COMPOSITIO MATHEMATICA

EDWARD THORP
ROBERT WHITLEY

Poincaré's conjecture and the distribution of digits in tables

Compositio Mathematica, tome 23, nº 2 (1971), p. 233-250
<http://www.numdam.org/item?id=CM_1971__23_2_233_0>

COMPOSITIO MATHEMATICA, Vol. 23, Fasc. 2, 1971, pag. 233-250
Wolters-Noordhoff Publishing
Printed in the Netherlands

POINCARÉ'S CONJECTURE AND THE DISTRIBUTION OF DIGITS IN TABLES [1]

by

Edward Thorp and Robert Whitley

1. Introduction

Let $E_i(n)$ be the number of even digits and $O_i(n)$ the number of odd digits occurring in the ith column of the table of logarithms of the first n integers to the base 10. In 1899 Henri Poincaré [3, page 193] expressed the belief that $\lim_n E_i(n)/n$ and $\lim_n O_i(n)/n$ exist and equal $\frac{1}{2}$. Poincaré provided mathematical support for this conjecture in [2]. Franel [1] disproved Poincaré's conjecture by showing that these limits do not exist. Franel also showed that the sequence of arithmetic means of the digits in the ith column has no limit, and gave information on the set of accumulation points of these sequences.

We replace Franel's analytic approach by elementary counting arguments. We find the derivations simpler, the ideas transparent, and the methods more general. We generalize to an arbitrary base of logarithms, derive the exact set of accumulation points for these sequences, and show that the limit does not exist for many sequences not treated by Franel.

Franel showed that although the frequency of the digit l in the ith decimal place of $\log 1, \log 2, \cdots, \log n$ does not tend to a limit as n tends to infinity, the frequency of the digit l in the ith decimal place of 1^a, $2^a, \cdots, n^a, 0 < a < 1$, does tend to the expected value of $1/10$ as n tends to infinity. We determine exactly which monotone functions of slow growth have the property that the frequency of the digit l in the ith place tends to a limit as n tends to infinity. This extends some of the results on the logarithm table to a wide class of function tables.

A similar problem considered by Polya and Szegö [5] is whether the sequence of numbers $f(n)-[f(n)]$ is uniformly distributed in $[0, 1]$. We extend their results by showing that, under suitable hypotheses, this occurs if and only if the frequency of the occurrence of the digit $l = 0, 1, \cdots, 9$ in the ith decimal place of $f(1), \cdots, f(n)$ tends to $1/10$ as n tends to infinity *for each i*.

[1] Research partially sponsored by the Air Force Office of Scientific Research, Office of Aerospace Research, United States Air Force, under AFOSR Grant Nos. AF-AFOSR 1113–66 and AF-AFOSR 70-1870A.

2. Poincaré's conjecture: distribution of even and odd digits

We begin with a new disproof of Poincaré's conjecture. Let t be a real number, $t = \sum_{i=-\infty}^{\infty} t_i\, 10^{-i}$, where the t_i are uniquely defined by the stipulation that when there are two choices for the decimal expansion, i.e. $\lim t_i = 9$ or $\lim t_i = 0$, then the latter is chosen.

Consider $\{(\log_a k)_i : 1 \leqq k \leqq n\}$, the ith column counting right from the units place as 0, in the table of the first n logarithms to the base $a > 1$. Let $E_i(n)$ be the number of even digits in the set and let $O_i(n)$ be the number of odd digits.

THEOREM 1. *The set of accumulation points for the sequence* $\{E_i(n)/n\}$ *and also for the sequence* $\{O_i(n)/n\}$ *is the interval* $[1/(a^t+1),\, a^t/(a^t+1)]$, *where* $t = 10^{-i}$. *Thus the sequences have no limit.*

PROOF. Since $O_i(n)/n = 1-E_i(n)/n$, the results for $O_i(n)/n$ follow from those for $E_i(n)/n$. It suffices to consider the $i = 0$ or units column and establish the result for arbitrary $a > 1$. To then establish the theorem for the ith column, note that $\log_{(a^t)} x = y$ if and only if $\log_a x = ty = 10^{-i}y$. Thus the ith column of the $\log_a n$ table is the units column of the $\log_{(a^t)} n$ table and the theorem for the ith column is obtained from the units case by replacing a by a^t throughout. We shall use this device frequently without further comment.

Let $E(n)$ be the number of even digits in $\{(\log_a k)_0 : 1 \leqq k \leqq n\}$. Let $w_n = E(n)/n$, $n \leqq 1$. Then $(\log_a n)_0$ is even if and only if there is a j such that $a^{2j} \leqq n < a^{2j+1}$. In this case $E(n) = E(n-1)+1$, $n \geqq 2$, so

$$w_n = \frac{E(n-1)+1}{(n-1)+1} = \frac{w_{n-1}}{1+1/(n-1)} + \frac{1}{n} \geqq w_{n-1}$$

where the last inequality follows because $0 \leqq w_{n-1} \leqq 1$ and $1/(n-1)$ is defined and positive. The last equality also yields $w_n \leqq w_{n-1}+1/n$. Thus $0 \leqq w_n - w_{n-1} \leqq 1/n$, $n \geqq 2$, when $(\log_a n)_0$ is even.

When $(\log_a n)_0$ is odd, i.e. there is no j such that $a^{2j} \leqq n < a^{2j+1}$, then $0 \leqq w_{n-1} - w_n < 1/(n-1)$, $n \geqq 2$ because for these n,

$$E(n) = E(n-1), \quad n \geqq 2, \quad \text{whence } w_n = \frac{E(n)}{(n-1)+1} = \frac{w_{n-1}}{1+1/(n-1)}$$

so $w_n(1+1/(n-1)) = w_{n-1}$ and $w_{n-1} - w_n = w_n/(n-1) < 1/(n-1)$.

Therefore the integers $\langle a^{2j} \rangle$ are the local minima for w_n and the integers $\langle a^{2j+1} \rangle$ are the local maxima, where $\langle t \rangle$ is defined for real t and is the greatest integer less than t. The sequence is monotonic between consecutive relative extrema. Furthermore, $\lim(w_n - w_{n-1}) = 0$. From this it follows that the set of accumulation points of $\{w_n\}$ is the closed interval $[m, M]$, where

$$m = \underline{\lim}_n \{w_n : n = \langle a^{2j} \rangle : j \geqq 0\}$$

and

$$M = \overline{\lim}_n \{w_n : n = \langle a^{2j+1} \rangle : j \geqq 0\}.$$

We first calculate m, by estimating the w_n involved. We use the fact that the number $N[a, b)$ of integers in the interval $[a, b)$ satisfies $b-a-1 < N[a,b) < b-a+1$.

We have $E(\langle a^{2p} \rangle) = \sum_{j=0}^{p-1} N[a^{2j}, a^{2j+1})$. For $t \geqq 0$, let $d(t)$ be an arbitrary number in $\{u : |u| \leqq t\}$. Then $N[a, b) = b-a+d(1)$ and

$$E(\langle a^{2p} \rangle) = \sum_{j=0}^{p-1} (a^{2j+1} - a^{2j}) + d(p)$$

$$= (a-1)\frac{a^{2p}-1}{a^2-1} + d(p) = \frac{a^{2p}-1}{a+1} + d(p).$$

Thus

$$\frac{E(\langle a^{2p} \rangle)}{\langle a^{2p} \rangle} = \frac{a^{2p}-1+(a+1)d(p)}{(a+1)(a^{2p}+d(1))}$$

and $\lim_{p\to\infty} E(\langle a^{2p} \rangle)/\langle a^{2p} \rangle$ exists and equals $1/(a+1)$ so $m = 1/(a+1)$.

Since $E(\langle a^{2p-1} \rangle) = E(\langle a^{2p} \rangle)$, we have

$$\frac{E(\langle a^{2p-1} \rangle)}{\langle a^{2p-1} \rangle} = \frac{a^{2p}-1+(a+1)d(p)}{(a+1)(a^{2p-1}+d(1))}$$

and this tends to $a/(a+1)$ as p increases, whence $M = a/(a+1)$. This completes the proof.

Franel [1, page 293] obtains these accumulation points (limited like all his results to base 10) but he does not say these are the only accumulation points.

3. The distribution of single digits and of subsets of digits

Let $N(a, i, l, n)$ be the number of occurrences of the digit l in $\{(\log_a k)_i : 1 \leqq k \leqq n\}$.

THEOREM 2. *The set of accumulation points of the sequence* $N(a, i, l, n)/n = u_n$ *is the closed interval with endpoints* $(a^t-1)/(a^{10t}-1)$ *and* $(a^{10t} - a^{9t})/(a^{10t}-1)$, *where* $t = 10^{-i}$. *In particular* $\lim_n N(a, i, l, n)/n$ *does not exist for any* $a > 1, i = 0, \pm 1, +2, \cdots, l = 0, 1, \ldots, 9$.

PROOF OF THEOREM 2: Let $N(a, l, n)$ be the number of occurrences of the digit l in $\{(\log_a k)_0 : 1 \leqq k \leqq n\}$. Let $w_n = N(a, l, n)/n$. Observe that l occurs in the units place of $\log_a n$ iff $a^{10j+l} \leqq n < a^{10j+l+1}$ for some $j = 0, 1, 2, \cdots$.

As in the proof of Theorem 1, we observe that for such n, $0 \leqq w_n - w_{n-1} \leqq 1/n$ when $n \geqq 2$. If instead n does not satisfy $a^{10j+l} \leqq n < a^{10j+l+1}$ for any j, then we can show $0 \leqq w_{n-1} - w_n < 1/(n-1)$. The local minima of $\{w_n\}$ are at the integers $\langle a^{10j+l} \rangle$ and the local maxima are at $\langle a^{10j+l+1} \rangle$. The set of accumulation points is $[m, M]$, where

$$m = \varliminf_n \{w_n : n = \langle a^{10j+l} \rangle : j \geqq 0\}$$

and

$$M = \varlimsup_n \{w_n : n = \langle a^{10j+l+1} \rangle : j \geqq 0\}.$$

Next we find m. In the notation of Theorem 1,

$$N(a, l, \langle a^{10p+l} \rangle) = \sum_{j=0}^{p-1} N[a^{10j+l}, a^{10j+l+1}) = a^l(a-1) \sum_{j=0}^{p-1} a^{10j} + d(p)$$
$$= a^l(a-1)(a^{10p}-1)/(a^{10}-1) + d(p).$$

Thus

$$\frac{N(a, l, \langle a^{10p+l} \rangle)}{\langle a^{10p+l} \rangle} = \frac{a^l(a-1)(a^{10p}-1)/(a^{10}-1)+d(p)}{a^{10p+l}+d(1)}$$

and the limit as $p \to \infty$ is $(a-1)/(a^{10}-1) = m$. Similarly,

$$\frac{N(a, l, \langle a^{10p+l+1} \rangle)}{\langle a^{10p+l+1} \rangle} = \frac{a^l(a-1)(a^{10(p+1)}-1)/(a^{10}-1)+d(p+1)}{a^{10p+l+1}+d(1)}$$

and the limit is $(a-1)a^9/(a^{10}-1) = (a^{10}-a^9)/(a^{10}-1) = M$. This establishes the theorem for the units place and completes the proof.

COROLLARY 3. *If A_i is the set of accumulation points of $\{N(a, i, l, n)/n\}$ then $\bigcap_{i=1}^{\infty} A_i = 1/10$.*

Proceeding as in the proof of Theorem 1 readily yields the following result [5, page 73, problem 180].

THEOREM 4. *If f is any non-constant function defined on $\{l = 0, 1, \cdots, 9\}$, then*

$$\lim_{n \to \infty} \frac{1}{n} \sum_{k=1} f((\log_a k)_i)$$

does not exist for any $a > 1$ or any $i = 0, \pm 1, \cdots$. The set of accumulation points is $[m, M]$ where general formulas for m and M can be obtained.

REMARK. The non-existence of the limits in Theorems 1, 2, and 5 below follow by special choice of f. In the case of even integers, choose $f(j) = 1$, j even and $f(j) = 0$ otherwise. For single integers l, choose $f(l) = 1$ and $f(j) = 0$ otherwise. For the arithmetic mean, choose $f(j) = j, j = 0, \cdots, 9$.

4. The arithmetic means

Let $\sum(a, i, n)$ be the sum of the digits in the ith place of $\{\log_a k : 1 \leqq k \leqq n\}$. Let $M(a, i, n) = \sum(a, i, n)/n$ be the corresponding arithmetic mean.

THEOREM 5. *The set* A_i *of accumulation points of* $\{M(a, i, n) : 1 \leqq n\}$ *is the interval* $[m, M]$ *with*

$$m = \min_{0 \leqq l \leqq 4} \left\{ l + \frac{10a^{(9-l)t}}{a^{10t}-1} - \frac{1}{a^t - 1} \right\}$$

and

$$M = 9 - \frac{1}{a^t - 1} + \frac{10}{a^{10t}-1}.$$

We have $m < M$ *so* $\lim_n M(a, i, n)$ *does not exist for any* $a > 1$ *or any* $i = 0, \pm 1, \pm 2, \cdots$. *However,* $\bigcap_{i=1}^{\infty} A_i = 4.5$.

PROOF. Let $\sum(n)$ be the sum and let w_n be the mean for the units ($i = 0$) place. We first show that w_n is monotone on each of the intervals $[a^j, a^{j+1})$, $i \geqq 0$. Choose $n \in [a^{10k+l}, a^{10k+l+1})$, i.e., $\log_a n$ has l in the units place. Then

$$w_n = \frac{\sum(n)}{n} = \frac{\sum(n-1)+l}{n} = \frac{(n-1)w_{n-1}+l}{n}$$

so w_n is a convex combination of w_{n-1} and l. Thus if $w_{n-1} > l$, then $w_{n-1} > w_n > l$; if $w_{n-1} = l$, then $w_{n-1} = w_n = l$; and if $w_{n-1} < l$, then $w_{n-1} < w_n < l$, and w_n is monotone on each $[a^{10k+l}, a^{10k+l+1})$, tending towards l on each such subinterval. Thus the relative maxima and minima of $\{w_n\}$ are all among the numbers $w_{\langle a^j \rangle}, j \geqq 1$. If we write

$$w_n = w_{n-1} + \frac{l - w_{n-1}}{n},$$

we have $|w_n - w_{n-1}| \leqq 9/n$. We conclude that the set of accumulation points is the closed interval $[m, M]$ with

$$m = \varliminf_j w_{\langle a^j \rangle} \quad \text{and} \quad M = \varlimsup_j w_{\langle a^j \rangle}.$$

There is an infinite subsequence of the $\langle a^j\rangle$, say n_k, such that $\lim_k n_k = m$. An infinite number of the $\{n_k\}$ are in one of the ten sets $\{\langle a^{10k+l}\rangle : k \geqq 0\}$, $l = 0, \cdots, 9$. Thus $m = \underline{\lim}_k w_{\langle a^{10k+l}\rangle}$ for some fixed l. Hence

$$m = \min_{0\leqq l\leqq 9} \underline{\lim}_k w_{\langle a^{10k+l}\rangle}.$$

Similarly,

$$M = \max_{l\leqq 0\leqq 9} \overline{\lim}_k w_{\langle a^{10k+l}\rangle}.$$

Thus we wish next to determine the $w_{\langle a^{10k+l}\rangle}$.

For convenience we find $w_{\langle a^{10k+l+1}\rangle}$ instead. We have

$$\sum \langle a^{10k+l+1}\rangle = \sum_{p=0}^{l} p \sum_{j=0}^{k} N[a^{10j+p}, a^{10j+p+1}) + \sum_{p=l+1}^{9} p \sum_{j=0}^{k-1} N[a^{10j+p}, a^{10j+p+1}).$$

The first double sum equals

$$\sum_{j=0}^{k} a^{10j} \sum_{p=0}^{l} p(a^{p+1}-a^p) + \sum_{j=0}^{k} \sum_{p=0}^{l} pd(1)$$
$$= (\sum_{j=0}^{k} a^{10j}) \left(la^{l+1} - a\,\frac{a^l-1}{a-1}\right) + d\left(\frac{l(l+1)(k+1)}{2}\right)$$
$$= \left(\frac{a^{10(k+1)}-1}{a^{10}-1}\right)\left(\frac{la^{l+2}-(l+1)a^{l+1}+a}{a-1}\right) + d(l(l+1)(k+1)/2).$$

The second double sum can be found similarly, or by using the first double sum and substituting:

$$\left(\frac{a^{10k}-1}{a^{10}-1}\right)\left(\frac{9a^{11}-la^{l+2}-10a^{10}+(l+1)a^{l+1}}{a-1}\right) + d(45k).$$

Dividing each double sum by $\langle a^{10k+l+1}\rangle$, we note that each $\lim_{k\to\infty}$ exists and we find

$$\lim_{k\to\infty} w_{\langle a^{10k+l+1}\rangle} = \frac{la^{11}-(l+1)a^{10}+10a^{10-l}-10a^{9-l}-la+(l+1)}{(a-1)(a^{10}-1)}$$
$$= f_l(a) = l + \frac{10a^{9-l}}{a^{10}-1} - \frac{1}{a-1}.$$

Thus

$$\min_{0\leqq l\leqq 9} w_{\langle a^{10k+l+1}\rangle} = m \quad \text{and} \quad \max_{0\leqq l\leqq 9} w_{\langle a^{10k+l+1}\rangle} = M.$$

Substituting a^t for a and using L'Hospital's rule, we find $\lim_{t\downarrow 0} f_l(a^t) = 4.5$, which agrees with the result in [1] for $a = 10$. Thus $\bigcap_{i=1}^{\infty} A_i = 4.5$.

Now let $k(a) = la^{11}-(l+1)a^{10}+10a^{10-l}-10a^{9-l}-la+(l+1)$, the numerator of $f_l(a)$. We have $k_9(a) = a(9a^{10}-10a^9+1) = ak_0(a)$ which shows $f_9(a) > f_0(a)$. Letting

$$g_l(a) \equiv f_l(a)-f_{l-1}(a) = 1-\frac{10a^{9-l}(a-1)}{a^{10}-1} = \frac{(a-1)}{a^{10}-1}\left(\sum_{j=0}^{9} a^j-10a^{9-l}\right)$$

we see that $g_l(a)$ is monotone strictly increasing in l. Thus $f_l(a)$ either strictly increases as l increases or it first decreases, then increases. Hence $f_9(a)$ is the unique maximum and

$$M = f_9(a) = \frac{10a^{10}}{a^{10}-1}-\frac{a}{a-1},$$

which is almost exactly $8+8/9$ for $a = 10$.

If $l \geqq 5$ it follows that

$$\sum_{j=0}^{9} a^j > 10a^{9-l} \quad \text{so} \quad \min_{0\leqq l\leqq 9} f_l(a) = \min_{0\leqq l\leqq 4} f_l(a).$$

Franel [1, page 291] gives intervals of accumulation points for $l = 0, \cdots, 9$, for the case $a = 10$. If we let $\alpha = 1$ in his formula and replace his q_i by a^t, we get our $f_l(a)$. Thus the convex hull of his set of accumulation points is the complete set. Whether he has all the points does not seem easy to check. In any case, he does not seem to know, and his information (twenty endpoints) contains much that is superfluous (only six are needed).

Letting

$$g_u(a) = 1-\frac{10a^{9-u}(a-1)}{a^{10}-1},$$

let $u(a)$ be the value of u for fixed a such that $g_u(a) = 0$. We find

$$u(a) = 9-\frac{\log_{10}\left(\sum_{j=0}^{9} a^j\right)-1}{\log_{10} a}$$

and we see that $m = f_{\langle u(a)\rangle}(a)$. As $a \to \infty$, $\langle u(a)\rangle = 0$ and as $a\downarrow 1$, $\langle u(a)\rangle = 4$.

REMARK. The type of theorem and method of proof generalize to $\sum(a, i, n, S)/n$ where $\sum(a, i, n, S)$ is the sum of all digits in the set S which occur in the ith place of $\{\log_a k : 1 \leqq k \leqq n\}$.

5. Further generalizations

The preceding theorems and proofs do not crucially depend on the choice of 10 as the base for the number system. They will hold in their essentials when we use any integer $b \geqq 2$ as the base. Now suppose that in the previous work we replace a single ith column by a set of consecutive columns $i_1, i_1+1, \cdots, i_1+m, m \geqq 0$. By changing the logarithm base from a to $\alpha = a^{b^{-m-i_1}}$, these become columns $-m, -m+1, \cdots, -1, 0$. By changing the base of the number system from b to b^{m+1}, these $m+1$ columns become the units column. Therefore we have reduced a problem involving $m+1$ columns to a problem involving only the units column.

This suggests generalizing the problems previously considered to m arbitrary columns $i_1 < i_2 < \cdots < i_m$, $m \geqq 1$. By using the transformations $a \to a^{b^{-i_m}}$ and $b \to b^{i_m - i_1 + 1}$, we reduce to the previously analyzed units column case. We now use Theorem 4 and associated methods to complete the work. As an example of the results we have:

THEOREM 6. *Poincaré's conjecture generalized to m arbitrary columns of the table* $\{\log_a k : 1 \leqq k \leqq n\}$ *does not hold.*

In this connection we note the attempt by Fisher and Yates (reported in [4, page 484] to construct random numbers by selecting digits from the $15-19$ columns of a table of twenty place logarithms to the base 10.

A column of the logarithm table is not normal and in general a row is not: for some a and k, for instance, $\log_a k$ is rational. However considerations like those in [2, 3] suggest that for paths through the logarithm table that move down and sufficiently rapidly to the right, so that $\lim_{i,j} [a^{jt}, a^{(j+1)t}) = 0$, $t = 10^{-i}$, then the number associated with the path is normal. In particular we suggest the:

CONJECTURE: The diagonal sequence $\{(\log_a i)_i : i \geqq 1\}$ is normal for every base $a > 1$ and every integer number base $b \geqq 2$.

REMARK. Mark Finkelstein suggested the following to us. Let $w_n^{(l)}$ be the fraction of initial digits which are l in the table $\{\log_a k : 1 \leqq k \leqq n\}$. Then the set of accumulation points for $\{w_n^{(l)}\}$ is [0, 1].

PROOF. Let $N(l, n)/n = w_n^{(l)}$ and note that the initial digit of $\log_a k$ is l iff $a^{l \cdot 10^j} \leqq k < a^{(l+1)10^j}$ for some $j \geqq 0$. Then $w_n^{(l)}$ increases on intervals $[a^{l \cdot 10^j}, a^{(l+1)10^j})$ and decreases otherwise. Hence $w_n^{(l)}$ has relative maxima at $\{\langle a^{(l+1)10^j} \rangle : j \geqq 0\}$ and relative minima at $\{\langle a^{l \cdot 10^j} \rangle : j \geqq 0\}$. We have

$$w_{\langle a^{(l+1)10^j} \rangle} = \frac{N(l, \langle a^{(l+1)10^j} \rangle)}{\langle a^{(l+1)10^j} \rangle} > \frac{a^{(l+1)10^j} - a^{l \cdot 10^j} - 1}{a^{(l+1)10^j}}$$

and it follows that $\lim_{j \to \infty} w_{\langle a^{(l+1)10^j} \rangle} = 1$.

Similarly,

$$w_{\langle a^l \cdot 10^j\rangle} = \frac{N(l, \langle a^{l \cdot 10^j}\rangle)}{\langle a^{l \cdot 10^j}\rangle} = \frac{N(l, \langle a^{l \cdot 10^{j-1}}\rangle)}{\langle a^{l \cdot 10^j}\rangle} \leqq \frac{a^{l \cdot 10^{j-1}}}{a^{l \cdot 10^j}-1}$$

and

$$\lim_{j \to \infty} w_{\langle a^l \cdot 10^j\rangle} = 0.$$

That every point in [0, 1] is an accumulation point for $\{w_n^{(l)}\}$, $1 \leqq l \leqq 9$, now follows.

6. Extensions to monotone slowly increasing functions

THEOREM 7. *Let f be a strictly increasing function mapping* $[0, \infty)$ *onto itself with* $\lim_{x\to\infty} f(x)/x = 0$. *Let $N[f, i, l, n]$ be the number of occurrences of the digit l in the ith decimal place (counting to the right with the units place as zero) of the numbers $\{f(k) : 1 \leqq k \leqq n\}$. Then the cluster points of the sequences*

$$\left\{\frac{N[f, i, l, n]}{n}\right\}$$

are exactly the points in the interval

$$\left[\varliminf \sum_{j=0}^{n} \frac{b_j - a_j}{a_{n+1}},\ \varlimsup \sum_{j=0}^{n+1} \frac{b_j - a_j}{b_{n+1}}\right],$$

where

$$a_j = f^{-1}\left(\frac{10j+l}{10^i}\right) \quad \text{and} \quad b_j = f^{-1}\left(\frac{10j+l+1}{10^i}\right).$$

Consequently $\lim_n N[f, i, l, n]/n$ *exists if*

(a) $\lim (a_n/b_n) = 1$ *and*

(b) *either*

$$\lim_n \sum_{j=0}^{n} \frac{b_j - a_j}{a_{n+1}} \text{ exists or } \lim_n \sum_{j=0}^{n+1} \frac{b_j - a_j}{b_{n+1}} \text{ exists.}$$

Conversely, if $\lim_n N[f, i, l, n]/n$ *exists, then both the limits in* (b) *exist (and are equal). If this limit* $\lim_n N[f, i, l, n]/n$ *is not* 1, (a) *holds as well; if it is* 1, (a) *may or may not hold.*

PROOF. For any integer i, the ith decimal place of $f(x)$ is the units decimal place of $10^i f(x)$. Suppose that Theorem 7 has been established for the units decimal place, i.e. the case $i = 0$. Applying this result to the function $(10^i f)^{-1}(x) = f^{-1}(x/10^i)$, establishes the results of Theorem 7 for the ith decimal place of $f(x)$. Thus we only need to prove the theorem for the special case $i = 0$ and that is what we will do.

The digit l occurs in the units place of $f(n)$ if and only if $f(n) = 10j+l+y$, $0 \leqq y < 1$, where j is a non-negative integer, i.e. if and only if

$$(1) \qquad f^{-1}(10j+l) \leqq n < f^{-1}(10j+l+1), \qquad j = 0, 1, \cdots,$$

For simplicity we set $a_j = f^{-1}(10j+l)$ and $b_j = f^{-1}(10j+l+1)$.

Let $w_n = N[f, 0, l, n]/n$. For n satisfying (1) we see that

$$w_n = \frac{N[f, 0, l, n-1]+1}{n} = \frac{w_{n-1}}{1+1/(n-1)} + \frac{1}{n}$$
$$= \left(1 - \frac{1}{n}\right) w_{n-1} + \frac{1}{n} = w_{n-1} + \frac{1-w_{n-1}}{n}$$

whence $w_n \geqq w_{n-1}$ and $w_n \leqq w_{n-1}+1/n$. For n not satisfying (1),

$$w_n = \frac{N[f, 0, l, n-1]+0}{n} = \frac{w_{n-1}}{1+1/(n-1)}$$

so

$$w_{n-1} = w_n + \frac{w_n}{n-1},$$

whence $w_n \leqq w_{n-1}$ and

$$w_{n-1} \leqq w_n + \frac{1}{n-1}.$$

The sequence $\{w_n\}$ is therefore monotonic between relative extrema with the integers at which local minima (maxima) occur being of the form $\langle a_n \rangle (\langle b_n \rangle)$, where $\langle t \rangle$ denotes the greatest integer less than t. Furthermore, since $\lim(w_n - w_{n-1}) = 0)$, the set of accumulation points of $\{w_n\}$ is the entire interval $[m, M]$ where

$$m = \underline{\lim}\, [w_n : n = 1, 2, \cdots] = \underline{\lim}\, [w_n : n = \langle a_j \rangle, j = 1, 2, \cdots] \text{ and}$$
$$M = \overline{\lim}\, [w_n : n = 1, 2, \cdots] = \overline{\lim}\, [w_n : n = \langle b_j \rangle, j = 1, 2, \cdots].$$

We estimate,

$$(2) \qquad \sum_{j=0}^{n-1} (b_j - a_j - 1) \leqq N[f, 0, l, \langle a_n \rangle] \leqq \sum_{j=0}^{n-1} (b_j - a_j + 1).$$

Divide (2) by $\langle a_n \rangle$. Since $f(n)/n$ tends to zero as n tends to infinity, by hypothesis, it follows that $n/f^{-1}(n)$ and thus $n/\langle a_n \rangle$ also tends to zero as n tends to infinity. Hence both sides of the inequality are close to

$$m_{n-1} = \sum_{j=0}^{n-1} \frac{b_j - a_j}{a_n}$$

for large n. Thus $\underline{\lim}\, w_n = \underline{\lim}\, m_n$.

Similarly

$$\sum_{j=0}^{n} (b_j - a_j - 1) \leqq N[f, 0, l, \langle b_n \rangle] \leqq \sum_{j=0}^{n} (b_j - a_j + 1). \tag{3}$$

Dividing (3) by $\langle b_n \rangle$ we see that for large n, $w_{\langle b_n \rangle}$ is close to

$$M_{n-1} = \sum_{j=0}^{n} \frac{b_j - a_j}{b_n}.$$

Hence the interval of accumulation points of $\{w_n\}$ is as described in the theorem.

We have seen that the sequence $\{w_n\}$ clusters in the interval $[\underline{\lim}\, m_n, \overline{\lim}\, M_n]$, where

$$m_n = \sum_{j=0}^{n} \frac{b_j - a_j}{a_{n+1}} \quad \text{and} \quad M_n = \sum_{j=0}^{n+1} \frac{b_j - a_j}{b_{n+1}}.$$

We have

$$M_n - m_n = \left(1 - \frac{a_{n+1}}{b_{n+1}}\right)(1 - m_n) = \left(1 - \frac{b_{n+1}}{a_{n+1}}\right)(M_n - 1).$$

Thus if (a) of the theorem holds, $M_n - m_n \to 0$ so if (b) also holds, $\lim_n N[f, 0, l, n]/n = \lim M_n = \lim m_n$.

Conversely, suppose $\lim_n N[f, 0, l, n]/n$ exists. Then $\lim_n M_n = \lim_n m_n$ so $\lim_n (M_n - m_n) = 0$, whence

$$\left(1 - \frac{a_{n+1}}{b_{n+1}}\right)(1 - m_n) \to 0.$$

If $\lim_n m_n < 1$, then

$$\frac{a_{n+1}}{b_{n+1}} \to 1,$$

establishing (a).

If $\lim N[f, 0, l, n]/n = 1$, then (a) may or may not hold. In fact, a_n/b_n may have any limit between 0 and 1 inclusive or may have no limit. For example, let $f(x)$ be defined by choosing $g(x) = f^{-1}(x)$ to have slope 1 if $x \notin \bigcup_{j=0}^{\infty} [10j, 10j+1)$ and $g(x)$ to have slope 10^j if $x \in [10j, 10j+1), j = 0, 1, \cdots$. We specify g is continuous and $g(0) = 0$, which completely determines g, hence f. It follows readily that $\lim_{n \to \infty} f(x)/x = 0$, that $\lim_n N[f, 0, 0, n]/n = 1$, and that $a_n/b_n \to 1/10$. By replacing the slopes 10^j by suitable constants c_j, the sequence a_n/b_n can be made to have any limit between 0 and 1 inclusive, or no limit at all.

REMARK. Theorem 7 and the proof generalize readily to f such that for some $c > 0$ and $d > 0$, f is a strictly increasing mapping of $[c, \infty)$ onto $[d, \infty)$ and $\lim_x f(x)/x = 0$, the domain of f^{-1} being $[d, \infty)$. Theorem 8 and Theorem 9 below similarly generalize.

Neither condition (a) nor condition (b) is alone enough to establish the existence of the limits $\lim_n N[f, i, l, n]/n$. To see that (b) will not suffice, let $f(x) = \log_a(1+x)$, $a > 1$. For the 0th decimal place,

$$\lim \sum_{j=0}^{n} \frac{b_j - a_j}{a_{n+1}}$$

exists and equals $(a-1)/(a^{10}-1)$ but $\lim a_n/b_n = 1/a$. So, although (b) is satisfied, the limits $\lim_n N[f, 0, l, n]/n$ do not exist. (Note that for no l can we have $\lim N[f, 0, l, n]/n = 1$, for then for all other l the limits would be 0 and thus condition (a) would hold.) To see that condition (a) will not suffice, define f by constructing its inverse g as follows. Let

$$g(10n+l) = n^2 + c(l, n)[(n+1)^2 - n^2],\ l = 0, 1, \cdots, 9,$$

where $0 < c(0, n) < c(1, n) < \cdots < c(9, n) < 1$, and g linear for all other values in $[0, \infty)$. Then g is a monotone function with

$$\lim \frac{g(10n+l)}{10n+l} = \infty \quad \text{and} \quad \lim \frac{g(10n+l+1)}{g(10n+l)} = 1$$

so that condition (a) is satisfied. However the sum

$$m_{n-1} = \sum_{j=0}^{n-1} \frac{b_j - a_j}{a_n} = \sum_{j=0}^{n-1} \frac{g(10j+l+1) - g(10j+l)}{g(10n)}$$

is the ratio of the growth of g over intervals of the form $[10j+l, 10j+l+1]$, $j = 0, 1, \cdots, n-1$, to the total growth of g over $[0, 10n]$. (This is asymptotic to the frequency of the occurrence of the digit l in the units place of $f(k)$ for those integers k with $0 \leqq f(k) \leqq 10n$.) So for any given l_0 the sum

$$\sum_{j=0}^{n-1} \frac{b_j - a_j}{a_n}$$

can be made to converge or not converge by appropriately choosing the constants $c(j, l_0)$.

In light of the difficulty of checking condition (b) of Theorem 7, it is useful to know that condition (a) suffices for many functions. We show this below in Theorem 8.

THEOREM 8. *Let f be a differentiable function mapping* $[0, \infty)$ *onto itself with f' strictly decreasing to zero. Then*

$$\overline{\lim_n}\, N[f, i, l, n]/n \geqq 1/10 \geqq \underline{\lim_n}\, N[f, i, l, n]/n.$$

The limit $\lim_n N[f, i, l, n]/n$ *exists if and only if* (a) $\lim a_n/b_n = 1$; *in this case the limit is* 1/10. *The limits* $\lim_n N[f, i, l, n]/n$ *all exist* (*and are* 1/10) *for* $l = 0, 1, \cdots, 9$ *and* $i = 0, \pm 1, \pm 2, \cdots$ *if and only if*

$$\text{(a}'\text{)} \lim_{x\to\infty} \frac{f^{-1}(x+\varepsilon)}{f^{-1}(x)} = 1 \qquad \text{for each } \varepsilon > 0.$$

Proof. For any given integer k and $n > k$,

$$\begin{aligned} \frac{f^{-1}(n)}{n} &= \frac{1}{n}\sum_{j=1}^{n-1}(f^{-1}(j+1)-f^{-1}(j)) + \frac{f^{-1}(1)}{n} \\ &\geqq \frac{1}{n}\sum_{j=k}^{n-1}(f^{-1}(j+1)-f^{-1}(j)) \geqq \frac{1}{n}\sum_{j=k}^{n-1}\frac{1}{f'(j)} \\ &\geqq \frac{n-k-1}{n}\,\frac{1}{f'(k)}. \quad \text{Thus } \frac{f^{-1}(n)}{n} \to \infty \text{ as } n \to \infty. \end{aligned}$$

Thus the hypotheses of Theorem 7 are satisfied by the function f.

As in Theorem 7 we will only consider the units place of f since the result for the ith decimal place follows by considering $10^i f$.

Let l be a given integer with $0 \leqq l \leqq 9$. Recall the notation of Theorem 7 : $g = f^{-1}$, $a_n = g(10n+l)$, $b_n = g(10n+l+1)$,

$$m_n = \sum_{j=0}^{n} \frac{b_j - a_j}{a_{n+1}} \quad \text{and} \quad M_n = \sum_{j=0}^{n+1} \frac{b_j - a_j}{b_{n+1}}.$$

Using the fact that $g(n+1)-g(n)$ is an increasing function of n we compute

$$\begin{aligned} b_j - a_j &> 1/10 \sum_{p=1}^{10} g(10(j-1)+l+p+1) - g(10(j-1)+l+p) \\ &= 1/10(b_j - b_{j-1}) \end{aligned}$$

and also

$$b_j - a_j < 1/10 \sum_{p=0}^{9} g(10j+l+p+1) - g(10j+l+p) = (a_{j+1} - a_j)/10.$$

Thus

$$(4) \qquad (b_j - b_{j-1})/10 \leqq b_j - a_j \leqq (a_{j+1} - a_j)/10$$

and

$$(5) \qquad (b_n - b_0)/10 \leqq \sum_{j=0}^{n} b_j - a_j \leqq (a_{n+1} - a_0)/10.$$

Dividing the right hand side of (5) by a_{n+1} we see that $\underline{\lim}\ m_n \leqq 1/10$ and dividing the left hand side of (5) by b_n we see that $\overline{\lim}\ M_n \geqq 1/10$. Hence $\overline{\lim}_n N[f, i, l, n]/n \geqq 1/10 \geqq \underline{\lim}_n N[f, i, l, n]/n$.

Suppose that $\lim a_n/b_n = 1$. Then from the left hand side of (4), $\lim b_{n-1}/a_n = 1$. It now follows from the left hand side of (5) that $\underline{\lim}\ m_n \geqq 1/10$ and, from the right hand side of (5), that $\overline{\lim}\ M_n \leqq 1/10$. Then by Theorem 7,

$$\lim_n N[f, 0, l, n]/n = 1/10.$$

On the other hand suppose that the limit $\lim_n N[f, 0, l, n]/n$ exists. As we saw above, $\lim_n N[f, 0, l, n]/n$ must be 1/10 and $\lim a_n/b_n = 1$ by Theorem 7.

The rest of the theorem follows directly.

REMARK. Theorems 7, 8 and 9 are related to a theorem given by Koksma [9, pages 88–89, Satz 3]. In Theorem 7, for instance, the consequence of (a) can be deduced from Koksma's Satz 3 for $l = 0$, when additional hypotheses are placed on f to satisfy the hypotheses of Satz 3. That part of Theorem 8 which asserts that if (a') holds then the limits all exist and are 1/10 can be deduced from Satz 3 if we add the hypothesis that $g'(x)/g(x)$ tends to zero *monotonically* as $x \to \infty$. In the absence of such an additional hypothesis, Theorem 8 and Satz 3 each apply to functions not covered by the other. The relation between Theorem 9 and Satz 3 is similar.

We see from Theorem 8 that if the frequency of all digits $0, \cdots, 9$ in the ith place of $f(1) \cdots f(n)$ exists for a function f satisfying the hypothesis of the theorem, then the frequency is 1/10 and it also is 1/10 for every digit in every decimal place.

Franel's results that the limits $\lim N[f, i, l, n]/n$ do not exist for $f(x) = \log_{10} x$ and are 1/10 for $f(x) = x^a$, $0 < a < 1$, follow immediately from Theorem 8 (as extended by the Remark following Theorem 7).

7. The distribution of $f(n)$ modulo one

In [5, problems 174 and 182] Polya and Szegö discuss the problem of when, for a certain class of functions f, the sequence $x_n = f(n) - [f(n)]$ is uniformly distributed in the interval [0, 1]. A sequence $\{x_n\}$ is *uniformly distributed* in [0, 1] if for each Riemann integrable function f on [0, 1],

$$\frac{f(x_1) + \cdots + f(x_n)}{n} \to \int_0^1 f(t)dt.$$

We show that their results, which consist of one condition which is necessary and another which is sufficient, follow from the simple necessary and sufficient condition of Theorem 8.

THEOREM 9. *Let f be a differentiable function mapping* $[0, \infty)$ *onto itself with f' strictly decreasing to zero. Then the sequence $x_n = f(n) - [f(n)]$ is uniformly distributed in* $[0, 1]$ *if and only if the frequency of the occurrence of the digit l in each decimal place of $f(1), \cdots, f(n)$ tends to* $1/10$ *as n tends to infinity, i.e. iff* $\lim_n N[f, i, l, n]/n = 1/10$ *for all i and* $l = 0, 1, \cdots, 9$.

PROOF. Suppose that $x_n = f(n) - [f(n)]$ is uniformly distributed in $[0, 1]$. The ith digit of $f(n)$ is l iff x_n belongs to

$$S = \{10^{-i}[10j+l, 10j+l+1) : j = 0, 1, \cdots, 10^{i-1}-1\}.$$

Letting C_S be the characteristic function of S,

$$N[f, i, l, n]/n = \frac{1}{n}\sum_{j=1}^{n} C_S(x_j) \to \int_0^1 C_S(t)dt = 1/10.$$

Conversely suppose that $\lim N[f, i, l, n]/n = 1/10$ for all i and $l = 0, 1, 2, \cdots, 9$. We need to show that for $x_n = f(n) - [f(n)]$

$$\frac{h(x_1) + \cdots + h(x_n)}{n} \to \int_0^1 h(t)dt$$

for all Riemann integrable functions h. For each $\varepsilon > 0$ there are functions h_1 and h_2, both linear combinations of characteristic functions of intervals of the form $[a, b)$, a and b rational, with $h_1 \leqq h \leqq h_2$ and

$$\varepsilon + \int h_1 \geqq \int h \geqq -\varepsilon + \int h_2 .$$

It follows that it suffices to establish (7) for linear combinations of characteristic functions of intervals with rational endpoints. By linearity it is enough to show that (7) holds for $C_{[a, b]}$, a and b rational and another use of linearity shows that it will do to show that (7) holds when

$$a = (10j+l)10^{-i} \quad \text{and} \quad b = (10j+l+1)10^{-i},$$

where $l = 0, 1, \cdots, 9$, i is a positive integer and j is a non-negative integer with $10j+l+1 \leqq 10^i$, i.e. $j \leqq 10^{i-1}-1$. Then x_n lies in $[(10j+l)/10^i, (10j+l+1)/10^i)$ iff $(10j+l)10^{-i} \leqq x_n = f(n) - [f(n)] < (10j+l+1)10^{-i}$ iff $10j+l+p(10^i) \leqq 10^i f(n) < 10j+l+1+p(10^i)$, where $p = [f(n)]$. For $0 \leqq j \leqq 10^{i-1}-1$, let $I_j(n)$ be the number of integers in the set

$$A_j(n) = \left\{m : m \leqq n, f(m) \in \bigcup_{p=0}^{\infty} [p+(10j+l)10^{-i}, p+(10j+l+1)10^{-i})\right\},$$

i.e. the number of integers $m \leqq n$ with

$$p10^i + 10j + l \leqq 10^i f(m) < p10^i + 10j + l + 1$$

for some $p = 0, 1, \cdots$. An integer $k \leqq n$ is one for which the ith digit of $f(k)$ is l iff m belongs to $\bigcup_{j=0}^{10^{i-1}-1} A_j(n)$; thus

$$(8) \qquad N[f, i, l, n] = I_0(n) + I_1(n) + \cdots + I_{10^{i-1}-1}(n)$$

We want to show that for the function $h = C_{[a,b]}$ we have

$$\frac{h(x_1) + \cdots + h(x_n)}{n} \to 10^{-i}$$

and we see that

$$\frac{h(x_1) + \cdots + h(x_n)}{n} = \frac{I_j(n)}{n}.$$

Thus, using (8), we must show that if $N[f, i, l, n]/n \to 1/10$, then each of the 10^{i-1} pieces $I_0(n)/n, I_1(n)/n, \cdots, I_{10^{i-1}-1}(n)/n$ contributes an equal share to the total frequency.

Let $g = f^{-1}$ and let $N[a, b)$ be the number of integers in $[a, b)$. Then

$$(9) \quad I_j(n)/n = \frac{1}{n} \sum_{p=0}^{p_0} N[g(p+(10j+l)10^{-i}), g(p+(10j+l+1)10^{-i})) + \Delta/n,$$

where $p_0 = \max\{p : g(p+(10j+l+1)10^{-i}) \leqq n\}$, i.e. $p_0 = [f(n)]$. In general Δ is not 0. However, we are interested in those n which are relative maxima or relative minima and in those cases Δ is 0.

Now $I_j(n)/n$ has relative maxima only at n of the form $\langle g(m+10j+l+1)10^{-i}) \rangle$ and for such n the sum (9) becomes

$$\sum_{p=0}^{m} \frac{N[g(p+(10j+l)10^{-i}), g(p+(10j+l+1)10^{-i}))}{\langle g(m+(10j+l+1)10^{-i}) \rangle}.$$

Multiplying this by 10^i and estimating as in Theorem 8 yields

$$(10) \quad \begin{aligned} &\limsup_n 10^i I_j(n)/n \\ &= \limsup_m \sum_{p=0}^{m} 10^i \frac{(g(p+(10j+l+1)10^{-i} - g(p+(10j+l)10^{-i}))}{g(m+(10j+l+1)10^{-i})}. \end{aligned}$$

Since g' is monotone increasing,

$$\begin{aligned} g(p+(10j+l+1)10^{-i}) - g(p+(10j+l)10^{-i}) \\ \leqq g(p+(10j+l+1+k)60^{-i}) \\ -g(p+(10j+l+k)10^{-i}) \quad \text{for} \quad k \geqq 0. \end{aligned}$$

We use this fact as we did in Theorem 8 to see that (10) is bounded above by:

$$\limsup_m \sum_{p=0}^{m} \sum_{k=0}^{10^{i-1}} \frac{g(p+(10j+l+1+k)10^{-i})-g(p+(10j+l+k)10^{-i})}{g(m+(10j+l+1)10^{-i})}$$
$$= \limsup_m \frac{g(m+1+(10j+l)10^{-i})-g((10j+l)10^{-i})}{g(m+(10j+l+1)10^{-i})} = 1,$$

since

$$\lim_{x\to\infty} \frac{g(x+\varepsilon)}{g(x)} = 1 \qquad \text{for each } \varepsilon > 0,$$

from Theorem 8.

Thus we have shown that $\lim_n \sup I_j(n)/n \leqq 10^{-i}$ for $j = 0, 1, \cdots, 10^{i-1}-1$. This, together with

$$I_0(n)/n+I_1(n)/n+ \cdots +I_{10^{i-1}-1}(n)/n = N[f, i, l, n]/n$$

and

$$\lim N[f, i, l, n]/n = 1/10,$$

show that $\lim I_j(n)/n = 10^{-i}$ for $j = 0, 1, \cdots, 10^{i-1}-1$. This completes the proof of the theorem.

COROLLARY 10. [5, *problems* 174 *and* 182] *Let f have the following properties for $t \geqq 1$:*

1) *f is (continuously) differentiable,*
2) *f is monotone increasing to ∞ as $t \to \infty$,*
3) *f' is monotone decreasing to* 0 *as $t \to \infty$.*

Then

4) (a) *If $tf'(t) \to \infty$ as $t \to \infty$, the sequence $x_n = f(n)-[f(n)]$ is uniformly distributed in* [0, 1].
   (b) *If $tf'(t) \to 0$ as $t \to \infty$, the sequence $x_n = f(n)-[f(n)]$ is not uniformly distributed in* [0, 1].

PROOF. An immediate consequence of Theorem 8 and Theorem 9 as generalized in the remark following the proof of Theorem 7 since, as noted in [5], for each $\varepsilon > 0$ in case (a)

$$\lim_{x\to\infty} \frac{f^{-1}(x+\varepsilon)}{f^{-1}(x)} = 1$$

and in case (b)

$$\lim_{x\to\infty} \frac{f^{-1}(x-\varepsilon)}{f^{-1}(x)} = 0.$$

We thank the referee for his many helpful suggestions.

REFERENCES

J. Franel

[1] À propos des tables des logarithmes. Vierteljschr. Naturforsch. Ges. Zürich, 1917, pp. 286–295.

H. Poincaré

[2] Revue générale des Sciences, April 15, 1899, pp. 262–269.

H. Poincaré

[3] Science and Hypothesis. Unabridged republication of the first English translation, Dover, 1952, page 193.

R. G. Stoneham

[4] A study of 60,000 digits of the transcendental '*e*'. Amer. Math. Mo. 72, May 1965, pp. 483–500. See pp. 484–485.

Polya and Szegö

[5] Aufgaben und Lehrsätze aus der Analysis, Springer, 1925, vol. I, pp. 72–74, problems 174–184 and pp. 238–241.

Sir R. A. Fisher and Frank Yates

[6] Statistical tables for biological, agricultural and medical research, Edinburgh, 1938.

B. Jansson

[7] Random Number Generators, Stockholm, 1966.

[8] Compositio Mathematica, 1964, vol. 16, pp. 1–203.

J. F. Koksma

[9] Diophantische Approximationen, Ergebnisse der Math. IV, 4, Springer, Berlin, 1936.

(Oblatum 13–X–70)

Mathematics Department
University of California
Irvine, California