NOTE ON NORMAL NUMBERS

ARTHUR H. COPELAND AND PAUL ERDÖS

D. G. Champernowne[1] proved that the infinite decimal

$$0.123456789101112\cdots$$

was normal (in the sense of Borel) with respect to the base 10, a normal number being one whose digits exhibit a complete randomness. More precisely a number is normal provided each of the digits $0, 1, 2, \cdots, 9$ occurs with a limiting relative frequency of $1/10$ and each of the 10^k sequences of k digits occurs with the frequency 10^{-k}. Champernowne conjectured that if the sequence of all integers were replaced by the sequence of primes then the corresponding decimal

$$0.12357111317\cdots$$

would be normal with respect to the base 10. We propose to show not only the truth of his conjecture but to obtain a somewhat more general result, namely:

THEOREM. *If $a_1, a_2, \cdots$ is an increasing sequence of integers such that for every $\theta<1$ the number of a's up to N exceeds N^θ provided N is sufficiently large, then the infinite decimal*

$$0.a_1a_2a_3\cdots$$

is normal with respect to the base β in which these integers are expressed.

On the basis of this theorem the conjecture of Champernowne follows from the fact that the number of primes up to N exceeds $cN/\log N$ for any $c<1$ provided N is sufficiently large. The corresponding result holds for the sequence of integers which can be represented as the sum of two squares since every prime of the form $4k+1$ is also of the form x^2+y^2 and the number of these primes up to N exceeds $c'N/\log N$ for sufficiently large N when $c'<1/2$.

The above theorem is based on the following concept of Besicovitch.[2]

DEFINITION. *A number A (in the base β) is said to be (ϵ, k) normal if any combination of k digits appears consecutively among the digits of A with a relative frequency between $\beta^{-k}-\epsilon$ and $\beta^{-k}+\epsilon$.*

Presented to the Society, September 17, 1945; received by the editors June 30, 1945, and, in revised form, January 3, 1946.

[1] J. London Math. Soc. vol. 8 (1933) pp. 254–260.

[2] Math. Zeit. vol. 39 (1935) pp. 146–147.

We prove the following lemma.

LEMMA. *The number of integers up to N (N sufficiently large) which are not (ϵ, k) normal with respect to a given base β is less than N^δ where $\delta = \delta(\epsilon, k, \beta) < 1$.*

First we prove the lemma for $(\epsilon, 1)$ normality. Let x be such that $\beta^{x-1} \leqq N < \beta^x$. Then there are at most

$$\beta \sum\nolimits_1 \beta_k + \beta \sum\nolimits_2 \beta_k$$

numbers up to N among whose digits there are less than $x(1-\epsilon)/\beta$ 0's, 1's, and so on, or more than $x(1+\epsilon)/\beta$ 0's, 1's, and so on, where $\beta_k = (\beta-1)^{x-k} C_{x,k}$ and where the summations $\sum_1$ and $\sum_2$ are extended over those values of k for which $k < (1-\epsilon)x/\beta$ and $k > (1+\epsilon)x/\beta$, respectively. The remaining numbers must have between $x(1-\epsilon)$ and $x(1+\epsilon)$ digits and hence for these remaining numbers the relative frequencies of 0's, 1's, 2's, and so on, must lie between $(1-\epsilon)/\beta(1+\epsilon)$ and $(1+\epsilon)/\beta(1-\epsilon)$. We have to show that $\beta(\sum_1 \beta_k + \sum_2 \beta_k) < N^\delta$. The following inequalities result from the fact that the terms of the binomial expansion increase up to a maximum and then decrease.

$$(1) \qquad \sum\nolimits_1 \beta_k < (x+1)\beta_{r_1}, \qquad \sum\nolimits_2 \beta_k < (x+1)\beta_{r_2},$$

where

$$(2) \qquad r_1 = [(1-\epsilon)x/\beta], \qquad r_2 = [(1+\epsilon)x/\beta]$$

and where $[(1-\epsilon)x/\beta]$ is the largest integer less than or equal to $(1-\epsilon)x/\beta$. Similarly for r_2. By repeated application of the relation

$$(3) \qquad \beta_{k+1}/\beta_k = (x-k)/(k+1)(\beta-1)$$

we obtain

$$\beta_{r_1} \rho_1^{\epsilon x/2} < \beta_{r'_1} < \beta^x$$

where

$$r'_1 = [(1-\epsilon/2)/\beta], \qquad \rho_1 = (x-r_1)/(r_1+1)\beta - 1)$$

and where $\rho_1 > 1$ for x sufficiently large. It follows that

$$\beta_{r_1} < (\rho_1^{-\epsilon/2}\beta)^x$$

and similarly

$$\beta_{r_2} < (\rho_2^{-\epsilon/2}\beta)^x.$$

Hence

$$\beta\left(\sum\nolimits_1 \beta_k + \sum\nolimits_2 \beta_k\right) < \beta(x+1)\left\{(\rho_1^{-\epsilon/2}\beta)^x + (\rho_2^{-\epsilon/2}\beta)^x\right\}$$
$$< \beta^{\delta(n-1)} \leqq N^\delta$$

and the lemma is established for $(\epsilon, 1)$ normality.

The extension to the case of (ϵ, k) normality is accomplished by a method similar to that used by Borel[3] and we shall only outline the proof. Consider the digits $b_0, b_1, \cdots$ of a number $m \leqq N$ grouped as follows:

$$b_0, b_1, \cdots, b_{k-1}; b_k, \cdots, b_{2k-1}; b_{2k}, \cdots, b_{3k-1}; \cdots.$$

Each of these groups represents a single digit of m when m is expressed in the base β^k. Hence there are at most N^δ integers $m \leqq N$ for which the frequency among these groups of a given combination of k digits falls outside the interval from $\beta^{-k}-\epsilon$ to $\beta^{-k}+\epsilon$.

The same holds for

$$b_1, b_2, \cdots, b_k; b_{k+1}, \cdots, b_{2k}; \cdots,$$

and so on. This gives our result.

To prove the theorem consider the numbers $a_1, a_2, \cdots$ of the increasing sequence up to the largest a less than or equal to N where $N=\beta^n$. At least $N^\theta - N^{(1-\epsilon)}$ of these numbers have at least $n(1-\epsilon)$ digits since by hypothesis there are at least N^θ of the numbers in this sequence and since at most $\beta^{n(1-\epsilon)} = N^{1-\epsilon}$ of them have fewer than $n(1-\epsilon)$ digits. Hence these numbers altogether have at least $n(1-\epsilon)(N^\theta - N^{1-\epsilon})$ digits. Let f_N be the relative frequency of the digit 0. It follows from the lemma that the number of a's for which the frequency of the digit 0 exceeds $\beta^{-1}+\epsilon$ is at most N^δ and hence

$$f_N < \beta^{-1} + \epsilon + \frac{nN^\delta}{n(1-\epsilon)(N^\theta - N^{1-\epsilon})}$$
$$= \beta^{-1} + \epsilon + \frac{N^{\delta-\theta}}{(1-\epsilon)(1 - N^{1-\epsilon-\theta})}.$$

Since we are permitted to take θ greater than δ and greater than $1-\epsilon$ it follows that $\lim_{N\to\infty} f_N$ is at most $\beta^{-1}+\epsilon$ and hence at most β^{-1}. Of course we have allowed N to become infinite only through values of the form β^n but this restriction can readily be removed. A similar result holds for the digits $1, 2, \cdots, \beta-1$ and hence each of these digits

[3] Ibid. p. 147.

must have a limiting relative frequency of exactly β^{-1}. In a similar manner it can be shown that the limiting relative frequency of any combination of k digits is β^{-k}. Hence the theorem is proved.

We make the following conjectures. First let $f(x)$ be any polynomial. It is very likely that $0.f(1)f(2) \cdots$ is normal. Besicovitch[4] proved this for $f(x)=x^2$. In fact he proved that the squares of almost all integers are (ϵ, k) normal. This no doubt holds for polynomials.

Second let $\beta_1, \beta_2, \cdots, \beta_r$ be integers such that no β is a power of any other. Then for any $\eta>0$ and large enough r the number of integers $m \leq n$ which are not (ϵ, k) normal for any of the bases β_i, $i \leq r$, is less than n^η. We cannot prove this conjecture.

UNIVERSITY OF MICHIGAN

[4] Ibid. p. 154.