SOME FURTHER STATISTICAL PROPERTIES OF THE DIGITS IN CANTOR'S SERIES

By

P. ERDŐS (Budapest), corresponding member of the Academy, and

A. RÉNYI (Budapest), member of the Academy

Dedicated to G. ALEXITS *on the occasion of his 60th birthday*

Introduction

Let $q_1, q_2, \ldots, q_n, \ldots$ be an arbitrary sequence of positive integers, restricted only by the condition $q_n \geqq 2$. We can develop every real number x $(0 \leqq x \leqq 1)$ into Cantor's series

$$x = \sum_{n=1}^{\infty} \frac{\varepsilon_n(x)}{q_1 q_2 \cdots q_n} \tag{1}$$

where the n-th "digit" $\varepsilon_n(x)$ may take on the values $0, 1, \ldots, q_n - 1$ $(n = 1, 2, \ldots)$. The representation (1) is clearly a straightforward generalization of the ordinary decimal (or q-adic) representation of real numbers, to which it reduces if all q_n are equal to 10 (or to q, resp.).

In a recent paper [3] (see also [2] for a special case of the theorem) it has been shown that the classical theorem of BOREL [1] (according to which for almost all real numbers x the relative frequency of the numbers $0, 1, \ldots, 9$ among the first n digits of the decimal expansion of x tends for $n \to +\infty$ to $\frac{1}{10}$) can be generalized for all those representations (1) for which $\sum_{n=1}^{\infty} \frac{1}{q_n}$ is divergent. The generalization obtained in [2] can be formulated as follows: Let $f_n(k, x)$ denote the number of those among the digits $\varepsilon_1(x), \varepsilon_2(x), \ldots, \varepsilon_n(x)$ which are equal to k $(k = 0, 1, \ldots)$, i. e. put

$$f_n(k, x) = \sum_{\substack{\varepsilon_j(x) = k \\ 1 \leqq j \leqq n}} 1. \tag{2}$$

Let us put further

$$Q_n = \sum_{j=1}^{n} \frac{1}{q_j} \tag{3a}$$

and

$$Q_{n,k} = \sum_{\substack{j=1 \\ q_j > k}}^{n} \frac{1}{q_j}. \tag{3b}$$

Then for all non-negative integers k for which

$$\lim_{n \to +\infty} Q_{n,k} = +\infty, \tag{4}$$

we have for almost all x

$$\lim_{n\to+\infty} \frac{f_n(k,x)}{Q_{n,k}} = 1. \tag{5}$$

For those values of k for which $Q_{n,k}$ is bounded, $f_n(k,x)$ is bounded for almost all x. (For other related results see [4] and [5].)

In the present paper we consider the behaviour of

$$M_n(x) = \operatorname*{Max}_{(k)} f_n(k,x), \tag{6}$$

i. e. of the frequency of the most frequent number among the first n digits.

We shall discuss the three most important types of behaviour of $M_n(x)$:

Type 1. $\lim\limits_{n\to\infty} \frac{M_n(x)}{Q_n} = 1$ for almost all x. This is the case if q_n is constant or bounded, but also if e. g. $q_n \sim cn^\beta$ with $c>0$ and $0<\beta<1$ (see Theorem 1).

Type 2. $\lim\limits_{n\to+\infty} \frac{M_n(x)}{Q_n} = C$ for almost all x where $1<C<+\infty$. This is the case e. g. if $q_n \sim cn$ with $c>0$ (see Theorem 2).

Type 3. $\lim\limits_{n\to+\infty} \frac{M_n(x)}{Q_n} = +\infty$ for almost all x. This is the case e. g. if $q_n \sim n(\log n)^\alpha$ with $0<\alpha\leqq 1$ (see Theorem 3).

There exist, of course, sequences q_n for which $\lim\limits_{n\to\infty} \frac{M_n(x)}{Q_n}$ does not exist for almost all x, but we do not consider such cases in the present paper. We shall deal with the case when $\sum \frac{1}{q_n} < +\infty$ and with some other questions on Cantor's series in another paper.

All results obtained are based on the evident fact that the digits $\varepsilon_n(x)$, considered as random variables on the probability space $[\Omega, \mathcal{A}, \mathbf{P}]$, where Ω is the interval $(0,1)$, $\mathcal{A}$ the set of all measurable subsets of Ω and $\mathbf{P}(A)$ is for $A\in\Omega$ the Lebesgue measure of A, are independent and have the probability distribution

$$\mathbf{P}(\varepsilon_n(x)=k) = \frac{1}{q_n} \qquad (k=0,1,\ldots,q_n-1). \tag{7}$$

§ 1. Type 1 behaviour of $M_n(x)$

In case q_n is bounded, $q_n \leqq K$, we have by (5)

$$\lim_{n\to\infty} \frac{N_n(0,x)}{Q_n} = 1 \quad \text{and} \quad \overline{\lim_{n\to\infty}} \frac{N_n(k,x)}{Q_n} \leqq 1 \quad \text{for} \quad k \geqq 1$$

and thus, as in this case $M_n(x) = \max\limits_{0 \leqq k < K} f_n(k, x)$, we obtain for almost all x

$$\lim_{n \to +\infty} \frac{M_n(x)}{Q_n} = 1. \tag{8}$$

We shall show that (8) is valid under more general conditions. We prove in this direction the following

THEOREM 1. *If*

$$\lim_{n \to +\infty} \frac{Q_n}{\log n} = +\infty, \tag{9}$$

then we have for almost all x

$$\lim_{n \to +\infty} \frac{M_n(x)}{Q_n} = 1. \tag{10}$$

PROOF OF THEOREM 1. Let $\mathcal{A}$ denote the set of those numbers n for which $q_n < n^3$. Let us denote the elements of the complementary set $\overline{\mathcal{A}}$ of $\mathcal{A}$ by n_j ($n_j < n_{j+1}$; $j = 1, 2, \ldots$), then we have $n_j \geqq j$ and therefore $q_{n_j} \geqq n_j^3 \geqq j^3$.

Then we have for any k

$$\sum_{j \in \overline{\mathcal{A}}} \mathbf{P}(\varepsilon_j(x) = k) = \sum_j \mathbf{P}(\varepsilon_{n_j}(x) = k) = \sum_j \frac{1}{q_{n_j}} \leqq \sum_{j=1}^{\infty} \frac{1}{j^3} < +\infty$$

and therefore, by the Borel—Cantelli lemma for almost every x, every k occurs only a finite number of times in the sequence $\varepsilon_{n_j}(x)$. On the other hand, the probability that a number k occurs more than once in the sequence $\varepsilon_{n_j}(x)$ $(j = 1, 2, \ldots)$ does not exceed

$$W_k = \sum_{\substack{q_{n_i} > k \\ q_{n_j} > k \\ j > i}} \frac{1}{q_{n_i} q_{n_j}}$$

and we have

$$\sum_{k=0}^{\infty} W_k = \sum_{i<j}\sum \frac{\min(q_{n_i}, q_{n_j})}{q_{n_i} q_{n_j}} = \sum_{i=1}^{\infty} \sum_{j>i} \frac{1}{q_{n_j}} \leqq \sum_{i=1}^{\infty} \sum_{j=i+1}^{\infty} \frac{1}{j^3} < +\infty.$$

Thus, using again the Borel—Cantelli lemma, it follows that for almost all x only a finite number of integers k may occur more than once in the sequence $\varepsilon_{n_i}(x)$. This, together with what has been proved above, implies that for almost every x in the sequence $\varepsilon_{n_i}(x)$ only a finite number of values occur more than once and these values occur also only a finite number of times. By other words, in proving Theorem 1 we may suppose that

$$q_n < n^3 \quad \text{for all values of } n \tag{11}$$

without the restriction of generality.

Clearly, we have

$$\frac{M_n(x)}{Q_n} \geqq \frac{f_n(0,x)}{Q_n}$$

and thus, taking into account that owing to (9) condition (4) is fulfilled for $k=0$, it follows by (5) that

$$\varliminf_{n\to+\infty} \frac{M_n(x)}{Q_n} \geqq 1.$$

Thus to prove Theorem 1 it suffices to show that for almost all x

$$\varlimsup_{n\to+\infty} \frac{M_n(x)}{Q_n} \leqq 1. \tag{12}$$

As by (4) we have for any k_0

$$\varlimsup_{n\to+\infty} \frac{\operatorname{Max}_{0\leqq k\leqq k_0} f_n(k,x)}{Q_n} \leqq 1,$$

(12) will be proved if we show that for any $\varepsilon>0$ and for some k_0 which may depend on ε, putting

$$M_n^{(k_0)}(x) = \operatorname{Max}_{k>k_0} f_n(k,x), \tag{13}$$

we have

$$\varlimsup_{n\to\infty} \frac{M_n^{(k_0)}(x)}{Q_n} \leqq 1+\varepsilon. \tag{14}$$

To prove (14) we start by calculating the probability $\mathbf{P}(f_n(k,x)=j)$. In what follows $c_1, c_2, \dots$ denote positive absolute constants. We evidently have

$$\mathbf{P}(f_n(k,x)=j) = \left(\sum_{\substack{1\leqq i_1<i_2<\dots<i_j\leqq n \\ q_{i_r}>k;\ r=1,2,\dots,j}} \frac{1}{(q_{i_1}-1)\cdots(q_{i_j}-1)} \right) \cdot \prod_{\substack{h=1 \\ q_h>k}}^{n} \left(1-\frac{1}{q_h}\right). \tag{15}$$

It follows that

$$\mathbf{P}(f_n(k,x)=j) \leqq e^{-Q_{n,k}} \frac{(Q^*_{n,k})^j}{j!} \tag{16}$$

where

$$Q^*_{n,k} = \sum_{\substack{j\leqq n_j \\ q_j>k}} \frac{1}{q_j-1}. \tag{17}$$

Using the well-known identity

$$e^{-\lambda} \sum_{j=N}^{\infty} \frac{\lambda^j}{j!} = \frac{1}{N!} \int_0^{\lambda} t^N e^{-t}\,dt$$

we obtain for $0<\lambda<\dfrac{N}{1+\varepsilon}$

$$e^{-\lambda} \sum_{j=N}^{\infty} \frac{\lambda^j}{j!} \leqq \frac{c_1}{\varepsilon\sqrt{N}} e^{-\frac{(N-\lambda)^2}{2N}}. \tag{18}$$

Thus we obtain for $0<\varepsilon<1$, in view of

$$Q^{*}_{n,k} \leqq Q_{n,k}\left(1+\frac{1}{k}\right) \leqq Q_n\left(1+\frac{1}{k}\right), \tag{19}$$

that

$$\mathbf{P}(f_n(k,x) \geqq (1+\varepsilon) Q_n) \leqq \frac{c_1}{\varepsilon\sqrt{Q_n}} e^{\frac{Q_n}{k}} e^{-\frac{Q_n\left(\varepsilon-\frac{1}{k}\right)^2}{4}}. \tag{20}$$

We obtain from (20) for $k \geqq \frac{8}{\varepsilon^2}$

$$\mathbf{P}(f_n(k,x) \geqq (1+\varepsilon) Q_n) \leqq \frac{c_1}{\varepsilon\sqrt{Q_n}} e^{-\frac{\varepsilon^2 Q_n}{16}}. \tag{21}$$

This implies, putting $k_0=\left[\frac{8}{\varepsilon^2}\right]+1$ and taking (11) into account,

$$\mathbf{P}(M_n^{(k_0)}(x) \geqq (1+\varepsilon) Q_n) \leqq \sum_{k=k_0}^{n^3} \mathbf{P}(f_n(k,x) \geqq (1+\varepsilon) Q_n) \leqq \frac{c_1 n^3}{\varepsilon\sqrt{Q_n}} e^{-\frac{\varepsilon^2 Q_n}{16}}. \tag{22}$$

As by (9) we have for $n \geqq n_0$ $Q_n > \frac{80}{\varepsilon^2}\log n$, it follows that

$$\mathbf{P}(M_n^{(k_0)}(x) \geqq (1+\varepsilon) Q_n) \leqq \frac{c_2}{n^2}. \tag{23}$$

Thus

$$\sum_{n=1}^{\infty} \mathbf{P}(M_n^{(k_0)}(x) \geqq (1+\varepsilon) Q_n) < +\infty \tag{24}$$

and therefore by the lemma of Borel—Cantelli, the inequality $M_n^{(k_0)}(x) \geqq (1+\varepsilon) Q_n$ can be satisfied for almost all x only for a finite number of values of n. This implies (14) for almost all x which proves Theorem 1.

§ 2. Type 2 behaviour of $M_n(x)$

In this § we shall prove the following rather surprising

THEOREM 2. *If*

$$0 < c_2 \leqq \frac{q_n}{n} \leqq c_3 \qquad (n=1,2,\ldots) \tag{25}$$

and

$$\lim_{n\to+\infty} \frac{Q_n}{\log n} = \alpha > 0, \tag{26}$$

then we have for almost all x

$$\lim_{n\to+\infty} \frac{M_n(x)}{Q_n} = y(\alpha) \tag{27}$$

where $y=y(\alpha)>1$ is the unique (real) solution of the equation

$$y \log y = \frac{1}{\alpha}. \tag{28}$$

PROOF OF THEOREM 2. We start from the inequality, which follows simply from Stirling's formula,

$$e^{-\lambda} \sum_{j=N}^{\infty} \frac{\lambda^j}{j!} \leqq c_4 \left(\frac{\lambda e}{N}\right)^N e^{-\lambda} \tag{29}$$

for $N>\beta\lambda$ with fixed $\beta>1$ where c_4 depends on β.

Now evidently (25) and (26) imply that

$$Q_{n,k} = \alpha \log \frac{n}{k} + o(\log n). \tag{30}$$

Thus, by virtue of (16), we have, if $Y>y(\alpha)$ where $y(\alpha)$ denotes the solution of the equation (28), for any ε with $0<\varepsilon<\alpha Y \log Y-1$ and $n \geqq n_0(\varepsilon)$

$$\sum_{k=1}^{c_3 n} \mathbf{P}(f_n(k,x) \geqq YQ_n) \leqq \frac{c_5}{n^{\alpha Y \log Y - 1 - \varepsilon}}. \tag{31}$$

Thus

$$\mathbf{P}(M_n(x) > YQ_n) \leqq \frac{c_5}{n^\delta} \quad \text{for} \quad n \geqq n_0(\varepsilon) \tag{32}$$

where $\delta = \alpha Y \log Y - 1 - \varepsilon > 0$. It follows that

$$\sum_{s=1}^{\infty} \mathbf{P}(M_{2^s}(x) > YQ_{2^s}) < +\infty \tag{33}$$

and therefore by the Borel—Cantelli lemma the number of those values of s for which $M_{2^s}(x) > YQ_{2^s}$ is finite for almost every x. If $2^{s-1}<n<2^s$, let us choose an arbitrary number Y_1 such that $y(\alpha)<Y<Y_1$, then

$$\frac{M_n(x)}{Q_n} \leqq \frac{M_{2^s}(x)}{Q_{2^{s-1}}} \leqq \frac{Y_1}{Y} \frac{M_{2^s}(x)}{Q_{2^s}}$$

if $s \geqq s_0$. Thus, if for such an n $M_n(x) > Y_1 Q_n$, then $M_{2^s}(x) > YQ_{2^s}$. As the last inequality can be valid for almost all x only for a finite number of values of s, it follows that $M_n(x) > Y_1 Q_n$ is valid for almost all x only for a finite number of values of n. As Y_1 may be equal to any number greater than $y(\alpha)$, this implies that for almost all x

$$\overline{\lim_{n\to\infty}} \frac{M_n(x)}{Q_n} \leqq y(\alpha). \tag{34}$$

It remains to prove that we have also

$$\underline{\lim_{n\to\infty}} \frac{M_n(x)}{Q_n} \geqq y(\alpha) \tag{35}$$

for almost all x.

As for any sequence of positive numbers $b_1, b_2, \ldots, b_N$ we have

$$\sum_{1 \leqq i_1 < i_2 < \ldots < i_j \leqq N} b_{i_1} b_{i_2} \cdots b_{i_j} \geqq \frac{\left(\sum_{i=1}^{N} b_i\right)^j}{j!} - \frac{1}{2}\left(\sum_{i=1}^{N} b_i^2\right) \frac{\left(\sum_{i=1}^{N} b_i\right)^{j-2}}{(j-2)!},$$

we obtain from (15)

$$\mathbf{P}(f_n(k, x) = j) \geqq c_6 e^{-Q_{n,k}} \left(\frac{Q_{n,k}^j}{j!} - \frac{Q_{n,k}^{j-2} \sum_{q_i > k, i \leqq n} \frac{1}{q_i^2}}{2(j-2)!} \right). \tag{36}$$

Taking into account that

$$\sum_{i \leqq n, q_i > k} \frac{1}{q_i^2} \leqq \frac{c_7}{k}$$

and that for $j = yQ_n$ and $k \leqq n^{1-\varepsilon}$

$$\frac{j^2}{(Q_{n,k})^2} \leqq \frac{c_8}{\varepsilon^2}$$

if $1 < y < y(\alpha)$ where $y(\alpha)$ denotes the solution of (28) and $0 < \varepsilon < 1 - \alpha y \log y$, it follows that

$$\sum_{\log^2 n \leqq k < n} \mathbf{P}(f_n(k, x) \geqq yQ_n) \geqq c_9 n^\delta \quad \text{for} \quad n \geqq n_0(\varepsilon) \tag{37}$$

where $\delta = 1 - \alpha y \log y - \varepsilon > 0$.

Now it is easy to see that

$$\begin{aligned} &\mathbf{P}(f_n(k_1, x) = j_1, f_n(k_2, x) = j_2) \leqq \\ &\leqq \left(1 + \frac{j_1}{k_1}\right)\left(1 + \frac{j_2}{k_2}\right) \mathbf{P}(f_n(k_1, x) = j_1) \mathbf{P}(f_n(k_2, x) = j_2). \end{aligned} \tag{38}$$

It follows that for $k_1 \geqq \log^3 n$, $k_2 \geqq \log^3 n$ we have for any y with $1 < y < y(\alpha)$, where $y(\alpha)$ is the solution of the equation (28),

$$\begin{aligned} &\mathbf{P}(f_n(k_1, x) \geqq yQ_n, f_n(k_2, x) \geqq yQ_n) \leqq \\ &\leqq \mathbf{P}(f_n(k_1, x) \geqq yQ_n) \mathbf{P}(f_n(k_2, x) \geqq yQ_n) \left(1 + O\left(\frac{1}{\log^2 n}\right)\right). \end{aligned}$$

If we define $\eta_n = \eta_n(x)$ as the number of those values of k for which $\log^2 n \leqq k \leqq n$ and $f_n(k, x) \geqq yQ_n$, we have, denoting by $\mathbf{M}(\eta_n)$ the mean value and by $\mathbf{D}^2(\eta_n)$ the variance of η_n,

$$\mathbf{M}(\eta_n) \geqq c_9 n^\delta \tag{39}$$

and

$$\mathbf{D}^2(\eta_n) \leqq c_{10} \frac{\mathbf{M}^2(\eta_n)}{\log^2 n}. \tag{40}$$

It follows by the inequality of Chebyshev

$$\mathbf{P}(\eta_n=0)\leqq\mathbf{P}(|\eta_n-\mathbf{M}(\eta_n)|\geqq\mathbf{M}(\eta_n))\leqq\frac{c_{10}}{\log^2 n} \tag{41}$$

and thus

$$\sum_{n=1}^{\infty}\mathbf{P}(\eta_{2^n}=0)<+\infty. \tag{42}$$

It follows by the Borel—Cantelli lemma that we have for almost all x

$$M_{2^n}(x)\geqq y\,Q_{2^n} \quad \text{for} \quad n\geqq n_0(x).$$

Thus for any $\varepsilon>0$ and for $n\geqq n_1(x,\varepsilon)$ and $2^n\leqq N<2^{n+1}$ we have

$$M_N(x)\geqq M_{2^n}(x)\geqq y\,Q_{2^n}\geqq(y-\varepsilon)\,Q_N. \tag{43}$$

This implies that for almost all x

$$\varliminf_{n\to\infty}\frac{M_n(x)}{Q_n}\geqq y. \tag{44}$$

As y may be any number not exceeding $y(\alpha)$, we obtain from (44) that (35) is also valid for almost all x. Thus the proof of Theorem 2 is complete.

§ 3. Type 3 behaviour of $M_n(x)$

Now we shall prove a theorem which deals with conditions under which $\frac{M_n(x)}{Q_n}$ tends to $+\infty$ for almost every x.

THEOREM 3. *Let us suppose that*

$$\lim_{n\to+\infty}\frac{q_n}{n}=+\infty, \tag{45}$$

but at the same time

$$\lim_{n\to+\infty}Q_n=+\infty. \tag{46}$$

Then we have for almost every x

$$\lim_{n\to+\infty}\frac{M_n(x)}{Q_n}=+\infty. \tag{47}$$

PROOF OF THEOREM 3. The proof follows the same pattern as the second half of the proof of Theorem 2 (i. e. the proof of (35)).

We have from (45)

$$Q_n=\sum_{i=1}^{n}\frac{1}{q_i}=\sum_{i=1}^{n}\frac{1}{i}\,\frac{i}{q_i}=o(\log n), \tag{48}$$

further for any $A>0$

(49) $$\sum_{q_j<e^{AQ_n}} \frac{1}{q_j} = o(\log e^{AQ_n}) = o(Q_n)$$

and thus

(50) $$Q_{n,k} \geqq Q_n(1-o(1)) \quad \text{for} \quad k \leqq e^{AQ_n}.$$

It follows from (36) that for any $N>N_0>1$

(51) $$\mathbf{P}(f_n(k,x) \geqq NQ_n) \geqq e^{-N\log N \cdot Q_n}.$$

Now let us choose $A=3N\log N$, then we have

$$\sum_{Q_n^2 \leqq k \leqq e^{AQ_n}} \mathbf{P}(f_n(k,x) \geqq NQ_n) \geqq e^{2N\log N \cdot Q_n}.$$

On the other hand, we have from (38)

$$\mathbf{P}(f_n(k_1,x) \geqq NQ_n, f_n(k_2,x) \geqq NQ_n) \leqq$$
$$\leqq \mathbf{P}(f_n(k_1,x) \geqq NQ_n)\mathbf{P}(f_n(k_2,x) \geqq NQ_n)\left(1+O\left(\frac{1}{Q_n}\right)\right)$$

and thus, defining $\eta_n=\eta_n(x)$ as the number of those values of k for which $Q_n^2 \leqq k \leqq e^{AQ_n}$ and $f_n(k,x)>NQ_n$, we have $\mathbf{M}(\eta_n) \to +\infty$ and

$$\mathbf{D}^2(\eta_n) \leqq c_{12} \frac{\mathbf{M}^2(\eta_n)}{Q_n}.$$

Similarly as in the proof of Theorem 2 we obtain that

$$\overline{\lim_{n\to\infty}} \frac{M_n(x)}{Q_n} \geqq N$$

for almost all x. As N may be chosen arbitrarily large, Theorem 3 follows.

(Received 29 October 1958)

References

[1] É. Borel, Sur les probabilités dénombrables et leurs applications arithmétiques, *Rendiconti del Circ. Mat. di Palermo*, **26** (1909), pp. 247—271.

[2] A. Rényi, On a new axiomatic theory of probability, *Acta Math. Acad. Sci. Hung.*, **6** (1955), pp. 285—335.

[3] A. Rényi, A számjegyek eloszlása valós számok Cantor-féle előállításaiban, *Mat. Lapok*, **7** (1956), pp. 77—100.

[4] A. Rényi, Representations for real numbers and their ergodic properties, *Acta Math. Acad. Sci. Hung.*, **8** (1957), pp. 477—493.

[5] P. Erdős, A. Rényi and P. Szüsz, On Engel's and Sylvester's series, *Annales Univ. Sci. Budapest*, Sectio Math., **1** (1958), pp. 7—32.