ON DIGITAL DISTRIBUTION IN SOME INTEGER SEQUENCES

B. D. CRAVEN

(Received 7 September 1964, revised 15 January 1965)

1. Introduction

Although the harmonic series

$$1+\tfrac{1}{2}+\tfrac{1}{3}+\tfrac{1}{4}+\cdots$$

diverges, there is a sense in which it "nearly converges". Let N denote the set of all positive integers, and S a subset of N. Then there are various sequences S for which

$$T=\sum_{n\in S}\frac{1}{n} \tag{1}$$

converges, but for which the "omitted sequence" $N-S$ is, in an intuitive sense, sparse, compared with N. For example, Apostol [1] (page 384) quotes, without proof, the case where S is the set of all positive integers whose decimal representation does not involve the digit zero (e.g. $7\in S$ but $101\notin S$); then (1) converges, with $T<90$.

It is shown in this paper that Apostol's example is a special case of a general theorem on a class of sequences S for which T converges. From this it follows that certain integer sequences — in particular the sequence of prime numbers — include, for each integer d, a term whose representation to a given base contains any given digit at least d times. For example, there exists a prime p whose decimal representation contains at least 100 zeros. Although the existence proof for p is not constructive, an asymptotic bound for p is obtained, using the prime number theorem of Hadamard and de la Vallée Poussin.

2. Harmonically convergent sequences

An increasing sequence of positive integers $\{n_1, n_2, \cdots\}$, for which the series of reciprocals

$$\frac{1}{n_1}+\frac{1}{n_2}+\frac{1}{n_3}+\cdots \tag{2}$$

converges, will be called "harmonically convergent". The sum of the series (2) will then be called the "harmonic sum" of the sequence. A sequence for which the sum of reciprocals (2) diverges will be called "harmonically divergent".

A large class of harmonically convergent sequences is characterised by the following theorem.

THEOREM 1. *For integers* $b = 2, 3, 4, \cdots$, $d = 1, 2, 3, \cdots$, $t = 0, 1, 2, \cdots, b-1$, *let* $S(b, d, t)$ *denote the increasing sequence of all positive integers whose representation to base b involves the digit t at most* $(d-1)$ *times. Then* $S(b, d, t)$ *is harmonically convergent, and its harmonic sum is (strictly) less than* $b^d(1+d \log b)$.

PROOF. For each positive integer r, denote by $D(r)$ the set of b^d consecutive integers whose least member is $b^d r$.

If $r \notin S(b, d, t)$, then $D(r)$ contains no members of $S(b, d, t)$.

If $r \in S(b, d, t)$, then $D(r)$ contains at most (b^d-1) members of $S(b, d, t)$, since one member of $D(r)$ has the digit t in each of its last d positions. Let $C(r)$ denote the sum of the reciprocals of these at most (b^d-1) integers. Then, for $r \in S(b, d, t)$, $C(r)$ is (strictly) less than $(b^d-1)(b^d r)^{-1}$. In particular, if r_0 is the least member of $S(b, d, t)$, then

$$\Delta \equiv (b^d-1)(b^d r_0)^{-1} - C(r_0) > 0.$$

Denote also

$$C(0) = \tfrac{1}{1}+\tfrac{1}{2}+\tfrac{1}{3}+ \cdots + \frac{1}{b^d-1}.$$

Let T_q denote the sum of the reciprocals of the first q members of $S(b, d, t)$. Let $\sum_r'$ denote summation only over values of r which belong to $S(b, d, t)$. Then, for all $q > (r_0+1)b^d$,

$$T_q < C(0) + \sum_{r=1}^{q}{}' C(r);$$

since the right side includes all terms of T_q, plus additional positive terms. Therefore

$$T_q < \{1+\log (b^d-1)\} + \{(b^d-1)b^{-d}T_q - \Delta\},$$

so that

$$T_q < b^d(1+d \log b - \Delta).$$

Hence $S(b, d, t)$ is harmonically convergent, and its harmonic sum $T(b, d, t)$ satisfies the inequality

$$T(b, d, t) < b^d(1+d \log b).$$

https://doi.org/10.1017/S1446788700027750 Published online by Cambridge University Press

3. Harmonically divergent sequences

Let $W = \{n_1, n_2, \cdots\}$ denote any harmonically divergent sequence. For $j = 1, 2, 3, \cdots$, define the functions

$$G(j) = \frac{1}{n_1} + \frac{1}{n_2} + \cdots + \frac{1}{n_j} \tag{3}$$

$$H(j) = \frac{1}{n_1} + \frac{1}{n_2} + \cdots + \frac{1}{n_r} \tag{4}$$

where n_r is the largest member of W not exceeding j.

Since $G(j)$ and $H(j)$ are increasing functions, their inverses $G^{-1}(x)$ and $H^{-1}(x)$ are defined for values of x which fall in the ranges of G and H respectively. For values of x which do not, let x' denote the largest number, not exceeding x, which lies in the range of G; then define $G^{-1}(x) = G^{-1}(x')$. Similarly define $H^{-1}(x)$.

THEOREM 2. *Let W be a harmonically divergent sequence. Let $b \geqq 2$ and $d \geqq 1$ be integers. Then for every choice of the integers $b' = 2, 3, \cdots, b$* and *$t = 0, 1, 2, \cdots, b'-1$, there is a member $n_i = n_i(b, d, t)$ of W whose representation to base b' contains the digit t at least d times, and such that*

$$n_i \leqq H^{-1}(T(b, d, t)) \tag{5}$$

$$i \leqq G^{-1}(T(b, d, t)). \tag{6}$$

PROOF. Let b, d, t be given. Then by Theorem 1, W is not contained in the set $S(b, d, t)$, so W includes an integer n whose representation to base b contains the digit t at least d times. Again by Theorem 1, a partial sum (3) which exceeds $T(b, d, t)$ must contain such a number n, so (5) is proved, for $b' = b$. A similar proof applies to (6).

For $b' < b$, (5) and (6) thus hold, if b' replaces b. Now if t and d are given, $T(b, d, t)$ increases as b increases, because an increase of b replaces each integer in the "omitted sequence" by a greater integer. So (5) and (6) hold as stated, since H^{-1} and G^{-1} are increasing functions.

If $b' = b$, the bounds (5) and (6) are "best possible", in the sense that for any positive ε, there is a harmonically divergent sequence for which

$$n_i > H^{-1}(T(b, d, t) - \varepsilon)$$

and

$$i > G^{-1}(T(b, d, t) - \varepsilon).$$

It suffices to take a harmonically divergent sequence W which contains only terms of $S(b, d, t)$, until the partial sum of the series of reciprocals exceeds $T(b, d, t) - \varepsilon$.

Combining (5) and (6) with the bound of Theorem 1 proves the

COROLLARY. *With symbols as in Theorem 2,*

$$n_i \leqq H^{-1}(b^d(1+d\log b)) \tag{7}$$

$$i \leqq G^{-1}(b^d(1+d\log b)). \tag{8}$$

4. Applications

As one application of theorem 2, let α and c satisfy $0<\alpha<1$ and $c>0$. Then the sequence $\{n_j\}$, where $n_j=[cj^\alpha]$, and $[x]$ denotes "greatest integer $\leqq x$", is harmonically divergent. For this sequence,

$$G(j) \geqq \int_1^{j+1} \frac{dx}{cx^\alpha} = \frac{(j+1)^{1-\alpha}-1}{c(1-\alpha)};$$

so by (8),

$$i \leqq \{1+c(1-\alpha)b^d(1+d\log b)\}^{1/(1-\alpha)}-1. \tag{9}$$

Thus, for example, for every choice of $b' \leqq 10$ and $0 \leqq t < b$, there is an integer $n < 4.3\times 10^{14}$ (approx.), such that the representation of $[\pi n^{\frac{1}{2}}]$ to base b' contains digit t at least 6 times.

Similar conclusions apply to the sequences $\{[c\log j]\}$ and $\{[cj\log j]\}$, where $j=1, 2, 3, \cdots$.

Let P denote the sequence of prime numbers $\{p_1, p_2, \cdots\}$. Let $P_{\alpha\beta}$ denote the subsequence $\{P_{\alpha+h\beta}: h=0, 1, 2, \cdots\}$, for given integers α and β. It is well known (e.g. [2]) that $P_{\alpha\beta}$ is harmonically divergent, therefore Theorem 2 applies to $P_{\alpha\beta}$. To approximate to the bounds (7) and (8), let $a(n)=1$ when $n\in P_{\alpha\beta}$, $a(n)=0$ otherwise; let $A(n)=a(1)+a(2)+\cdots+a(n)$; then $A(n)$ equals the number of primes in $P_{\alpha\beta}$ which do not exceed n. Then for $P_{\alpha\beta}$,

$$H(2^k)=\sum_{j=1}^{2^k}\frac{a(j)}{j}=\frac{A(2^k)}{2^k}+\sum_{1}^{2^k-1}A(j)\cdot\left(\frac{1}{j}-\frac{1}{j+1}\right)$$

by Abel's transformation

$$\geqq \sum_{q=2}^{k}\sum_{j=2^{q-1}}^{2^q-1}A(j)\cdot\left(\frac{1}{j}-\frac{1}{j+1}\right)$$

$$>\sum_{q=2}^{k}A(2^{q-1})\cdot\left(\frac{1}{2^{q-1}}-\frac{1}{2^q}\right)$$

since $A(j)$ is non-decreasing

$$=\tfrac{1}{2}\sum_{2}^{k}A(2^{q-1})/2^{q-1}.$$

https://doi.org/10.1017/S1446788700027750 Published online by Cambridge University Press

From the Prime Number Theorem, $A(n)$ is given asymptotically by

$$A(n) \sim \frac{n}{\beta \log n} \quad \text{as} \quad n \to \infty. \tag{10}$$

Here the symbol $f(n) \sim g(n)$ means that

$$\lim_{n\to\infty} f(n)/g(n) = 1.$$

It will be convenient also to use the expressions "$f(n)$ is asymptotically less than $g(n)$" or "$g(n)$ is asymptotically greater than $f(n)$" to mean

$$\limsup_{n\to\infty} f(n)/g(n) \leqq 1.$$

If this holds, then for any $\varepsilon > 0$, $f(n) < (1+\varepsilon)g(n)$ for all n sufficiently large; $g(n)$ may thus also be termed an "asymptotic upper bound to $f(n)$", as $n \to \infty$.

Now from (10), for any $\varepsilon > 0$,

$$\frac{A(n)}{n} > \frac{1-\varepsilon}{\beta \log n} \quad \text{for all } n > n(\varepsilon). \tag{11}$$

Therefore

$$\tfrac{1}{2} \sum_{2}^{k} A(2^{q-1})/2^{q-1} > \frac{1-\varepsilon}{2\beta} \sum_{2}^{k} \frac{1}{\log 2^{q-1}} + B,$$

where the constant B represents the error arising from those terms in the summation to which inequality (11) does not apply; since the number of such terms depends on ε, but not on k, B does not depend on k.

Now if $k = [\log n/\log 2]$, then, for n sufficiently large,

$$\begin{aligned} H(n) \geqq H(2^k) &> B + \frac{1-\varepsilon}{2\beta \log 2} \sum_{2}^{k} \frac{1}{q-1} \\ &> B + \frac{(1-\varepsilon) \log ([\log n/\log 2])}{2\beta \log 2}. \end{aligned} \tag{12}$$

Let

$$L(n) = \{\log (\log n/\log 2)\}/\{2\beta \log 2\}.$$

Then, from (12), for any $\varepsilon > 0$,

$$\limsup_{n\to\infty} L(n)/H(n) \leqq 1/(1-\varepsilon).$$

Consequently, $H(n)$ is an asymptotically greater than $L(n)$, as $n \to \infty$. Therefore, from (7), an asymptotic upper bound $\bar{n}$ for n_i, as $d \to \infty$, is given by

$$2\beta \log 2 \cdot \left(b^d(1+d \log b)\right) = \log (\log \bar{n}/\log 2). \tag{13}$$

Let $P^*_{\alpha\beta}$ denote the set of prime numbers obtained by selecting arbitrarily exactly one prime from each subset

$$\{p_{\alpha+h\beta}, p_{\alpha+h\beta+1}, \cdots, p_{\alpha+(h+1)\beta-1}\},$$

where $h = 0, 1, 2, \cdots$. Then $P^*_{\alpha\beta}$ is also harmonically divergent, and the same asymptotic estimates, including (13), apply to $P^*_{\alpha\beta}$ as to $P_{\alpha\beta}$.

Similar results apply also to primes in arithmetic progression. Let y and z be relatively prime integers. Let Q denote the set of all primes $p \equiv z \pmod{y}$. Then LeVeque [3] shows that the number of primes in Q which do not exceed n is asymptotically

$$\frac{1}{\phi(y)} \int_2^n \frac{du}{\log u} \qquad \text{as } n \to \infty, \tag{14}$$

where $\phi(y)$ is Euler's function.

A similar discussion to that for $P_{\alpha\beta}$ then shows that Q is harmonically divergent, and the asymptotic bounds (12) and (13) apply also to Q, with $\beta = \phi(y)$.

As a numerical illustration of (13), set $b = 10$ and $d = 100$. Then for any base $\leqq 10$, there exists a prime p whose representation contains a given digit at least 100 times; and an upper bound $\bar{n}$ to p is asymptotically estimated by

$$\log_{10} \log_{10} \bar{n} = 1.4 \times 10^{102}.$$

Acknowledgement

My thanks are due to the referee for some improvements in the presentation of this paper.

References

[1] Apostol, T. M., Mathematical Analysis. (Addison-Wesley, 1957).
[2] Landau, E., Handbuch der Lehre von der Verteilung der Primzahlen, Vol. I.
[3] LeVeque, W. J., Topics in Number Theory, Vol. II. (Addison-Wesley, 1956).

University of Melbourne