Linear sampling procedures

In this section we shall consider a linear sampling procedure $A(a, d)$. $A(a, d)$ is a subset of positive integers of the form $a_n = a + (n-1)\cdot d$, where a and d are positive integers. The simplest case: $a=1$ and $d=1$ was considered by Flehinger, which corresponds to the complete enumeration of aggragates.

Firstly we consider the case for $d=1$ and arbitrary a. Secondaly the case for $a=1$ and any d. Summing up these two cases we shall treat a general case.

Let us define $A = A(a, 1)$, then the nth smallest integer $a_n = a + n - 1$. Suppose that the initial value a has g digits. The number $\delta(a)$ of

g-digits positive integers greater than or equal to a whose initial digits are less than or equal to k is

$$\delta(a)=\begin{cases}(k+1)\cdot 10^{g}-a\,, & \text{if } 10^{g-1}\leqq a<(k+1)\cdot 10^{g-1}\\ 0\,, & \text{if } (k+1)\cdot 10^{g-1}\leqq a<10^{g}\,.\end{cases}$$

Let N be the number of sampled integers by the procedure $A(a, 1)$, then $a_N=N+a$ and

$$\begin{aligned}A_N(A\,;k)&=\delta(a)+k\cdot 10^{g}\frac{(10^{j-g}-1)}{9}+N+a-10^{j}\\ &\qquad \text{for } 10^{j}-a\leqq N<(k+1)\cdot 10^{j}-a\,,\\ &=\delta(a)+k\cdot 10^{g}\frac{(10^{j-g+1}-1)}{9}\\ &\qquad \text{for } (k+1)\cdot 10^{j}-a\leqq N<10^{j+1}-a\,.\end{aligned}$$

Thus we obtain

$$\begin{aligned}P_N(A\,;k)&=A_N(A\,;k)/N\\ &=1-\frac{(9-k)\cdot 10^{j}}{9N}+\frac{9\delta(a)+9a-k\cdot 10^{g}}{9N}\\ &\qquad \text{for } 10^{j}-a\leqq N<(k+1)\cdot 10^{j}-a\,,\\ &=\frac{k\cdot 10^{j+1}}{9N}+\frac{9\delta(a)-k\cdot 10^{g}}{9N}\\ &\qquad \text{for } (k+1)\cdot 10^{j}-a\leqq N<10^{j}\,.\end{aligned}$$

If we replace N by $N+a$, the value $P_{N+a}(A\,;k)$ differs from $P_N(A\,;k)$ at most $O(N^{-1})$. Thus we can deduce

$$Q(\alpha\,;k)=\lim_{j\to\infty} P^{1}_{\alpha\cdot 10^{j}}(A\,;k)=\begin{cases}1-\dfrac{9-k}{9\alpha}\,, & 1\leqq\alpha<k+1\\[2ex] \dfrac{10k}{9\alpha}\,, & k+1\leqq\alpha<10\,.\end{cases}$$

Using the same argument as Flehinger, we have

$$\lim_{l\to\infty}\lim_{N\to\infty} P^{l}_N(A\,;k)=\log(k+1)\,.$$

Let us define $B=A(1, d)$ and the nth smallest integer $b_n=1+(n-1)\cdot d$. If a sampled integer b_N satisfies

$$10^{j}\leqq b_N<10^{j+1}\,,$$

then

$$\left[\frac{10^j}{d}\right] \leqq N < \left[\frac{10^{j+1}}{d}\right],$$

where $[x]$ is the greatest integer less than or equal to x. The number $A_N(B; k)$ of sampled integers with their first digit less than or equal to k is

$$\begin{aligned} A_N(B; k) &= \sum_{i=1}^{j-1} \left[\frac{k \cdot 10^i}{d}\right] + \left[\frac{b_N - 10^j}{d}\right] \\ &\qquad \text{for } \left[\frac{10^j}{d}\right] \leqq N < \left[\frac{(k+1) \cdot 10^j}{d}\right], \\ &= \sum_{i=1}^{j} \left[\frac{k \cdot 10^i}{d}\right] \\ &\qquad \text{for } \left[\frac{(k+1) \cdot 10^j}{d}\right] \leqq N < \left[\frac{10^{j+1}}{d}\right]. \end{aligned}$$

Then

$$\begin{aligned} P_N(B; k) &= 1 - \frac{(9-k) \cdot 10^j}{9dN} + O(N^{-1}) \\ &\qquad \text{for } \left[\frac{10^j}{d}\right] \leqq N < \left[\frac{(k+1) \cdot 10^j}{d}\right], \\ &= \frac{10k \cdot 10^j}{9dN} + O(N^{-1}) \\ &\qquad \text{for } \left[\frac{(k+1) \cdot 10^j}{d}\right] \leqq N < \left[\frac{10^{j+1}}{d}\right], \end{aligned}$$

since $[x] = x - \{x\}$ where $0 \leqq \{x\} < 1$. Thus we obtain

$$Q(\alpha; k) = \lim_{j \to \infty} P^1_{\alpha \cdot 10^j/d}(B; k) = \begin{cases} 1 - \dfrac{9-k}{9\alpha}, & 1 \leqq \alpha < k+1 \\[2ex] \dfrac{10k}{9\alpha}, & k+1 \leqq \alpha < 10. \end{cases}$$

Replacing α by α/d in Flehinger's calculation, we have

$$\lim_{l \to \infty} \lim_{N \to \infty} P^l_N(B; k) = \log(k+1).$$

For a general sampling procedure $A(a, d)$, we can conclude by almost the same arguement as used before:

THEOREM 2.1. *$\lim_{l \to \infty} \lim_{N \to \infty} P^l_N(A(a, d); k) = \log(k+1)$, where $A(a, d) = \{a + (n-1) \cdot d;\ n = 1, 2, \cdots\}$ for any positive integer a and d.*

3. Geometrical Sampling procedures

In this section we shall consider a geometrical sampling procedure $G=G(c, r)$, which is a subset of positive integers of the form $g_n=c\cdot r^{n-1}$ where c and $r\neq 1$ are positive integers. Put $B_N(G(c, r); k)$ be the number of sampled integers g_n for $1\leqq n\leqq N$ with their first digit equal to k, then we have

THEOREM 3.1. $\lim_{N\to\infty}\dfrac{B_N(G(c, r); k)}{N}=\log(k+1)-\log k$, *except for the case* $r=10^m$ *for some positive integer* m.

COROLLARY. $\lim_{N\to\infty}\dfrac{A_N(G(c, r); k)}{N}=\log(k+1)$, *except for the same case as in Theorem* 3.1.

In order to prove Theorem 3.1 we need some lemmas.

LEMMA 3.1. *Let r be a positive integer greater than* 1. *Then* $\log r$ *is irrational except for* $r=10^m$ *for some nonnegative integer* m.

LEMMA 3.2. *Let r be a positive integer greater than* 1 *and not of the form* 10^m *for some nonnegative integer* m. *Then the sequence* $\{n\cdot\log r\}_{n=1,2,\dots}$ *is uniformly distributed mod* 1.

(*L. Kuipers and H. Niederreiter* [11], *Example* 2.1.)

LEMMA 3.3. *If the sequence* $\{x_n\}_{n=1,2,\dots}$ *is uniformly distributed mod* 1, *then the sequence* $\{x_n+\alpha\}_{n=1,2,\dots}$, *where α is a real constant, is uniformly distributed mod* 1. (*Kuipers and Niederreiter, Lemma* 1.1.)

PROOF OF THEOREM 3.1. The necessary and sufficient condition for the first digit g_n to be k is

$$k\cdot 10^m\leqq g_n<(k+1)\cdot 10^m$$

for some nonnegative integer m, where $g_n=c\cdot r^{n-1}$. Taking logarithms these inequalities above, we have

$$\log k+m\leqq\log c+(n-1)\cdot\log r<\log(k+1)+m\,.$$

Suppose that r is not of the form 10^l for l a positive integer. Then the sequence $\{(n-1)\cdot\log r\}_{n=1,2,\dots}$ is uniformly distributed mod 1 by Lemma 3.1 and Lemma 3.2. From Lemma 3.3, the sequence $\{\log c+(n-1)\cdot\log r\}_{n=1,2,\dots}$ is uniformly distributed. Taking account of the necessary and sufficient condition, we obtain

$$\lim_{N\to\infty}\frac{B_N(G(c, r); k)}{N}=\log(k+1)-\log k\,.$$

If $r=10^m$ for some nonnegative integer m, then the first digit of $g_n=c\cdot 10^{n-1}$ is identical to the first digit of c, that is, the logarithmic law does not hold for any geometrical sampling procedure $G(c, 10^m)$. (Q.E.D.)

For a linear sampling procedure $A(a, d)$, the sequence of sampled integers $\{a_n\}$ satisfies a linear recurrence formula of first order:

$$a_{n+1}=a_n+d .$$

For a geometrical sampling procedure $G(c, r)$ $(r\neq 1)$, the sequence of sampled integers $\{g_n\}$ satisfies

$$g_{n+1}=r\cdot g_n .$$

It is natural to ask for a general linear recurrence formula of first order:

$$h_{n+1}=r\cdot h_n+s ,$$

where $r\neq 1$ and s are positive integers. Then

$$h_n=c\cdot r^{n-1}+d ,$$

where $d=s/(1-r)$ and $c=h_1-d$. The first significant digit of h_n is identical to that of $c\cdot r^{n-1}$ for every sufficiently large n. Thus we obtain

THEOREM 3.2. *Let $C=C(r, s, t)$ be a sampling procedure for which the sequence of sampled integer $\{c_n\}$ satisfies the linear recurrence formula of first order*

$$c_{n+1}=r\cdot c_n+s .$$

Then

$$\lim_{N\to\infty} P_N(C; k)=\log(k+1) ,$$

except for the case $r=10^m$ with m some nonnegative integer.

4. Linear recurrence sampling procedures

In the preceding sections we have proved that linear recurrence integer sequences of first order obey Benford's law in the sense of Banach limit. In this section we shall consider a linear recurrence sampling procedure $L(d, \boldsymbol{a}, \boldsymbol{c})$ which is a subset of positive integers and the nth smallest integer of $L(d, \boldsymbol{a}, \boldsymbol{c})$ is denoted by u_n. The sequence of sampled integers $\{u_n\}$ satisfies the following linear recurrence formula of order d:

(4.1) $$u_{n+d}=a_{d-1}\cdot u_{n+d-1}+a_{d-2}\cdot u_{n+d-2}+\cdots+a_0\cdot u_n\,, \qquad (n\geqq 1)\,,$$

and also the initial conditions:

(4.2) $$u_1=c_1\,,\quad u_2=c_2\,,\quad \cdots,\quad \text{and}\quad u_d=c_d\,,$$

where

$$\boldsymbol{a}=(a_{d-1}, a_{d-2}, \cdots, a_0) \quad\text{and}\quad \boldsymbol{c}=(c_1, c_2, \cdots, c_d)$$

are d-dimensional integral vectors.

We consider the case: $d=2$. Duncan [6], Kuipers [10] and Washington [16] proved that for $\boldsymbol{a}=(1, 1)$, Fibonacci and Lucas numbers satisfy Benford's law. Now we consider the following recurrence formula:

$$u_{n+2}=a_2\cdot u_{n+1}+a_1\cdot u_n\,, \qquad n\geqq 1\ (a_1\cdot a_2\neq 0)\,,$$

and its characteristic equation is

(4.3) $$\lambda^2=a_2\cdot\lambda+a_1\,.$$

THEOREM 4.1. *If the characteristic equation* (4.3) *has two real distinct roots* α *and* β *with* $|\alpha|\geqq|\beta|$ *and* α *is not of the form* $\pm 10^m$ *for any nonnegative integer* m, *then* $\{u_n\}_{n=1,2,\ldots}$ *obeys Benford's law.*

PROOF. In this case, the discriminant of (4.3)

(I) $D=a_2^2+4a_1>0$,

then the nth term u_n can be represented by

$$u_n=A\cdot\alpha^{n-1}+B\cdot\beta^{n-1}\,, \qquad n\geqq 1\,,$$

where A and B are constants depending only on a_1, a_2, u_1 and u_2.

Suppose further that $|\alpha|>|\beta|$. Then we have, for sufficiently large n,

$$\begin{aligned}\log u_n&=\log(A\cdot\alpha^{n-1}+B\cdot\beta^{n-1})\\ &=\log A\cdot\alpha^{n-1}\left\{1+\frac{B}{A}\left(\frac{\beta}{\alpha}\right)^{n-1}\right\}\\ &=\log|A|+(n-1)\cdot\log|\alpha|+\log\left|1+\frac{B}{A}\left(\frac{\beta}{\alpha}\right)^{n-1}\right|\,,\end{aligned}$$

and

$$\lim_{n\to\infty}\log\left|1+\frac{B}{A}\left(\frac{\beta}{\alpha}\right)^{n-1}\right|=0\,.$$

LEMMA 4.1. *If the sequence* $\{x_n\}_{n=1,2,\ldots}$ *is uniformly distributed mod 1, and if* $\{y_n\}$ *is a sequence with the property* $\lim_{n\to\infty}(x_n-y_n)=\gamma$, *a real*

constant, then $\{y_n\}_{n=1,2,\ldots}$ *is uniformly distributed mod* 1.
(*Kuipers and Niederreiter, Theorem* 1.2.)

LEMMA 4.2. *If* α *is an algebraic number of degree* 2 *or a rational number except for the case* $\alpha=\pm 10^m$ *for any nonnegative integer* m, *the sequence* $\{(n-1)\cdot\log\alpha\}_{n=1,2,\ldots}$ *is uniformly distributed mod* 1.

Combining Lemma 3.2, Lemma 4.1 and Lemma 4.2, we see that the sequence $\{\log u_n\}_{n=1,2,\ldots}$ is uniformly distributed mod 1, which means immediately that Benford's law holds for the sequence $\{u_n\}_{n=1,2,\ldots}$.

If $0<\alpha=|\alpha|=|\beta|$, then $\beta=-\alpha$ (because $\alpha\neq\beta$), and

$$u_n=\begin{cases} A_1\cdot\alpha^{2m}, & \text{for } n=2m+1, \\ B_1\cdot\alpha^{2m}, & \text{for } n=2m+2, \end{cases}$$

where A_1 and B_1 are constants depending only on a_1, a_2, u_1 and u_2. From Lemma 4.2 we see that two sequences $\{\log u_n\}_{n=1,3,5,\ldots}$ and $\{\log u_n\}_{n=2,4,6,\ldots}$ are both uniformly distributed mod 1, from which we can easily deduce that the sequence $\{\log u_n\}_{n=1,2,3,\ldots}$ is uniformly distributed mod 1. This signifies that Benford's law holds for the sequence $\{u_n\}_{n=1,2,3,\ldots}$, which completes the proof. (Q.E.D.)

(II) $D=a_2^2+4a_1=0$.

In this case the characteristic equation (4.1) has only one real root α and u_n can be represented as

$$u_n=(A\cdot n+B)\cdot\alpha^{n-1}, \qquad n\geqq 1,$$

where A and B are constants depending only upon a_1, a_2, u_1 and u_2. Then we get, for sufficiently large n,

$$\begin{aligned}\log u_n&=\log(A\cdot n+B)\cdot\alpha^{n-1}\\&=\log\left\{A\cdot n\cdot\alpha^{n-1}\left(1+\frac{B}{A\cdot n}\right)\right\}\\&=\log|A|+\log n+(n-1)\cdot\log|\alpha|+\log\left|1+\frac{B}{A\cdot n}\right|.\end{aligned}$$

LEMMA 4.3. *Let* $f(x)$ *be a function defined for* $x\geqq 1$ *that is differentiable for sufficiently large* x. *If*

$$\lim_{x\to\infty} f'(x)=\theta \quad (\textit{irrational}),$$

then the sequence $\{f(n)\}_{n=1,2,\ldots}$ *is uniformly distributed mod* 1.
(*Kuipers and Niederreiter, Exercise* 3.5.)

By this Lemma 4.3, we know that the sequence $\{\log n+(n-1)\cdot$

$\log|\alpha|\}_{n=1,2,\dots}$ is uniformly distributed mod 1. Then we obtain

THEOREM 4.2. *If the characteristic equation* (4.1) *has a double real root* α *which is not of the form* $\pm 10^m$ *for any nonnegative integer* m, *then* $\{u_n\}_{n=1,2,\dots}$ *obeys Benford's law.*

Now we consider the general recurrence sampling procedure $L(d, \boldsymbol{a}, \boldsymbol{c})$ with its recurrence formula (4.1) and the initial conditions (4.2). The characteristic equation of (4.1) is

$$\lambda^d = a_{d-1}\cdot\lambda^{d-1} + a_{d-2}\cdot\lambda^{d-2} + \cdots + a_1\cdot\lambda + a_0 . \tag{4.4}$$

THEOREM 4.3. *The characteristic equation* (4.4) *has roots* $\alpha_1, \alpha_2, \cdots, \alpha_p$ *with multiplicity* $m_1, m_2, \cdots, m_p$, *respectively. Suppose that*

$$|\alpha_1| > |\alpha_2| \geqq |\alpha_3| \geqq \cdots \geqq |\alpha_p| \tag{4.5}$$

and α_1 *is a real root, which is not of the form* $\pm 10^m$ *for any nonnegative integer* m. *Then the linear recurrence sampling procedure* $L(d, \boldsymbol{a}, \boldsymbol{c})$ *obeys Benford's law.*

PROOF. The nth smallest sampled integer of $L(d, \boldsymbol{a}, \boldsymbol{c})$ u_n can be represented by

$$u_n = b_1(n-1)\cdot\alpha_1^{n-1} + b_2(n-1)\cdot\alpha_2^{n-2} + \cdots + b_p(n-1)\cdot\alpha_p^{n-1} ,$$

where $b_1, b_2, \cdots$ and b_p are polynomials of degree at most $m_1-1, m_2-1, \cdots$ and m_p-1, respectively, which depend on $\boldsymbol{a}$, $\boldsymbol{c}$, $\alpha_1, \alpha_2, \cdots$ and α_p.

Let us put

$$b_1(n-1) = e_l\cdot n^l + e_{l-1}\cdot n^{l-1} + \cdots + e_1\cdot n + e_0 ,$$

where $e_l \neq 0$ and $0 \leqq l \leqq m_1 - 1$. Then

$$\begin{aligned}\log u_n = &\log|b_1(n-1)| + (n-1)\cdot\log|\alpha_1| \\ &+\log\left|1 + \frac{b_2(n-1)\cdot\alpha_2^{n-1}}{b_1(n-1)\cdot\alpha_1^{n-1}} + \cdots + \frac{b_p(n-1)\cdot\alpha_p^{n-1}}{b_1(n-1)\cdot\alpha_1^{n-1}}\right| .\end{aligned}$$

From (4.5), the third term tends to zero as n goes to infinity. Here

$$\begin{aligned}\log|b_1(n-1)| &= \log\left(|e_l\cdot n^l|\cdot\left|1 + \frac{e_{l-1}}{e_l\cdot n} + \cdots + \frac{e_1}{e_l\cdot n^{l-1}} + \frac{e_0}{e_l\cdot n^l}\right|\right) \\ &= \log|e_l| + l\cdot\log n + O(n^{-1}) ,\end{aligned}$$

and $\log|\alpha_1|$ is irrational and from Lemma 4.3, and so the sequence $\{\log u_n\}_{n=1,2,\dots}$ is uniformly distributed mod 1, which completes the proof. (Q.E.D.)

Remark. If the root α_2 is also real, then we may substitute the condition (4.5) by

$$|\alpha_1| \geqq |\alpha_2| > |\alpha_3| \geqq \cdots \geqq |\alpha_p| . \tag{4.6}$$

In the next section we only treat polynomial sampling procedures which are special cases of $L(d; \boldsymbol{a}, \boldsymbol{c})$ but have their characteristic equation of special form. For example, a sequence of square sampled integer $\{n^2\}$ satisfies the recurrence formula:

$$u_{n+3} = 3 \cdot u_{n+2} - 3 \cdot u_{n+1} + u_n . \tag{4.7}$$

The characteristic equation of (4.7) is

$$(\lambda - 1)^3 = 0 ,$$

which has only one root $1 = 10^0$ and this is an exceptional case for Theorem 4.3.

5. Polynomial sampling procedures

In this section we consider a polynomial sampling procedure $P(d, \boldsymbol{a})$. The nth smallest sampled integer

$$p_n = a_d \cdot n^d + a_{d-1} \cdot n^{d-1} + \cdots + a_1 \cdot n + a_0 ,$$

where $a_d \neq 0$, a_{d-1}, $\cdots$, a_1 and a_0 are integers.

Firstly we consider a monomial sequence

$$p_n = a_d \cdot n^d ,$$

then

$$\log p_n = \log |a_d| + d \cdot \log n .$$

LEMMA 5.1. *The sequence* $\{c \cdot \log n\}_{n=1,2,\ldots}$, *where* c *is a constant, is not uniformly distributed mod* 1.

(*Kuipers and Niederreiter, Exercise* 2.13)

By using the contraposition of Lamma 3.3, we conclude that, for any monomial sequence P_n, $\{\log P_n\}_{n=1,2,\ldots}$ is not uniformly distributed mod 1. For $d = a_d = 1$, Flehinger pointed that $P_N(C; k)$ does not have the limit as n tends to the infinity, so that the Banach limit of $P_N(C; k)$ is indispensable.

For a general polynomial sampling procedure $P(d, \boldsymbol{a})$,

$$\begin{aligned} \log p_n &= \log (a_d \cdot n^d + a_{d-1} \cdot n^{d-1} + \cdots + a_1 \cdot n + a_0) \\ &= \log |a_d| + d \cdot \log n + \log (1 + O(n^{-1})) . \end{aligned}$$

Combining Lemma 5.1 and the contraposition of Lemma 4.1, we obtain

Remark. Any polynomial integer sequence $\{P_n\}_{n=1,2,\dots}$ does not necessarily obey Benford's law.

On the other hand, we have

THEOREM 5.1. *For every polynomial sampling procedure $P(d, \boldsymbol{a})$,*

$$\lim_{l\to\infty}\lim_{N\to\infty} P_N^l(P(d, \boldsymbol{a}); k) = \log(k+1).$$

PROOF. Let us consider a monic monomial sampling procedure $P_1(d, \boldsymbol{a})$ with $p_n = n^d$ ($P_1(d, \boldsymbol{a})$ is written simply n^d for abbreviation).

The number of $p_n = n^d$ with the first digits less than or equal to k satisfying

$$10^i \leqq n^d < 10^{i+1}, \qquad \text{for } i \text{ nonnegative integer},$$

is equal to

$$\{(k+1)^{1/d}-1\}\cdot 10^{1/d} + O(1).$$

Then we have

$$\begin{aligned} P_N^1(P_1; k) &= P_N^1(n^d; k) \\ &= 1 - \frac{10^{j/d}\{10^{1/d}-(k+1)^{1/d}\}}{N(10^{1/d}-1)} + O(N^{-1}) \\ &\qquad \text{for } 10^{j/d} \leqq N < (k+1)^{1/d}\cdot 10^{j/d} \\ &= \frac{10^{1/d}\{(k+1)^{1/d}-1\}\cdot 10^{j/d}}{N(10^{1/d}-1)} + O(N^{-1}) \\ &\qquad \text{for } (k+1)^{1/d}\cdot 10^{j/d} \leqq N < 10^{(j+1)/d}. \end{aligned}$$

For another monic polynomial sampling procedure $P_2(d, \boldsymbol{a})$ with $p_n = n^d + \sum_{i=0}^{d-1} a_i \cdot n^i$, we obtain

$$P_N^1(P_1; k) = P_N^1(P_2; k) + O(N^{-1}).$$

$O(N^{-1})$ term gives no affection on the calculation afterwards, so we may consider only a monic polynomial sampling procedure of the type n^d.

Let us define

$$\begin{aligned} (5.1) \qquad Q^1(n^d; \alpha, k) &= \lim_{j\to\infty} P^1_{\alpha\cdot 10^{j/d}}(n^d; k) \\ &= 1 - \frac{\{10^{1/d}-(k+1)^{1/d}\}}{\alpha\cdot(10^{1/d}-1)} \\ &\qquad \text{for } 1 \leqq \alpha < (k+1)^{1/d}, \\ &= \frac{10^{1/d}\{(k+1)^{1/d}-1\}}{\alpha\cdot(10^{1/d}-1)} \\ &\qquad \text{for } (k+1)^{1/d} \leqq \alpha < 10^{1/d}. \end{aligned}$$

We can observe that, for all l, $\lim_{j\to\infty} P^l_{\alpha\cdot 10^{j/d}}(n^d; k)$ exists and we put this limit equal to $Q^l(n^d; \alpha, k)$.

We now derive a formula for $Q^l(n^d; \alpha, k)$ for $l>1$.

$$
\begin{aligned}
(5.2)\quad & Q^l(n^d; \alpha, k) \\
&= \lim_{j\to\infty} P^l_{\alpha\cdot 10^{j/d}}(n^d; k) \\
&= \lim_{j\to\infty} \frac{1}{\alpha\cdot 10^{j/d}} \sum_{M=1}^{\alpha\cdot 10^{j/d}} P^{l-1}_M(n^d; k) \\
&= \lim_{j\to\infty} \frac{1}{\alpha\cdot 10^{j/d}} \left[\sum_{i=0}^{j-1} \sum_{10^{i/d}\leqq M<10^{(i+1)/d}} P^{l-1}_M(n^d; k) + \sum_{M=10^{j/d}}^{\alpha\cdot 10^{j/d}} P^{l-1}_M(n^d; k)\right] \\
&= \frac{1}{\alpha}\left[\sum_{s=1}^{\infty} 10^{-s/d} \int_1^{10^{1/d}} Q^{l-1}(n^d; \beta, k)d\beta + \int_1^{\alpha} Q^{l-1}(n^d; \beta, k)d\beta\right] \\
&= \frac{1}{\alpha}\left[\frac{1}{10^{1/d}-1} \int_1^{10^{1/d}} Q^{l-1}(n^d; \beta, k)d\beta + \int_1^{\alpha} Q^{l-1}(n^d; \beta, k)d\beta\right].
\end{aligned}
$$

Let us set

$$
\begin{aligned}
(5.3)\quad Q^l(n^d; \alpha, k) &= 1 - \frac{1}{\alpha}\sum_{i=0}^{l-1} \frac{c_{l-i}(n^d; k)\cdot(\ln \alpha)^i}{i!}, \\
&\qquad \text{for } 1\leqq\alpha<(k+1)^{1/d}, \\
&= \frac{1}{\alpha}\sum_{i=0}^{l-1} \frac{d_{l-i}(n^d; k)\cdot(\ln \alpha)^i}{i!}, \\
&\qquad \text{for } (k+1)^{1/d}\leqq\alpha<10^{1/d}.
\end{aligned}
$$

For $l=1$, wh have from (5.1)

$$
c_1(n^d; k) = \frac{10^{1/d}-(k+1)^{1/d}}{10^{1/d}-1},
$$

$$
d_1(n^d; k) = \frac{10^{1/d}\{(k+1)^{1/d}-1\}}{10^{1/d}-1}.
$$

Suppose that the expression for $Q^l(n^d; \alpha, k)$ (5.3) is valid for $l-1$, and substituting (5.3) into another formula for $Q^l(n^d; \alpha, k)$ (5.2) for $l>1$, we obtain

$$
\begin{aligned}
c_l(n^d; k) &= \frac{10^{1/d}-(k+1)^{1/d}}{10^{1/d}-1} \\
&\quad + \frac{1}{10^{1/d}-1}\sum_{i=1}^{l-1} \frac{[\ln (k+1)^{1/d}]^i}{i!}\cdot[c_{l-i}(n^d; k)+d_{l-i}(n^d; k)] \\
&\quad - \frac{1}{10^{1/d}-1}\sum_{i=1}^{l-1} \frac{[\ln 10^{1/d}]^i}{i!}\cdot d_{l-i}(n^d; k),
\end{aligned}
$$

$$
d_l(n^d; k) = \frac{10^{1/d}\{(k+1)^{1/d}\}}{10^{1/d}-1}
$$

$$-\frac{10^{1/d}}{10^{1/d}-1}\sum_{i=1}^{l-1}\frac{[\ln(k+1)^{1/d}]^i}{i!}\cdot[c_{l-i}(n^d;k)+d_{l-i}(n^d;k)]$$
$$+\frac{1}{10^{1/d}-1}\sum_{i=1}^{l-1}\frac{[\ln 10^{1/d}]^i}{i!}\cdot d_{l-i}(n^d;k)\,.$$

Consider the generating functions associated with $\{c_l(n^d;k)\}$ and $\{d_l(n^d;k)\}$:

$$C(n^d;Z,k)=\sum_{l=1}^{\infty}c_l(n^d;k)\cdot Z^l\,,$$

$$D(n^d;Z,k)=\sum_{l=1}^{\infty}d_l(n^d;k)\cdot Z^l\,.$$

Then they satisfy

$$C(n^d;Z,k)=\frac{10^{1/d}-(k+1)^{1/d}}{10^{1/d}-1}\cdot\frac{Z}{1-Z}$$
$$+\frac{1}{10^{1/d}-1}\cdot[C(n^d;Z,k)+D(n^d;Z,k)]\cdot[(k+1)^{Z/d}-1]$$
$$+\frac{1}{10^{1/d}-1}\cdot D(n^d;Z,k)\cdot[10^{Z/d}-1]\,.$$

By solving this functional equation for $C(n^d;Z,k)$ and $D(n^d;Z,k)$, we get

$$C(n^d;Z,k)=\frac{Z}{1-Z}\cdot\frac{10^{(1-Z)/d}-(k+1)^{(1-Z)/d}}{10^{(1-Z)/d}-1}\,,$$

$$D(n^d;Z,k)=\frac{Z}{1-Z}\cdot\frac{10^{(1-Z)/d}\cdot\{(k+1)^{(1-Z)/d}-1\}}{10^{(1-Z)/d}-1}\,.$$

We also consider the generating function associated with the set of functions $\{Q^l(n^d;\alpha,k)\}$:

$$\Phi(n^d;Z,\alpha,k)=\sum_{l=1}^{\infty}Q^l(n^d;\alpha,k)\cdot Z^l\,.$$

From the expression for $Q^l(n^d;\alpha,k)$ (5.3), we obtain

$$\Phi(n^d;Z,\alpha,k)=\begin{cases}\dfrac{Z}{1-Z}-\dfrac{C(n^d;Z,k)}{\alpha^{1-Z}}\,, & \text{for } 1\leqq\alpha<(k+1)^{1/d}\\[2ex] \dfrac{D(n^d;Z,k)}{\alpha^{1-Z}}\,, & \text{for } (k+1)^{1/d}\leqq\alpha<10^{1/d}\,.\end{cases}$$

The function $\Phi(n^d;Z,\alpha,k)$ has simple poles only at the points:

$$1+\frac{2^{d+1}\pi n i}{\ln 10}\,,\qquad n=0,\,\pm1,\,\pm2,\,\cdots\,.$$

Let Γ_1 is a circle centered at the origin with radius R, where $1<R<2$, and Γ_2 is a circle centered at 1 with radius $R-1$. Then

$$Q^l(n^d;\alpha)=\frac{1}{2\pi i}\int_{\Gamma_1}\frac{1}{Z^{l+1}}\cdot\Phi(n^d;Z,\alpha,k)dZ-\frac{1}{2\pi i}\int_{\Gamma_2}\frac{1}{Z^{l+1}}\cdot\Phi(n^d;Z,\alpha,k)dZ$$
$$=O(R^{-(l-1)})-\lim_{Z\to1}\frac{Z-1}{Z^{l+1}}\cdot\Phi(n^d;Z,\alpha,k).$$

Applying l'Hôpital's rule to the calculation of this limit, we obtain that, for $1\leqq\alpha<10^{1/d}$,

$$Q^l(n^d;\alpha)=O(R^{-(l-1)})+\ln(k+1)/\ln 10.$$

Thus we have

$$\lim_{l\to\infty}Q^l(n^d;\alpha)=\log(k+1),$$

for any α with $1\leqq\alpha<10^{1/d}$, which signifies that

$$\lim_{l\to\infty}\lim_{N\to\infty}P_N^l(n^d;k)=\log(k+1).$$

For a general polynomial sampling procedure $P(d,\boldsymbol{a})$ with $p_n=a_d\cdot n^d+\sum_{i=0}^{d-1}a_i\cdot n^i$ $(a_d\neq0)$, replacing α by α/a_d, we have

$$\lim_{l\to\infty}\lim_{N\to\infty}P_N^l(P(d;\boldsymbol{a});k)=\log(k+1),$$

which completes the proof of Theorem 5.1.

Acknowledgement

This work has been completed while I was visiting the Institute of Statistical Mathematics. I am grateful to Professor Chikio Hayashi of the Institute and also to the Shinsyu University for offering me this opportunity.

SHINSYU UNIVERSITY

REFERENCES

[1] Adler, R. L. and Konheim, A. G. (1963). Solution of Problem 4999, *Amer. Math. Monthly*, **70**, 218-219.

[2] Benford, F. (1938). The law of anomalous numbers, *Proc. Amer. Philos. Soc.*, **78**, 551-572.

[3] Brady, W. G. (1978). More on Benford's law, *Fibonacci Quart.*, **16**, 51-52.

[4] Brown, J. L. and Duncan, R. L. (1970). Modulo one uniform distribution of the sequence of logarithms of certain recursive sequences, *Fibonacci Quart.*, 8, 482-486.

[5] Diaconis, P. (1977). The distribution of leading digits and uniform distribution mod 1, *Ann. Prob.*, **5**, 72-81.

[6] Duncan, R. L. (1967). An application of uniform distribution to the Fibonacci numbers, *Fibonacci Quart.*, **5**, 137-140.
[7] Flehinger, B. J. (1966). On the probability that a random integer has initial digit A, *Amer. Math. Monthly*, **73**, 1056-1061.
[8] Furry, W. S. and Hurwitz, H. (1945). Distribution of numbers and distribution of significant figures, *Nature*, **155**, 52-53.
[9] Goudsmit, S. A. and Furry, W. H. (1944). Significant figures of numbers in statistical tables, *Nature*, **154**, 800-801.
[10] Kuipers, L. (1969). Remarks on a paper by R. L. Duncan concerning the uniform distribution mod 1 of the sequence of logarithms of the Fibonacci numbers, *Fibonacci Quart.*, **7**, 465-466, 473.
[11] Kuipers, L. and Niederreiter, H. (1974). *Uniform Distribution of Sequences*, Interscience-Wiley, New York-London-Sydney-Toronto.
[12] Kuipers, L. and Shiue, J-S. (1973). Remark on a paper by Duncan and Brown on the sequence of logarithms of certain recursive sequences, *Fibonacci Quart.*, **11**, 292-294.
[13] Pinkham, R. S. (1961). On the distribution of first significant digits, *Ann. Math. Stat.*, **32**, 1223-1230.
[14] Raimi, R. A. (1969). On the distribution of first significant figures, *Amer. Math. Monthly*, **76**, 342-348.
[15] Raimi, R. A. (1976). The first digit problem, *Amer. Math. Monthly*, **83**, 521-538.
[16] Washington, L. C. (1981). Benford's law for Fibonacci and Lucas numbers, *Fibonacci Quart.*, **19**, 175-177.
[17] Wlodarski, J. (1971). Fibonacci and Lucas numbers tend to obey Benford's law, *Fibonacci Quart.*, **9**, 87-88.