Math. J. Okayama Univ. 34 (1992), 225—232

SOME EXAMPLES OF BENFORD SEQUENCES

KAZUO GOTO

1. Introduction. It seems empirically that the first digits of random numbers do not occur with equal frequency. After making many counts from a large body of physical data, Farmer's Almanac, Cencus Reports etc, $\cdots$. F. Benford first noticed that the proportion of numbers with first significant digits equals to or less than k ($k = 1, 2, \cdots, 9$) is approximately $\log_{10}(k+1)$. Hence this logarithmic law for the first significant digits is called Benford's law. Tables 1 and 2 below are the examples. In Table 1,

$L(n) = \log_{10}(n+1) - \log_{10} n$,

$B(n) =$ the empirical frequency found by Benford (1938) is his ensemble of 20, 229 entries,

$P(n) =$ the frequency of leading digits n among the first hundred powers of 2, i.e., $2^0, 2^1, 2^2, \cdots, 2^{99}$.

Applying χ^2-test with $9-1 = 8$ degrees of freedom to Table 2, we obtain $\chi^2 = 5.59 < 15.5 = \chi^2_8(0.05)$. Therefore we should not have to reject the hypothesis: *the events do obey Benford's law.* Therefore we may say the events obey Benford's law.

In this paper we show another example of this type and also give a Benford sequence in the sense of natural density which is not a strong Benford sequence.

Table 1 (due to Raimi [3]).

n	1	2	3	4	5	6	7	8	9
Benford's law $L(n)$	.301	.176	.125	.097	.079	.067	.058	.051	.046
Benford's data $B(n)$	.306	.185	.124	.094	.080	.064	.051	.049	.047
Powers of two $P(n)$	.30	.17	.13	.10	.07	.07	.06	.06	.05

Table 2. Distribution of the first significant digits of the first page number of each paper appearing in the Bibliography of the book by Kuipers and Niederreiter [2] (total 866).

Range of n	1	2	3	4	5	6	7	8	9
Frequency of events	268	162	111	85	74	51	43	42	30
Expected frequency	261	153	108	84	69	58	50	44	40
Proportion	.3095	.1871	.1282	.0982	.0855	.0589	.0497	.0485	.0346
Expected proportion	.3010	.1761	.1249	.0969	.0792	.0669	.0580	.0512	.0458

2. Definitions. We denote the first significant (k-digit) b-adic expression by $d_b(x)$ (or shortly $d(x)$), i.e.

$$d_b(x) = [x/b^{[\log_b x]-k+1}] \quad \text{if} \quad x > 0.$$

Definition 1. We call that a sequence (x_n) obeys *Benford's law* if for $k = 1, 2, \cdots, b-1$

$$\lim_{N\to\infty}\frac{1}{N}\#\{n : 1 \leq n \leq N, d_b(x_n) = k\} = \log_b(k+1) - \log_b k.$$

Definition 2 (*cf.* [4]). Let $P = (p(n))$, $n = 1, 2, \cdots$, be a sequence of non-negative real numbers with $p(1) > 0$. For $N \geq 1$, we put $S(N) = p(1) + p(2) + \cdots + p(N)$. Then a sequence (x_n) is said to be $(M, p(n))$*-uniformly distributed mod 1*, if for every positive integer h,

$$\lim_{N\to\infty}\frac{1}{S(N)}\sum_{n=1}^{N} p(n)e^{2\pi i h x_n} = 0.$$

Definition 3. A positive sequence (a_n) is said to be *a strong Benford sequence* if $(\log a_n)$ is uniformly distributed mod 1.

Definition 4. A positive sequence (a_n) is said to be *a weak-Benford sequence* if $(\log a_n)$ is $(M, 1/n)$-uniformly distributed mod 1.

It was J. Cigler [*cf.* 1] who first proposed the notions of strong and resp. weak Benford sequence. He observed that if for a sequence (a_n) of positive reals, $(\log a_n)$ is uniformly distributed mod 1 (abbreviated, u.d. mod 1), then (a_n) obeys Benford's law in the sense of natural density of D_k in (a_n). This was proved by P. Diaconis in 1977 [1, Theorem 1].

Definition 5. We call that a sequence (x_n) obeys *k-th digit Benford's law to the base b* if for $j = b^{k-1}, b^{k-1}+1, \cdots, b^k - 1$

$$\lim_{N\to\infty}\frac{1}{N}\#\{n : 1 \leq n \leq N, d_b(X_n) = j\} = (\log_b(j+1) - \log_b j).$$

3. Results.

Theorem 1. *There exists a positive real sequence (u_n) which obeys Benford's law in the sense of natural density such that (u_n) is not a strong-Benford's sequence.*

Proof. We put $p(j) = \log_{10}(1+1/j)$ for $j = 1, 2, \cdots, 9$. Let (u_n) be a sequence of integers with repetition :

$$(u_n) = (\overbrace{\underbrace{1,1,\cdots,1}_{n_1(1)},\underbrace{2,2,\cdots,2}_{n_1(2)},\cdots,\underbrace{9,9,\cdots,9}_{n_1(9)}}^{\textit{first block}},\overbrace{\underbrace{1,1,\cdots,1}_{n_2(1)},\underbrace{2,2,\cdots,2}_{n_2(2)},\cdots,\underbrace{9,9,\cdots,9}_{n_2(9)}}^{\textit{second block}},\cdots,\overbrace{\underbrace{1,1,\cdots,1}_{n_m(1)},\cdots,\underbrace{j,j,\cdots,j}_{n_m(j)},\cdots,\underbrace{9,9,\cdots,9}_{n_m(9)}}^{m\textit{-th block}},\cdots),$$

where in each block, u_n takes the value j, $n_m(j) = [m(m+21)p(j)]$ times, for each j ($j = 1, 2, \cdots, 9$) where $[x]$ is the greatest integer $\leq x$. We remark that $n_m(j) \geq 1$ holds. If the term u_N is included in the m-th block, we may write

$$N = \textstyle\sum_{i=1}^{m-1}\sum_{j=1}^{9} n_i(j) + N_m,$$

where N_m is the number of elements u_n included in the m-th block and satisfies

$$1 \leq N_m \leq n_m(1)+n_m(2)+\cdots+n_m(9).$$

First we show that $N \sim m^2/2$ as $m \to \infty$. Suppose N_m satisfies

$$n_m(1)+\cdots+n_m(j-1)+1 \leq N_m \leq n_m(1)+\cdots+n_m(j),$$

for $j = 1, 2, \cdots, 9$. We consider the following three cases.

Case 1. For $k = 1, 2, \cdots, j-1$,

$$J_1 := \#\{u_n = k, 1 \leq n \leq N\} = \textstyle\sum_{i=1}^{m} n_i(k).$$

Case 2. For $k = j$,

$$\begin{aligned} J_2 &:= \#\{u_n = k, 1 \leq n \leq N\} \\ &= \textstyle\sum_{i=1}^{m-1} n_i(k) + N_m - (n_m(1)+\cdots+n_m(k-1)). \end{aligned}$$

Case 3. For $k = j+1, \cdots, 9$,

$$J_3 := \#\{u_n = k, 1 \leq n \leq N\} = \textstyle\sum_{i=1}^{m-1} n_i(k).$$

Now we estimate N. For $j = 1, 2, \cdots, 9$, since N_m satisfies

$$n_m(1)+\cdots+n_m(j-1)+1 \leq N_m \leq n_m(1)+\cdots+n_m(j),$$

we obtain

$$\begin{aligned} N &= \textstyle\sum_{i=1}^{m-1}\sum_{j=1}^{9} n_i(j) + N_m \leq \sum_{i=1}^{m-1}\sum_{j=1}^{9}(i+21)p(j) + \sum_{i=1}^{j}(m+21)p(i) \\ &= \textstyle\sum_{j=1}^{9} p(j)\{(m+20)(m+21)/2-6\} + (m+21)\sum_{i=1}^{9} p(i) \\ &\leq (m+20)(m+21)/2-6+(m+21). \end{aligned}$$

On the other hand,

$$\begin{aligned} N &\geq \sum_{i=1}^{m-1}\sum_{j=1}^{9}\{(i+21)p(j)-1\}+\sum_{i=1}^{j-1}\{(m+21)p(i)-1\} \\ &= \sum_{j=1}^{9}p(j)\sum_{i=1}^{m-1}(i+21)-9(m-1)+(m+21)\sum_{i=1}^{j-1}p(i)-(j-1) \\ &\geq (m+20)(m+21)/2-9(m-1)-(j-1). \end{aligned}$$

So we have $N \sim m^2/2$.

Next we show that J_1, J_2 and $J_3 \sim m^2p(k)/2$.

Case 1.

$$J_1 = \sum_{i=1}^{m} n_i(k) \geq \sum_{i=1}^{m}\{(i+21)p(k)-1\} = p(k)(m+21)(m+22)/2-m,$$

and

$$J_1 \leq \sum_{i=1}^{m}(i+21)p(k) = p(k)(m+21)(m+22)/2.$$

Consequently we have $J_1 \sim m^2p(k)/2$.

Case 2.

$$\begin{aligned} J_2 :&= \sum_{i=1}^{m-1} n_i(k)+N_m-(n_m(1)+\cdots+n_m(k-1)) \\ &\geq \sum_{i=1}^{m-1}\{(i+21)p(k)-1\} = p(k)(m+20)(m+21)/2-(m-1). \end{aligned}$$

$$\begin{aligned} J_2 &\leq \sum_{i=1}^{m-1}(i+21)p(k)+n_m(k) \leq \sum_{i=1}^{m-1}(i+21)p(k)+(m+21)p(k) \\ &= p(k)(m+20)(m+21)/2+(m+21)p(k). \end{aligned}$$

Consequently we have $J_2 \sim m^2p(k)/2$.

Case 3. Similarly, we have

$$\sum_{i=1}^{m-1} n_i(k) \sim m^2p(k)/2.$$

Hence J_1, J_2 and $J_3 \sim m^2p(k)/2$ $(N \to \infty)$.

Thus we obtain J_1/N, J_2/N and $J_3/N \to p(k)$ $(m \to \infty)$.

Therefore the sequence (u_n) obeys Benford's law in the sense of natural density. But obviously, the sequence $(\log u_n)$ only consists of the set $\{\log 1, \log 2, \log 3, \cdots, \log 9\}$. Thus the sequence $(\log u_n)$ is not uniformly distributed mod 1. This completes the proof.

For a k-th digit problem to the base b, we may have easily a similar result, i.e.,

Theorem 2. *Let (u_n) be a sequence of integers with repetition:*

$$(u_n) = (\overbrace{\underbrace{c, c, \cdots, c}_{n_1(1)}, \underbrace{c+1, c+1, \cdots, c+1}_{n_1(2)}, \cdots, \underbrace{b^k-1, \cdots, b^k-1}_{n_1(b^k-1)}}^{\textit{first block}}, \overbrace{\underbrace{c, c, \cdots, c}_{n_2(1)}, \cdots}^{\textit{second block}},$$

$$\cdots, \overbrace{\underbrace{c, c, \cdots, c}_{n_m(1)}, \cdots, \underbrace{j, j, \cdots, j}_{n_m(j)}, \cdots, \underbrace{b^k-1, \cdots, b^k-1}_{n_m(b^k-1)}}^{m\text{-th block}}, \cdots),$$

where $c = b^{k-1}$, $n_m(j) = [(m+a)p(j)]$, $p(j) = \log_b(1+1/j)$ *and* $a = (1-\log(b-1))^{-1}$, *then* (u_n) *obeys the k-th digit Benford's law to the base b in the sense of natural density, but* $(\log u_n)$ *is not uniformly distributed mod 1.*

R. E. Whitney [5] showed that the sequence of primes is a weak but not strong Benford sequence. We show an example of this type, whch is not a monotone sequence.

Example. *Let us define*

$$x_n = \begin{cases} n & \text{if } n \neq k^2 \text{ for every integer } k, \\ \exp((\log n)^2) & \text{otherwise.} \end{cases}$$

Then (x_n) *is a weak, but not strong Benford sequence.*

Proof. First we show that $(\log x_n)$ is not uniformly distributed mod 1. We estimate

$$\textstyle\sum_{1 \leq n \leq N} e^{2\pi ih \log x_n}.$$

By Euler summation formula, we have

(*) $\sum_{n=1}^{N} e^{2\pi ih \log n}$

$$= \int_1^N e^{2\pi ih \log t} dt + \frac{1}{2}(e^{2\pi ih \log N}+1) + \int_1^N \left(\{t\} - \frac{1}{2}\right) e^{2\pi ih \log t} 2\pi ih \frac{dt}{t}.$$

Now the second term $\ll 1$, the third term $\ll \int_1^N 1/t dt = \log N$ and

$$\text{the first term} = \int_1^N t^{2\pi ih} dt = \frac{1}{2\pi ih+1}(N^{2\pi ih+1}-1) = \Omega(N),$$

where $f = \Omega(g)$ and $f(x) \ll g(x)$ are meant by $f \neq o(g)$ and $|f/g| \leq$ constant as $x \to \infty$, respectively. So we have

$$\textstyle\sum_{n=1}^{N} e^{2\pi ih \log n} = \Omega(N) + O(\log N),$$

$$\textstyle\sum_{n=1, n=k^2}^{N} e^{2\pi ih \log n} = \sum_{k \leq \sqrt{N}} e^{4\pi ih \log k} = \Omega(\sqrt{N}) + O(\log N).$$

Next, in the same way as in (*), we have

(**) $D_N := \sum_{k \leq \sqrt{N}} e^{8\pi ih(\log k)^2}$

$$= \int_1^{\sqrt{N}} e^{8\pi ih(\log t)^2} dt + \frac{1}{2}(e^{2\pi ih(\log N)^2}+1)$$

$$+\int_1^{\sqrt{N}}\left(\{t\}-\frac{1}{2}\right)e^{2\pi ih(\log t)^2}16\pi ih\frac{\log t}{t}dt,$$

where the second term $O(1)$, and the third term $\ll \int_1^{\sqrt{N}}\log t/t\ dt \ll (\log N)^2$. To estimate the first term, we need the following lemma:

Lemma 1 (Van der Corput: *cf.* [2, Lemma 2.1]). *Suppose the real-valued function f has a monotone derivative f' on $[a, b]$ with $f'(x) \geq \lambda > 0$ or $f'(x) \leq -\lambda < 0$ for $x \in [a, b]$, and*

$$J = \int_a^b exp(2\pi if(x))dx.$$

Then

$$|J| < 1/\lambda.$$

We set $f(t) = 4h(\log t)^2$. Then by the lemma, we have in $(**)$

$$\text{the first term} = \int_1^{\sqrt{N}} e^{8\pi ih(\log t)^2}dt \ll \frac{N}{\log N}.$$

Hence $D_N \ll N/\log N$. Thus we have

$$\begin{aligned}\textstyle\sum_{n=1}^N e^{2\pi ih\log x_n} &= \textstyle\sum_{n=1}^N e^{2\pi ih\log n} - \sum_{n=1,n=k^2}^N e^{2\pi ih\log n} + \sum_{k\leq\sqrt{N}} e^{8\pi ih(\log k)^2}\\ &= \Omega(N)+O((\log N)^2).\end{aligned}$$

Therefore the sequence $(\log x_n)$ is not u.d. mod 1.

Finally we show that the sequence is $(M, 1/n)$-u.d. mod 1. We write

$$\begin{aligned}&\textstyle\sum_{1\leq n\leq N}\frac{1}{n}e^{2\pi ih\log x_n}\\ &= \textstyle\sum_{1\leq n\leq N}\frac{1}{n}e^{2\pi ih\log n} - \sum_{1\leq n\leq N, n=k^2}\frac{1}{n}e^{2\pi ih\log n} + \sum_{1\leq n\leq N, n=k^2}\frac{1}{n}e^{2\pi ih(\log n)^2}.\end{aligned}$$

First we estimate the following sum: By Euler summation formula, we have

$$\begin{aligned}&\textstyle\sum_{1\leq n\leq N}\frac{1}{n}e^{2\pi ih\log n}\\ &= \int_1^N\frac{1}{t}e^{2\pi ih\log t}dt+\frac{1}{2}(1+\frac{1}{N}e^{2\pi ih\log N})+\int_1^N\left(\{t\}-\frac{1}{2}\right)\frac{2\pi ih-1}{t^2}e^{2\pi ih\log t}dt,\end{aligned}$$

where the absolute value of the second term ≤ 1, and the third term $\ll \int_1^N t^{-2}dt \ll 1$. Besides,

$$\text{the first term} \ll \int_1^N t^{2\pi ih-1}dt = \frac{1}{2\pi ih}(N^{2\pi ih}-1) \ll 1.$$

So we have

$$\textstyle\sum_{1\leq n\leq N}\frac{1}{n}e^{2\pi ih\log n} = O(1).$$

Similarly,

$$\sum_{1\leq n\leq N, n=k^2}\frac{1}{n}e^{2\pi ih\log n} = \sum_{k\leq\sqrt{N}}\frac{1}{k^2}e^{4\pi ih\log k}$$

$$= \int_1^{\sqrt{N}}\frac{1}{t^2}e^{4\pi ih\log t}dt+\frac{1}{2}\left(1+\frac{1}{N}e^{2\pi ih\log N}\right)$$

$$+\int_1^{\sqrt{N}}\left(\{t\}-\frac{1}{2}\right)\frac{4\pi ih-2}{t^3}e^{4\pi ih\log t}dt.$$

Absolute value of this second term ≤ 1. Also,

$$\text{the third term} \ll \int_1^{\sqrt{N}}\frac{1}{t^3}dt = O(1), \text{ the first term} \ll \int_1^{\sqrt{N}}\frac{1}{t^2}dt = O(1).$$

Thus we have

$$\sum_{1\leq n\leq N, n=k^2}\frac{1}{n}e^{2\pi ih\log n} = O(1).$$

Finally we estimate

$$\sum_{1\leq n\leq N, n=k^2}\frac{1}{n}e^{2\pi ih(\log n)^2} = \sum_{1\leq n\leq\sqrt{N}}\frac{1}{k^2}e^{8\pi ih(\log k)^2}.$$

Since $D_N \ll N/\log N$, we easily find by partial summation that this sum is $O(1)$.

Consequently, we have

$$\sum_{n=1}^{N}\frac{1}{n}e^{2\pi ih\log x_n} = O(1) = o(\log N).$$

Therefore the sequence $(\log x_n)$ is $(M, 1/n)$-u.d. mod 1.

Remark. *Even if we assume that $(g(n))$ is a sequence such that $g(n) = o(n)$ and $g(n)/\log n \uparrow \infty$, $(g(n))$ is not always uniformly distributed mod 1 as shown by the following example;*

For $e^{\sqrt{k}} \leq n < e^{\sqrt{k+1}}$ $(k = 0, 1, \cdots)$, we put

$$g(n) = \left(\sqrt{k}+\frac{n}{8(k+1)^2}e^{-\sqrt{k+1}}\right)\log n.$$

Then $g(n)$ satisfies $g(n) = o(n)$ and $g(n)/\log n \uparrow \infty$.

For $e^{\sqrt{k}} \leq n < e^{\sqrt{k+1}}$ we have

$$\sqrt{k} \leq \frac{g(n)}{\log n} < \sqrt{k}+\frac{n}{8(k+1)^2}e^{-\sqrt{k+1}} < \sqrt{k}+\frac{1}{8(k+1)^2} < \frac{2k+1}{2\sqrt{k+1}}.$$

Hence $k \leq g(n) < k+1/2$. Therefore we obtain $\{g(n)\} \in [0, 1/2)$.

References

[1] P. Diaconis: The distribution of leading digits and uniform distribution mod 1, Ann. Prob. **5**

(1977), 72—81.
[2] L. KUIPERS and H. NIEDERREITER: Uniform Distribution of Sequences, John Wiley and Sons, New York, 1974.
[3] R. A. RAIMI: The first digit problem, Amer. Math. Monthly **83** (1976), 521—538.
[4] M. TSUJI: On the uniform distribution of numbers mod 1, J. Math. Soc. Japan **4** (1952), 313—322.
[5] R. E. WHITNEY: Initial digits for the sequence of primes, Amer. Math. Monthly **79** (1972), 150—152.

INFORMATION SCIENCE
SAKUYO JUNIOR COLLEGE
YAZU 2050-13, OKAYAMA 703, JAPAN

(*Received May 29, 1992*)

Added in Proof: *Current Address*;
DEPARTMENT OF MATHEMATICS
FACULTY OF EDUCATION
TOTTORI UNIVERSITY
TOTTORI 680, JAPAN