62. A Remark on a Theorem of Copeland-Erdös

By Iekata SHIOKAWA
Tokyo University of Education

(Comm. by Kôsaku YOSIDA, M. J. A., April 18, 1974)

Let $g\geq 2$ be a fixed integer. An infinite sequence $a_1a_2\cdots$ of non-negative integers not greater than $g-1$ is said to be normal to base g, if for every positive integer l and every sequence $B=b_1b_2\cdots b_l$ of digits $0,1,\cdots,g-1$, of length l we have

$$\lim_{n\to\infty}\frac{1}{n}N_n(B)=g^{-l},$$

where $N_n(B)$ is the number of indices i, $1\leq i\leq n$, for which $a_ia_{i+1}\cdots a_{i+l-1}=b_1b_2\cdots b_l$. Any positive integer n can be expressed uniquely in the form

$$n=\sum_{i=1}^{k}a_ig^{k-i}$$

where each $a_i=a_i(n)$ is one of $0,1,\cdots,g-1$, and $k=k(n)$ is the integer such that $g^{k-1}\leq n<g^k$, and we shall denote the sequence $a_1a_2\cdots a_{k(n)}$ by $B(n)$. An increasing sequence $\{m_1,m_2,\cdots\}$ of positive integers is said to be normal to base g, if the sequence of digits $B(m_1)B(m_2)\cdots$ is normal to base g. In 1946 Copeland-Erdös [1] proved that *any increasing sequence* $\{m_1,m_2,\cdots\}$ *of positive integers such that for every* $\theta<1$ *the number of* m_j*'s up to* x *exceeds* x^θ *provided* x *is sufficiently large, is normal to any base.* This theorem implies the normality (to any base) of the sequence of prime numbers, and this is the only known proof of this fact. In this paper we shall make a remark that the theorem of Copeland-Erdös is, in some sense, the best possible. Indeed we shall prove the following

Theorem. *For any fixed integer* $g\geq 2$ *and any fixed positive number* $\theta<1$ *we can construct a non-normal (to base* g*), increasing sequence of positive integers such that*

$$x^\theta<\sum_{m_j\leq x}1<g^2x^\theta$$

for all sufficiently large x.

To prove the theorem we need the following lemma.

Lemma. *Let* b *be any one of* $0,1,\cdots,g-1$, *and let* $\varepsilon<1/3$ *be any fixed positive number. Denote by* $T(b;k,\varepsilon)$ *the number of sequences* $B=b_1b_2\cdots b_k$ *of* 0's, 1's, $\cdots$, $g-1$*'s of length* k *such that* $N(b,B)>(g^{-1}+\varepsilon)k$, *where* $N(b,B)$ *be the number of* b*'s contained in the sequence* B. *Then we have*

$$T(b;k,\varepsilon)>g^k\exp(-16g\varepsilon^2k)$$

for all positive integers k satisfying $\varepsilon^2 k > \log k > 2$.

Proof. *Case* (i): Suppose that $k = ng$ and put

$$p(ng, l) = \binom{ng}{l}(g-1)^{ng-l}.$$

Then we have

$$T(b\,;k,\varepsilon) = \sum_{l > (g^{-1}+\varepsilon)ng} p(ng, l) = \sum_{j > \varepsilon ng} p(ng, n+j).$$

On the other hand, for $j = [\varepsilon ng] + 1$, we have

$$\frac{p(ng, n+j)}{p(ng, n)} = \frac{(ng-n)(ng-n-1)\cdots(ng-n-j+1)}{(n+1)(n+2)\cdots(n+j)(g-1)^j}$$
$$= \left(1+\frac{1}{n}\right)^{-1}\left(1+\frac{2}{n}\right)^{-1}\cdots\left(1+\frac{j}{n}\right)^{-1}\left(1-\frac{1}{n(g-1)}\right)\left(1-\frac{2}{n(g-1)}\right)\cdots\left(1-\frac{j-1}{n(g-1)}\right)$$
$$> \exp\left(-\frac{j(j+1)}{2n} - \frac{3}{2}\,\frac{j(j-1)}{2n(g-1)}\right) \ge \exp\left(-\frac{3}{2n}j^2\right).$$

At the same time we find easily

$$p(ng, n) > n^{-\frac{1}{2}} g^{ng}.$$

Thus we have

$$T(b\,;k,\varepsilon) > n^{-\frac{1}{2}} g^{ng} \exp\left(-\frac{3}{2n}(\varepsilon ng+1)^2\right)$$
$$\ge g^{ng}\exp\left(-\frac{27}{8}\varepsilon^2 ng - \frac{1}{2}\log n\right)$$
$$> g^{ng}\exp\left(-4\varepsilon^2 ng^2\right),$$

provided $\varepsilon ng \ge 2$ and $\varepsilon^2 ng > \log n$.

Case (ii): $k = ng + d$, $0 < d < g$. Let $B = b_1 b_2 \cdots b_k$ be any sequence of digits of length k and let $C = b_1 b_2 \cdots b_{ng}$. If $N(b, C) > (g^{-1} + 2\varepsilon)ng$ then $N(b, B) > (g^{-1}+\varepsilon)k$, provided $\varepsilon n \ge 2$. Therefore we obtain from the result in *Case* (i)

$$T(b\,;k,\varepsilon) \ge g^d T(b\,;ng, 2\varepsilon) > g^k \exp(-16g\varepsilon^2 k)$$

as required.

Proof of the theorem. Let $T^*(1\,;k,\varepsilon)$ be the number of integers m, $g^k \le m < 2g^k$, such that

$$N(1, B(m)) > (g^{-1}+\varepsilon)(k+1). \tag{1}$$

Consider the sequence consisting of the g^k possible arrangements of digits formed with 0's, 1's, $\cdots$, $g-1$'s and ranked in ascending order of magnitude, and denote it by

$$B_k(0), B_k(1), \cdots, B_k(l), \cdots, B_k(g^k - 1).$$

Then, for any m, $g^k \le m < 2g^k$, we have

$$N(1, B(m)) = 1 + N(1, B_k(l))$$

where $l = m - g^k$ and so

$$T^*(1;k,\varepsilon)\geq T(1;k,\varepsilon). \tag{2}$$

Now we choose a number $\varepsilon>0$ such that

$$1-16(\log g)^{-1}g\varepsilon^2>\theta \tag{3}$$

and let K be the least positive integer satisfying the following inequalities for all $k\geq K$:

$$\begin{aligned}&\varepsilon^2 k>\log k>2\\&(1-16(\log g)^{-1}g\varepsilon^2)k>\theta(k+2)>1.\end{aligned} \tag{4}$$

(By (3) such a positive integer K surely exists.) For any $k\geq K$, we set

$$\phi(k)=[g^{\theta(k+2)}-g^{\theta(k+1)}]+2$$

where $[x]$ denotes the integral part of x, and

$$\Phi(k)=\sum_{j=K}^{k}\phi(j),\qquad \Phi(K-1)=0.$$

Thus, taking account of (3) and (4), we have, by the lemma above,

$$T(1;k,\varepsilon)>\phi(k). \tag{5}$$

By (2) and (5), we can, for any $k\geq K$, choose $\phi(k)$ positive integers m in the interval $g^k\leq m<2g^k$ satisfying the condition (1), which we represent by m_j, $\Phi(k-1)<j\leq\Phi(k)$.

We shall prove that the increasing sequence of positive integers $\{m_1,m_2,\cdots\}$, constructed just above, has the properties mentioned in the theorem. It follows from the definition of the sequence $\{m_1,m_2,\cdots\}$ that

$$\begin{aligned}\sum_{m_j\leq x}1&\leq\sum_{k=1}^{k(x)}\sum_{g^{k-1}\leq m_j<g^k}1=\sum_{k=K+1}^{k(x)}\phi(k-1)\\&\leq\sum_{k=K+1}^{k(x)}(g^{\theta(k+1)}-g^{\theta k}+2)<g^{\theta(k(x)+1)}+2k(x)\\&\leq g^{2\theta}x^{\theta}+4\log x<g^2x^{\theta}\end{aligned}$$

for all sufficiently large x. Similarly we get

$$\sum_{m_j\leq x}1\geq\sum_{k=K+1}^{k(x)-1}(g^{\theta(k+1)}-g^{\theta k}+1)>x^{\theta}.$$

The non-normality to base g of this sequence is apparent, since we have by (1)

$$\limsup_{n\to\infty}\frac{1}{n}N_n(1)\geq\limsup_{j\to\infty}\frac{\sum_{i=1}^{j}N(1,B(m_i))}{\sum_{i=1}^{j}k(m_i)}>g^{-1}+\varepsilon.$$

This concludes the proof of our theorem.

Finally we remark that the key point of the proof of the Copeland-Erdös theorem is in the estimation of the number $T(b;k,\varepsilon)$ from above; more precisely,

$$T(b;k,\varepsilon)<g^k\exp(-cg\varepsilon^2k)$$

provided k is sufficiently large, where $c>0$ is an absolute constant. It is interesting that the fact that *almost all real numbers are normal (in the sense of E. Borel) to any base* can also be deduced from this inequality (see I. Niven [2]). An elegant proof of the above inequality can be found in the Niven's monograph [2] and our proof of the lemma has been carried out along almost the same lines as in [2].

References

[1] A. H. Copeland and P. Erdös: Notes on normal numbers. Bull. Amer. Math. Soc., **52**, 857–860 (1946).

[2] I. Niven: Irrational Numbers. The Carus Math. Monogr. No. 11. Math. Assoc. Amer., Washington, D. C. 1956. Especially, Chap. 8.