Časopis pro pěstování matematiky, roč. 96 (1971), Praha

ON SOME NEW PROPERTIES OF THE CANTOR SET

SRI SACHI BHUSAN SEN GUPTA, Kalyani

(Received September 16, 1969)

Introduction and notations. Suppose that the real number x is expressed in the Scale g (g is a positive integer >1)

$$x = \frac{c_1(x)}{g} + \frac{c_2(x)}{g^2} + \ldots + \frac{c_n(x)}{g^n} + \ldots \tag{1}$$

$0 \leqq c_i(x) < g$, $i = 1, 2, \ldots$ and that the digit b, $0 \leqq b \leqq g - 1$ occurs n_b times in the first n places of the expression (1) for x.

If $\lim_{n\to\infty} n_b/n$ exists and equal to β then we say that the digit b has frequency β. [See HARDY and WRIGHT [9]].

We say that x is simply normal in the scale g if $\lim_{n\to\infty} n_b/n = 1/g$, for each of the $(g - 1)$ possible values of b [See [9]].

Let

$$\sum_{n=1}^{\infty} d_n = d_1 + d_2 + d_3 + \ldots + d_n + \ldots \tag{2}$$

be an infinite series and let $\{k_n\}$ be an ascending sequence of positive integers; then the series

$$\sum_{n=1}^{\infty} d_{k_n} = d_{k_1} + d_{k_2} + \ldots + d_{k_n} + \ldots \tag{3}$$

is called a subseries of the series (2).

Let each number of the interval $(0, 1]$ be expressed in the scale 2 with infinitely many digits equal to 1.

Hence, if $x \in (0, 1]$, then

$$x = \sum_{k=1}^{\infty} \frac{\varepsilon_k(x)}{2^k} = \frac{\varepsilon_1(x)}{2} + \frac{\varepsilon_2(x)}{2^2} + \ldots \tag{4}$$

where $\varepsilon_k(x) = 0$ or 1, and $\varepsilon_k(x) = 1$, for infinitely many k.

We have correspondingly an infinite series

$$(x) = \sum_{k=1}^{\infty} \varepsilon_k(x)\, d_k\,, \tag{5}$$

which is a subseries of (2).

Also every subseries (3) of the series (2) can be obtained from (5), [by putting $\varepsilon_{k_n}(x) = 1$, $n = 1, 2, \ldots$ and $\varepsilon_k(x) = 0$ when $k \neq k_n$, $n = 1, 2, \ldots$].

Hence all subseries of (2) can be mapped onto (0, 1]. We say that certain property P is valid for almost all subseries of (2), if the corresponding set $\{x\}$, $x \in (0, 1]$, has the Lebesgue measure 1. For instance, we know that almost all subseries of a divergent series are divergent. [See [8]].

Let (5) be a subseries of the series (2), and let $p(n, x) = \sum_{k=1}^{n} \varepsilon_k(x)$. Then the numbers

$$p_1(x) = \liminf_{n\to\infty} \frac{p(n, x)}{n}\,, \quad p_2(x) = \limsup_{n\to\infty} \frac{p(n, x)}{n}$$

are called lower and upper asymptotic density respectively of the subseries (5) in the series (2).

If the limit $p(x) = \lim_{n\to\infty} (p(n, x)/n)$ $(= \underline{\lim}\, (p(n, x)/n) = \overline{\lim}\, (p(n, x)/n))$ exists, then we call this number asymptotic density of (5) in (2). Obviously $p_1(x), p_2(x), p(x) \in [0, 1]$ [See [12]].

Theorem 1. *For almost all points* $(x) = \sum_{k=1}^{\infty} (2\varepsilon_k(x)/3^k) = \sum_{k=1}^{\infty} (c_k(x)/3^k)$ *of the Cantor set* C, *each of the digits* 0, 2 *has the frequency* $\frac{1}{2}$.

[That is almost all points of C have nearly equal number of twos and zeros in the first n digits, where n is sufficiently large and each point is expressed in the ternary scale.]

Proof. We know the Theorem that almost all numbers are simply normal in any given scale g [See [9]].

It follows that almost all numbers of (0, 1] are simply normal in the scale 2 (i.e. $g = 2$).

That is, if $x = \sum_{k=1}^{\infty} (\varepsilon_k(x)/2^k) \in (0, 1]$, $\varepsilon_k(x) = 0$ or 1 and $\varepsilon_k(x) = 1$, for infinitely many k and if the digit 1 (or 0), (i.e. $b = 1$ or 0), occurs n_b times among the first n numbers $\varepsilon_1(x), \varepsilon_2(x), \ldots, \varepsilon_n(x)$, then

$$\lim_{n\to\infty} \frac{n_b}{n} = \frac{1}{2}\,, \quad \text{for almost all} \quad x \in (0, 1]\,. \tag{6}$$

Now consider the Cantor series $2/3 + 2/3^2 + \ldots + 2/3^n + \ldots$ we form the Cantor point

$$(x) = \sum_{k=1}^{\infty} \frac{2\varepsilon_k(x)}{3^k}, \quad \text{corresponding to} \quad x = \sum_{k=1}^{\infty} \frac{\varepsilon_k(x)}{2^k}.$$

It follows from (6) that the digit 2 (and also 0) has the frequency $\frac{1}{2}$ in the expression for (x), for almost all $(x) \in C$.

Hence the theorem.

Note 1. For any Cantor point $x = \cdot\delta 1 = \cdot\delta 022\ldots$ (Scale 3), (which is the left hand end point of an interval complementary to the Cantor set C, and δ is a finite complex of 0's and 2's), we have

$$\lim_{n\to\infty} \frac{n_2}{n} = 1, \quad \lim_{n\to\infty} \frac{n_0}{n} = 0$$

(n_b is the number of b's in the first n digits of $\cdot\delta 1$, $b = 2, 0$). For the Cantor point $x = \cdot\delta$, which is the right hand end point of a contiguous interval, $\lim_{n\to\infty} (n_2/n) = 0$ and $\lim_{n\to\infty} (n_0/n) = 1$.

Note 2. If we represent the numbers in $(0, 1]$ in the ternary scale as

$$x = \frac{c_1(x)}{3} + \frac{c_2(x)}{3^2} + \ldots + \frac{c_k(x)}{3^k} + \ldots, \quad \text{where} \quad c_i(x) = 0, 1, 2$$

and $N_n(r, x)$ as the number of $c_k(x)$ in the first n terms, each having the integral value r ($=0, 1, 2$), then we know that $\lim_{n\to\infty} (N_n(r, x)/n) = \frac{1}{3}$, for almost all x in $(0, 1]$, [9].

If we denote this set of simply normal numbers (of measure 1) by N_3, then we know that the set N_3 is of First Category [See [13]].

Also, if we denote the derived set of the sequence

$$\frac{N_1(r, x)}{1}, \frac{N_2(r, x)}{2}, \ldots, \frac{N_n(r, x)}{n}, \ldots \equiv \left\{\frac{N_n(r, x)}{n}\right\}$$

by $\{N_n(r, x)/n\}'_n$, it has been shown by Tibor Šalát [13] that, for all $x \in (0, 1]$, except for a set of the first Category (F.C.), [including N_3]

$$\left\{\frac{N_n(r, x)}{n}\right\}'_n = [0, 1], \quad \text{for each} \quad r (= 0, 1, 2).$$

If we now consider the perfect set C (the Cantor set) instead of the whole interval

[0, 1], where each point (in the scale 3) x is given as $x = \sum_{k=1}^{\infty} (2\varepsilon_k(x)/3^k)$, $\varepsilon_k(x) = 0, 1$, we have seen above in Theorem 1 that

$$\lim_{n\to\infty} \frac{N_n(r, x)}{n} = \frac{1}{2},$$

for each r ($= 0$ or 2) for almost all $x \in C$.

We can, therefore, say that 'Almost all numbers belonging to Cantor set C are simply normal' (with respect to C). We denote the set of such numbers by $N_{3,2}$. (It should be noticed that none of Cantor points can be simply normal with reference to the whole interval [0, 1] and the scale 3, as none of the Cantor points contain the digit 1, as $x = 1/3 = 0/3 + 2/3^2 + 2/3^3 + \ldots$, and so on.)

The question now arises, whether the other two properties mentioned above hold good for the Cantor set as well: That is

(i) Is the set $N_{3,2}$ ($=$ the set of simply normal numbers of Cantor set C, as defined above) of first category with respect to C?

(ii) Is it true that except for a set of first category (with respect to $\check{C}$) including $N_{3,2}$, for other points $x \in C$, which form a residual set (with respect to C),

$$\left\{\frac{N_n(r, x)}{n}\right\}_n' = [0, 1] \quad \text{for} \quad r = 0, 2?$$

Since C is mapped onto [0, 1], that the answers to both the above questions are in the affirmative may be conjectured from Tibor Šalát's Theorem [13]:

For all $x \in (0, 1]$,

$$\left[x = \sum_{k=1}^{\infty} \frac{\varepsilon_k(x)}{2^k}, \quad \varepsilon_k(x) = 0, 1\right] \left\{\frac{N_n(r, x)}{n}\right\}_n' = [0, 1],$$

with the exception of a set of the first category, for each r ($=0, 1$).

We give below a formal theorem:

Theorem 2. *For all $x \in C$, with the exception of points of a set of the first category in C*

$$\left\{\frac{N_n(r, x)}{n}\right\}_n' = [0, 1], \quad (r = 0, 2)$$

holds.

Proof. The proof of this theorem follows as a Corollary to the following theorem of P. Kostyrko [10]:

Let

$$a_n > 0, \quad A = \sum_{n=1}^{\infty} a_n < +\infty, \, a_n > R_n = \sum_{k=1}^{\infty} a_{n+k}, \quad (n = 1, 2, \ldots).$$

Let W denote the set of all numbers x of the form $x = \sum_{n=1}^{\infty} \varepsilon_n a_n$, where $\varepsilon_n = 1$ or -1 $(n = 1, 2, \ldots)$. Let $f(n, x)$ denote the number of k's, $k \leqq n$, for which $\varepsilon_k = 1$. Then for all $x \in W$ with the exception of points of a set of the first category we have,

$$\left\{\frac{f(n, x)}{n}\right\}_n' = [0, 1].$$

If we now put $a_n = 1/3^n$ $(n = 1, 2, \ldots)$, the conditions $a_n > 0$, $A = \sum a_n$ and $a_n > R_n (=1/2.3^n)$ are all satisfied. In view of the fact that the Cantor set C is obtained by a translation of $W (C = W + A = W + 1/2$, since $A = \sum(1/3^n) = 1/2)$, the above theorem follows from P. Kostyrko's result [10].

Theorem 3. *Almost all points of the Cantor set* C *have each an asymptotic density* $\frac{1}{2}$ *in the Cantor series*

$$\frac{2}{3} + \frac{2}{3^2} + \ldots + \frac{2}{3^k} + \ldots$$

Proof. Let x be a point of $(0, 1]$ given by

$$x = \sum_{k=1}^{\infty} \frac{\varepsilon_k(x)}{2^k} \tag{A}$$

where $\varepsilon_k(x) = 0$ or 1 and $\varepsilon_k(x) = 1$, for infinitely many k's.

We have correspondingly the Cantor point

$$(x) = \frac{2\varepsilon_1(x)}{3} + \frac{2\varepsilon_2(x)}{3^2} + \ldots + \frac{2\varepsilon_k(x)}{3^k} + \ldots \tag{B}$$

which is a subseries of $\sum_{k=1}^{\infty} (2/3^k)$.

Now, number of twos in the first n terms of (B) in the right hand side is the same as

$$\sum_{k=1}^{n} \varepsilon_k(x) = p(n, x) = n_b, \quad (b = 1).$$

Hence

$$\lim_{n \to \infty} \frac{p(n, x)}{n} = \lim_{n \to \infty} \frac{n_b}{n}.$$

Since by Theorem 148 page 125 [9], $\lim_{n\to\infty}(n_b/n) = \frac{1}{2}$ $(b = 1, 0)$, for almost all $x \in (0, 1]$, it follows that $\lim_{n\to\infty}(p(n, x)/n) = \frac{1}{2}$, for almost all $(x) \in \boldsymbol{C}$. Hence the theorem.

We know from Randolph's Theorem [11] that every point $\in [0, 1]$ lies midway between a pair of Cantor points. BOSE MAJUMDER [See [6]] gave an alternative proof of this theorem. He further showed that almost all points of $[0, 1]$ are each midway between a continuum number of pairs of Cantor points [6].

We now prove the following

Theorem 4. *Each point λ of $(0, 1)$ is the midpoint of a unique pair of Cantor points if and only if λ itself is a Cantor point.*

Proof. It has already been seen [6] that, taking

$$0 \leqq \tfrac{1}{2}(d + 1) = \lambda = \sum_{i=1}^{\infty} \frac{\delta_i}{3^i}, \quad \delta_i = \begin{Bmatrix} 0 \\ 1 \\ 2 \end{Bmatrix}, \quad \text{if} \quad d \in [-1, 1],$$

we get

$$\frac{d}{2} = \lambda - \tfrac{1}{2} = \sum_{i=1}^{\infty} \frac{v_i}{3^i}, \quad v_i = \begin{Bmatrix} -1 \\ 0 \\ 1 \end{Bmatrix}.$$

Generally this representation is unique. But if $\frac{1}{2}d$ (and hence $\lambda - \frac{1}{2}$) has more than one such representation, then there are only two such representations and $\frac{1}{2}d$ (and hence $\lambda - \frac{1}{2}$) is given by,

$$\frac{d}{2} = \begin{cases} \cdot v_1 v_2 \ldots v_{k-1}(-1)\,111 \ldots \\ \cdot v_1 v_2 \ldots v_{k-1}(0)\,(-1)\,(-1)\,(-1) \ldots \end{cases}$$

or else by

$$\frac{d}{2} = \begin{cases} \cdot v_1 v_2 \ldots v_{k-1}(0)\,(1)\,(1)\,(1) \ldots \\ \cdot v_1 v_2 \ldots v_{k-1}(1)\,(-1)\,(-1)\,(-1) \ldots \end{cases}, \quad v_i = \begin{Bmatrix} -1 \\ 0 \\ 1 \end{Bmatrix}.$$

Now since

$$d = \sum_{i=1}^{\infty} \frac{2v_i}{3^i} = \sum_{i=1}^{\infty} \frac{2(\beta_i - \alpha_i)}{3^i} = \sum_{i=1}^{\infty} \frac{2\beta_i}{3^i} - \sum_{i=1}^{\infty} \frac{2\alpha_i}{3^i} = y - x,$$

where

$$y \in \boldsymbol{C}, \quad x \in \boldsymbol{C}.$$

By choosing

$$\alpha_i = 1\,, \quad \beta_i = 0 \quad \text{if} \quad v_i = -1$$
$$\alpha_i = 0\,, \quad \beta_i = 1 \quad \text{if} \quad v_i = 1$$

and either

$$\begin{cases}\alpha_1 = 0\\ \beta_i = 0\end{cases} \quad \text{or} \quad \begin{cases}\alpha_i = 1\\ \beta_i = 1\end{cases} \quad \text{if} \quad v_i = 0\,.$$

Hence $d = \sum_{i=1}^{\infty} (2v_i/3^i)$ is uniquely representable as $d = y - x$, $y \in C$, $x \in C$, if and only if no v_i is a zero, i.e. if and only if no δ_i is an 1, that is if and only if $\lambda\,(= (d + 1)/2)$ is a Cantor point. And in this case $y - x = d$ or $y - x = 2\lambda - 1$ or $2\lambda = y + (1 - x)$ or $2\lambda = y + x'$ where $y \in C$, $x' \in C$ (as the Cantor set C is symmetrical).

Hence the theorem.

Corollary. *Each Cantor point is the arithmetic mean of a unique pair of Cantor points.*

We know that the set N_3 of simply normal numbers in $[0, 1]$ in the scale 3 has the measure 1 [9] and also the set T_c of numbers $d \in [0, 1]$, each being the difference of continuum number c of pairs of elements of the Cantor set C has the measure 1 [See Boas [1] and Bose Majumder [5]].

Hence the set $E = N_3 \cap T_c$ is also of measure 1 [See Bose Majumder and Das Gupta [7]]. We thus have the theorem:

Theorem 5. *Excepting possibly for a set of measure zero, every point in* $[0, 1]$ *which is expressible as the difference of a pair of Cantor points in continuum number of ways is necessarily a simply normal number in the scale* 3 *and vice versa.*

Note 1. That the two sets are not identical can be seen from the fact that there exists $d \in [0, 1]$ which belongs to T_c but does not belong to N_3. For instance, let $d = \cdot\delta$ (scale 3), where δ is a complex containing a finite number of zeros and twos and thus ending with a 2. This represents the right hand end point of a contiguous interval of the Cantor set C. As this representation of d does not contain any 1, it follows that this can not be a simply normal number. But it is known that [See [2], [3]] this d can be expressed as the difference of a pair of Cantor points in continuum number of ways. Hence $\cdot\delta \in T_c$, but $\cdot\delta \bar{\in} N_3$.

Note 2. Though T_c and N_3 are each of measure 1, it is interesting to note that T_c is a residual set [See [4]], but N_3 is a set of the first category [See [13]].

In conclusion, I offer my sincere gratitude to Dr. N. C. Bose Majumder of Calcutta University for his kind help and guidance in the preparation of the paper. I am also thankful to the Reviewer for his kind suggestions to improve this paper.