NOTE ON NORMAL NUMBERS

ARTHUR H. COPELAND AND PAUL ERDÖS

D. G. Champernowne\textsuperscript{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 \(N^\theta\) provided \(N\) is sufficiently large, then the infinite decimal

\[ 0.a_1a_2a_3 \cdots \]

is normal with respect to the base \(\beta\) 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.\textsuperscript{2}

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

\textsuperscript{1} J. London Math. Soc. vol. 8 (1933) pp. 254–260.
We prove the following lemma.

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

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

$$\beta \sum_1 \beta_k + \beta \sum_2 \beta_k$$

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 $\beta_k = (\beta - 1)^{-k} C_{x,k}$ 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 $\beta(\sum_1 \beta_k + \sum_2 \beta_k) < N^x$. The following inequalities result from the fact that the terms of the binomial expansion increase up to a maximum and then decrease.

1. $\sum_1 \beta_k < (x + 1)\beta r_1$, $\sum_2 \beta_k < (x + 1)\beta r_2$,

where

2. $r_1 = \lceil (1 - \epsilon)x/\beta \rceil$, $r_2 = \lceil (1 + \epsilon)x/\beta \rceil$

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

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

we obtain

$$\beta_{r_1}^{x^{x/2}} < \beta r_1 < \beta^x$$

where

$$r_1' = \lceil (1 - \epsilon/2)/\beta \rceil$$

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

$$\beta_{r_1} < (\rho_1^{-1/2}\beta)^x$$

and similarly

$$\beta_{r_2} < (\rho_2^{-1/2}\beta)^x$$

Hence
\[
\beta \left( \sum_1 \beta_k + \sum_2 \beta_k \right) < \beta(x + 1) \{ (e_{1/2})^x + (e_{1/3})^x \} \\
< \beta^x(x - 1) \leq N^x
\]

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

The extension to the case of \((e, k)\) normality is accomplished by a method similar to that used by Borel\(^8\) and we shall only outline the proof. Consider the digits \(b_0, b_1, \ldots\) of a number \(m \leq N\) grouped as follows:

\[b_0, b_1, \ldots, b_{k-1}; b_k, \ldots, b_{2k-1}; b_{2k}, \ldots, b_{3k-1}; \ldots\]

Each of these groups represents a single digit of \(m\) when \(m\) is expressed in the base \(\beta^k\). Hence there are at most \(N^k\) integers \(m \leq N\) 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 \(b_0, b_1, \ldots, b_{k}, b_{k+1}, \ldots, b_{2k}, \ldots\)

and so on. This gives our result.

To prove the theorem consider the numbers \(a_1, a_2, \ldots\) 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 \(N^\theta\) 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 \(N^\theta\) and hence

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

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

\(^8\) Ibid. p. 147.
must have a limiting relative frequency of exactly $\beta^{-1}$. In a similar manner it can be shown that the limiting relative frequency of any combination of $k$ digits is $\beta^{-k}$. Hence the theorem is proved.

We make the following conjectures. First let $f(x)$ be any polynomial. It is very likely that $0.1f(1)f(2)\cdots$ is normal. Besicovitch proved this for $f(x) = x^2$. In fact he proved that the squares of almost all integers are $(\epsilon, k)$ normal. This no doubt holds for polynomials.

Second let $\beta_1, \beta_2, \cdots, \beta_r$ be integers such that no $\beta$ 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 $(\epsilon, k)$ normal for any of the bases $\beta_i$, $i \leq r$, is less than $n^\eta$. We cannot prove this conjecture.

*University of Michigan*

*Ibid. p. 154.*