

NOTE ON NORMAL DECIMALS

H. DAVENPORT AND P. ERDÖS

1. **Introduction.** A real number, expressed as a decimal, is said to be *normal* (in the scale of 10) if every combination of digits occurs in the decimal with the proper frequency. If $a_1 a_2 \dots a_k$ is any combination of k digits, and $N(t)$ is the number of times this combination occurs among the first t digits, the condition is that

$$(1) \quad \lim_{t \rightarrow \infty} \frac{N(t)}{t} = \frac{1}{10^k}.$$

It was proved by Champernowne [2] that the decimal $\cdot 1234567891011 \dots$ is normal, and by Besicovitch [1] that the same holds for the decimal $\cdot 1491625 \dots$. Copeland and Erdős [3] have proved that if p_1, p_2, \dots is any sequence of positive integers such that, for every $\theta < 1$, the number of p 's up to n exceeds n^θ if n is sufficiently large, then the infinite decimal $\cdot p_1 p_2 p_3 \dots$ is normal. This includes the result that the decimal formed from the sequence of primes is normal.

In this note, we prove the following result conjectured by Copeland and Erdős:

THEOREM 1. *Let $f(x)$ be any polynomial in x , all of whose values, for $x = 1, 2, \dots$, are positive integers. Then the decimal $\cdot f(1)f(2)f(3) \dots$ is normal.*

It is to be understood, of course, that each $f(n)$ is written in the scale of 10, and that the digits of $f(1)$ are succeeded by those of $f(2)$, and so on. The proof is based on an interpretation of the condition (1) in terms of the equal distribution of a sequence to the modulus 1, and the application of the method of Weyl's famous memoir [6].

Besicovitch [1] introduced the concept of the (ϵ, k) normality of an individual positive integer q , where ϵ is a positive number and k is a positive integer. The condition for this is that if $a_1 a_2 \dots a_l$ is any sequence of l digits, where $l \leq k$, then the number of times this sequence occurs in q lies between

$$(1 - \epsilon)10^{-l}q' \quad \text{and} \quad (1 + \epsilon)10^{-l}q'$$

where q' is the number of digits in q . Naturally, the definition is only significant when q is large compared with 10^k . We prove:

THEOREM 2. *For any ϵ and k , almost all the numbers $f(1), f(2), \dots$ are (ϵ, k) normal; that is, the number of numbers $n \leq x$ for which $f(n)$ is not (ϵ, k) normal is $o(x)$ as $x \rightarrow \infty$ for fixed ϵ and k .*

Received February 9, 1951.

This is a stronger result than that asserted in Theorem 1. But the proof of Theorem 1 is simpler than that of Theorem 2, and provides a natural introduction to it.

2. Proof of Theorem 1. We defined $N(t)$ to be the number of times a particular combination of k digits occurs among the first t digits of a given decimal. More generally, we define $N(u, t)$ to be the number of times this combination occurs among the digits from the $(u + 1)$ th to the t th, so that $N(0, t) = N(t)$. This function is almost additive; we have, for $t > u$,

$$(2) \quad N(u, t) \leq N(t) - N(u) \leq N(u, t) + (k - 1),$$

the discrepancy arising from the possibility that the combinations counted in $N(t) - N(u)$ may include some which contain both the u th and $(u + 1)$ th digits.

Let g be the degree of the polynomial $f(x)$. For any positive integer n , let x_n be the largest integer x for which $f(x)$ has less than n digits. Then, if n is sufficiently large, as we suppose throughout, $f(x_n + 1)$ has n digits, and so have $f(x_n + 2), \dots, f(x_{n+1})$. It is obvious that

$$(3) \quad x_n \sim \alpha(10^{1/g})^n \quad \text{as } n \rightarrow \infty,$$

where α is a constant.

Suppose that the last digit in $f(x_n)$ occupies the t_n th place in the decimal $\cdot f(1)f(2) \dots$. Then the number of digits in the block

$$f(x_n + 1)f(x_n + 2) \dots f(x_{n+1})$$

is $t_{n+1} - t_n$, and is also $n(x_{n+1} - x_n)$, since each f has exactly n digits. Hence

$$(4) \quad t_{n+1} - t_n = n(x_{n+1} - x_n).$$

It follows from (3) that

$$(5) \quad t_n \sim \alpha n(10^{1/g})^n \quad \text{as } n \rightarrow \infty.$$

To prove (1), it suffices to prove that

$$(6) \quad N(t_n, t) = 10^{-k}(t - t_n) + o(t_n)$$

as $n \rightarrow \infty$, for $t_n < t \leq t_{n+1}$. For, by (2), we have

$$N(t) - N(t_n) = \sum_{r=0}^{n-1} N(t_r, t_{r+1}) + N(t_n, t) + R,$$

for a suitable fixed h , where $|R| < nk$. Since (6) includes as a special case the result

$$N(t_r, t_{r+1}) = 10^{-k}(t_{r+1} - t_r) + o(t_r),$$

we obtain (1).

In proving (6), we can suppose without loss of generality that t differs from t_n by an exact multiple of n . Putting $t = t_n + nX$, the number $N(t_n, t)$ is the number of times that the given combination of k digits occurs in the block

$$(7) \quad f(x_n + 1)f(x_n + 2) \dots f(x_n + X),$$

where $0 < X \leq x_{n+1} - x_n$. We can restrict ourselves to those combinations which occur entirely in the same $f(x)$, since the others number at most $(k-1) \cdot (x_{n+1} - x_n)$, which is $o(t_n)$ by (3) and (5).

The number of times that a given combination $a_1 a_2 \dots a_k$ of digits occurs in a particular $f(x)$ is the same as the number of values of m with $k \leq m \leq n$ for which the fractional part of $10^{-m}f(x)$ begins with the decimal $\cdot a_1 a_2 \dots a_k$. If we define $\theta(z)$ to be 1 if z is congruent (mod 1) to a number lying in a certain interval of length 10^{-k} , and 0 otherwise, the number of times the given combination occurs in $f(x)$ is

$$\sum_{m=k}^n \theta(10^{-m}f(x)).$$

Hence

$$N(t_n, t) = \sum_{z=x_{n+1}}^{x_n+X} \sum_{m=k}^n \theta(10^{-m}f(x)) + O(x_{n+1} - x_n),$$

the error being simply that already mentioned.

To prove (6), it suffices to prove that

$$(8) \quad \sum_{m=k}^n \sum_{z=x_{n+1}}^{x_n+X} \theta(10^{-m}f(x)) = 10^{-k}nX + o(n(x_{n+1} - x_n))$$

for $0 < X \leq x_{n+1} - x_n$. We shall prove that if δ is any fixed positive number, and $\delta n < m < (1 - \delta)n$, then

$$(9) \quad \sum_{z=x_{n+1}}^{x_n+X} \theta(10^{-m}f(x)) = 10^{-k}X + o(x_{n+1} - x_n)$$

uniformly in m . This suffices to prove (8), since the contribution of the remaining values of m is at most $2\delta nX$, where δ is arbitrarily small. We have

$$(10) \quad X \leq x_{n+1} - x_n < \alpha(10^{1/\theta})^{n+1},$$

and we can also suppose that

$$(11) \quad X > (x_{n+1} - x_n)^{1-\delta} > \beta(10^{1/\theta})^{n(1-\delta)},$$

where β is a constant, since (9) is trivial if this condition is not satisfied.

The proof of (9) follows well-known lines. One can construct [6; 4, pp. 91-92, 99] for any $\eta > 0$, functions $\theta_1(z)$ and $\theta_2(z)$, periodic in z with period 1, such that $\theta_1(z) \leq \theta(z) \leq \theta_2(z)$, having Fourier expansions of the form

$$\begin{aligned} \theta_1(z) &= 10^{-k} - \eta + \sum_{\nu} A_{\nu}^{(1)} e(\nu z), \\ \theta_2(z) &= 10^{-k} + \eta + \sum_{\nu} A_{\nu}^{(2)} e(\nu z). \end{aligned}$$

Here the summation is over all integers ν with $\nu \neq 0$, and $e(w)$ stands for $e^{2\pi iw}$. The coefficients A_{ν} are majorized by

$$|A_\nu| \leq \min\left(\frac{1}{|\nu|}, \frac{1}{\eta\nu^2}\right).$$

Using these functions to approximate $\theta(10^{-m}f(x))$ in (9), we see that it will suffice to estimate the sum

$$S_{n,m,\nu} = \sum_{x=x_n+1}^{x_n+X} e(10^{-m}\nu f(x)).$$

We can in fact prove that

$$(12) \quad |S_{n,m,\nu}| < CX^{1-\zeta}$$

for all m and ν satisfying

$$(13) \quad \delta n < m < (1-\delta)n, \quad 1 \leq \nu < \eta^{-2},$$

where C and ζ are positive numbers depending only on δ , η and on the polynomial $f(x)$. This is amply sufficient to prove (9), since $X \leq x_{n+1} - x_n$.

The inequality (12) is a special case of Weyl's inequality for exponential sums. The highest coefficient in the polynomial $10^{-m}\nu f(x)$ is $10^{-m}\nu c/d$, where c/d is the highest coefficient in $f(x)$, and so is a rational number. Write

$$10^{-m}\nu \frac{c}{d} = \frac{a}{q},$$

where a and q are relatively prime integers. Let $G = 2^{g-1}$. Then, by Weyl's inequality¹,

$$(14) \quad |S_{n,m,\nu}|^G < C_1 X^\epsilon q^\epsilon (X^{G-1} + X^G q^{-1} + X^{G-g} q)$$

for any $\epsilon > 0$, where C_1 depends only on g and ϵ . In the present case, we have

$$q \leq 10^m d < 10^{(1-\delta)n} d,$$

and

$$q \geq 10^m \nu^{-1} c^{-1} > 10^{\delta n} \eta^{-2} c^{-1}.$$

This relates the magnitude of q to that of n . Relations between n and X were given in (10) and (11), and it follows that

$$C_2 X^{\epsilon \delta} < q < C_3 X^{\epsilon(1-\delta/2)},$$

where C_2 and C_3 depend only on η , c , d , and g . Using these inequalities for q in (14), we obtain a result of the form (12).

3. Proof of Theorem 2. We again consider the values of x for which $f(x)$ has exactly n digits, namely those for which $x_n < x \leq x_{n+1}$. We denote by $T(x)$ the number of times that a particular digit combination $a_1 a_2 \dots a_l$ (where $l \leq k$) occurs in $f(x)$. Then, with the previous notation,

$$T(x) = \sum_{m=1}^n \theta(10^{-m}f(x)).$$

¹The most accessible reference is [5, Satz 267]. The result is stated there for a polynomial with one term, but the proof applies generally.

We proved earlier that (putting $X = x_{n+1} - x_n$),

$$\sum_{x=x_n+1}^{x_n+X} T(x) \sim 10^{-l} n X \quad \text{as } n \rightarrow \infty.$$

Now our object is a different one; we wish to estimate the number of values of x for which $T(x)$ deviates appreciably from its average value, which is $10^{-l}n$.

For this purpose, we shall prove that

$$(15) \quad \sum_{x=x_n+1}^{x_n+X} T^2(x) \sim 10^{-2l} n^2 X \quad \text{as } n \rightarrow \infty.$$

When this has been proved, Theorem 2 will follow. For then

$$\begin{aligned} \sum_{x=x_n+1}^{x_n+X} (T(x) - 10^{-l}n)^2 &= \sum T^2(x) - 2(10^{-l}n) \sum T(x) + 10^{-2l}n^2 X \\ &= o(10^{-2l}n^2 X) \quad \text{as } n \rightarrow \infty. \end{aligned}$$

Hence the number of values of x with $x_n < x \leq x_{n+1}$, for which the combination $a_1 a_2 \dots a_l$ does *not* occur between $(1 - \epsilon)10^{-l}n$ and $(1 + \epsilon)10^{-l}n$ times, is $o(x_{n+1} - x_n)$ for any fixed ϵ . Since this is true for each combination of at most k digits, it follows that $f(x)$ is (ϵ, k) normal for almost all x .

To prove (15), we write the sum on the left as

$$(16) \quad \sum_{x=x_n+1}^{x_n+X} \sum_{m_1=1}^n \sum_{m_2=1}^n \theta(10^{-m_1}f(x))\theta(10^{-m_2}f(x)).$$

Once again, we can restrict ourselves to values of m_1 and m_2 which satisfy

$$(17) \quad \delta n < m_1 < (1 - \delta)n, \quad \delta n < m_2 < (1 - \delta)n,$$

since the contribution of the remaining terms is small compared with the right hand side of (15) when δ is small. For a similar reason, we can impose the restriction that

$$(18) \quad m_2 - m_1 > \delta n.$$

Proceeding as before, and using the functions $\theta_1(z)$ and $\theta_2(z)$, we find that it suffices to estimate the sum

$$(19) \quad S(n, m_1, m_2, \nu_1, \nu_2) = \sum_{x=x_n+1}^{x_n+X} e((10^{-m_1}\nu_1 + 10^{-m_2}\nu_2)f(x)),$$

for values of ν_1 and ν_2 which are not both zero, and satisfy $|\nu_1| < \eta^{-2}$, $|\nu_2| < \eta^{-2}$. If either ν_1 or ν_2 is zero, the previous result (7) applies. Supposing neither zero, we write the highest coefficient again as

$$(10^{-m_1}\nu_1 + 10^{-m_2}\nu_2) \frac{c}{d} = \frac{a}{q}.$$

In view of (17) and (18), we have

$$q \leq 10^{m_2}d < 10^{(1-\delta)n}d < C_3 X^{(1-\delta)\sigma} d.$$

We observe that a cannot be zero, since

$$10^{-m_2}|v_2| < 10^{-m_1-\delta n}|v_2| < \frac{1}{2}10^{-m_1}|v_1|,$$

provided that $2\eta^2 < 10^{\delta n}$, which is so for large n . Hence

$$q > \frac{2}{3}10^{m_1}|v_1|^{-1}c^{-1} > C_4X^{\delta q}.$$

It now follows as before from Weyl's inequality that

$$|S(n, m_1, m_2, v_1, v_2)| < CX^{1-\xi},$$

where again C and ξ are positive numbers depending only on δ , η , and the polynomial $f(x)$. Using this in (16), we obtain (15).

REFERENCES

1. A. S. Besicovitch, *The asymptotic distribution of the numerals in the decimal representation of the squares of the natural numbers*, Math. Zeit., vol. 39 (1934), 146-156.
2. D. G. Champernowne, *The construction of decimals normal in the scale of ten*, J. London Math. Soc., vol. 8 (1933), 254-260.
3. A. H. Copeland and P. Erdős, *Note on normal numbers*, Bull. Amer. Math. Soc., vol. 52 (1946), 857-860.
4. J. F. Koksma, *Diophantische Approximationen* (Ergebnisse der Math., IV, 4; Berlin, 1936).
5. E. Landau, *Vorlesungen über Zahlentheorie* (Leipzig, 1927).
6. H. Weyl, *Über die Gleichverteilung von Zahlen mod. Eins*, Math. Ann., vol. 77 (1916), 313-352.

University College, London
The University, Aberdeen