数理解析研究所講究録
第 496 巻 1983 年 1-11

NORMAL NUMBERS AND DIMENSION

KENJI NAGASAKA

SHINSYU UNIVERSITY

1. INTRODUCTION.

Arithmetical sequences are, in general, infinite sequences of positive integers, of symbols on finite alphabet, or of real (complex) numbers. Once an arithmetical sequence is exposed, several problems will arise: How to generate the sequence? What is the essential property of the given sequence? How to characterise the sequence? How is the amount of sequences of given type? If we know quite well the structure of the sequence, we can reply to all these problems easily.

Even in high-school mathematics syllabus, the simplest cases are already given, such as, arithmetical progression, geometrical series, linear recurrence sequences of second order, etc... But distribution properties or number-theoretical properties, like divisibility, are on question, then we are soon confronted with certain difficulty.

By using Fourier analysis (harmonic analysis), we cannot say too much on number-theoretical properties, but we are able to give a fairly good discription of distribution properties of arithmetical sequences, which indicates the degree of complexity of them.

Let us start with too extremes with respect to distribution properties. Periodic sequences have extremely simple structure and we can conclude almost all questions on them. Starting from the periodicity several attempts have been made toward more complicated sequences.

Random sequences are supposed to be another extreme. The notion of random sequences is in itself contradictory. It is always possible to clarify an aspect or several aspects of randomness, but impossible to explain every property of the notion of randomness, to which we won't enter in this symposium .

Between two extremes, we are interested in constructing steps which characterize the degree of randomness in some sense. Several methods and their results shall be exposed in the sequel in this Proceedings (Kokyu-Roku) . In this report I will explain normal numbers, one extreme, and dimension as a tool of steps between two extremes.

2. NORMAL NUMBERS.

In 1909 Emile Borel defined normal numbers. At that time modern probability theory based on Lebesgue measure theory did not exist. Borel kept his style to write down his papers even after the famous Kolmogorov's memorial work "Grundlagen der Wahrscheinlichkeitstheorie" . Anyway Borel is one of the founder of probability theory. In his paper "Les probabilités dénombrables et leurs applications arithmétiques" , he distiguished probabilities into three types: discontinuous probabilities, denumerable probabilities and continuous (or geometrical) probabilities, and he discussed the second case. His discussion was essentially Borel-Catelli's lemmas.

Secondly he considered the decimal fractions, that is $x \in I_0 = [0,1)$ and x is represented by

$$x = \sum_{n=1}^{\infty} \frac{a_n(x)}{10^n} \quad \text{or} \quad x = \sum_{n=1}^{\infty} \frac{b_n(x)}{q^n}$$

where q is an integer superior to 1 (expansion to base q).

He supposed on decimal fraction as following:

i) the digits $\{a_n(x)\}$ $n = 1,2,\ldots$ (or $\{b_n(x)\}$) are independent

ii) each digit takes one of the values 0, 1,..., 9 (or q-1) with equal probability,

and he studied with his supposition the probability for which a decimal fraction is contained in a given interval in $I_0 = [0,1)$ and concluded that this probability is equal to the length of the given interval.

Then he took his attention to the value of a digit, say 3, and the probability for which a digit takes the value 3 is, under his supposition, 1/10 and $\{a_n(x)\}$ $n = 1,2,...$ are independent hence this is the divergent case of Borel-Cantelli's lemma. The conclusion is that the value 3 appears infinitely many times with probability one.

This led him to define normal numbers. Let us fix the notation. A real number x is represented in q-adic expansion where q is an integer superior to 1 and fixed (we call it base):

$$x = \sum_{n=-\infty}^{\infty} \frac{a_n(x)}{q^n} \qquad 0 \leq a_n(x) \leq q-1$$

only finite numbers of terms $\{a_{-1}(x), a_{-2}(x), ...\}$ are nonnull.

$$\{x\} = x - [x] = \sum_{n=1}^{\infty} \frac{a_n(x)}{q^n},$$ the fractional part of x.

Let us define

$$A_N(j;x) = \#\{ n \le N ; a_n(x) = j \}$$

and

$$A_N(B_k;x) = \#\{ n \le N ; a_n(x) = b_1, a_{n+1}(x) = b_2, \ldots a_{n+k-1}(x) = b_k \},$$

where $B_k = b_1 b_2 \ldots b_k$ is a block on a finite alphabet $\{ 0, 1, \ldots q-1 \}$.

DEFINITION. A real number x is said to be "simply normal to base q" if, for any integer j, $0 \le j \le q-1$,

$$\lim_{N\to\infty} \frac{1}{N} A_N(j;\{x\}) = 1/q .$$

Borel stated that almost all real numbers are simply normal to base q.

DEFINITION I. A real number x is said to be "normal to base q" if, for any positive integer k and for any block B_k,

$$\lim_{N\to\infty} \frac{1}{N} A_N(B_k;\{x\}) = 1/q^k .$$

DEFINITION II. A real number x is said to be "normal to base q" if all of the numbers $x, qx, q^2x, \ldots q^kx, \ldots$ are simply normal to bases $q, q^2, q^3, \ldots$

Borel's original definition was the second one and he stated the first definition as a property. The first problem to solve was equivalence between these two definitions.

In 1949, D.D.Wall tried this problem in his Ph.D. Thesis in Berkeley without success but he obtained other necessary and sufficient conditions on normality. In 1951 Niven and Zuckerman proved this equivalence. Their proof was simplified by Cassels in 1953. Pillai got the following theorem in 1940:

THEOREM (PILLAI). A real number x is said to be normal to base q if x is simply normal to the bases $q, q^2, q^3, \ldots$ Using the result of Niven-Zuckerman, the proof of Pillai's theorem was simplified by Maxfield in 1952. Long refined this result replacing the bases $q, q^2, q^3, \ldots$ by $q^{m_1}, q^{m_2}, q^{m_3}, \ldots,$ where $1 \leq m_1 < m_2 < m_3 \ldots$ an infinite sequence of positive integers, and can't by any finite m_i's.

Afterwards all of these necessary and sufficient conditions were proved using ergodic arguement by Christian Batut and myself. Michel Mendès France gave another necessary and sufficient condition on normal numbers:

THEOREM (MENDES FRANCE). Let $\phi_n(x)$ $n = 0,1,2,\ldots$ be the Rademacher functions. Suppose that x is not dyadic rational. Then x is normal to the base 2 iff

$$\lim_{N\to\infty} \frac{1}{N}\sum_{n=0}^{N-1} \phi_{n+k_1}(x) \cdots \phi_{n+k_s}(x) = 0$$

holds for all $s \geq 1$ and all distinct nonnegative integers $k_1, \ldots k_s$.

This result was the first his publication (1), with which he entered into the society of working mathematicians and since then he is quite productive (see the list of his publications).

Now we discuss on the explicit construction of normal numbers. Sierpinski 1917 and Lebesgue 1917 gave a partially explicit construction of normal numbers. A simple example is due to Champernowne in 1933 as following: 0.123456789101112131415..., simply successive arrangement of every positive integer and normal to the base 10. Shiokawa and Uchiyama obtained neat results on dyadic Champernowne numbers in 1975.

Let p_n be the n-th prime, $0.p_1p_2\ldots p_n\ldots$ is normal to the base 10 proved by Copeland-Erdös, which led a general construction of normal numbers to the base 10, that is, the successive arrangement of $[n^\theta]$ $\theta>1$ ($n = 1,2,\ldots$). The best estimation of the sum of digits of this integer sequence $[n^\theta]$ $\theta > 1$ was obtained by Shiokawa. Champernowne numbers are transcendental, which was proved by K.Mahler.

It is not known that number-theoretically interesting irrationals, such as e, π, $\sqrt{2}$, log 2 are normal or not. An answer related to this problem was given by Kamae and Mendès France et al. (26) in 1977 during Kamae's stay in Bordeaux but this is not the direct answer to the original normality of algebraic numbers.

By the way the set of all simply normal numbers, normal numbers are of full Lebesgue measure but these two sets are of the first category (see papers of Salat, Schweiger and mine).

3. DIMENSION.

The concept of dimension has been considered as a obvious notion before Cantor's one to one correspondence between the unit interval and a square in 1877. Peano constructed so-called Peano Curve which maps continuously the unit interval to the whole of a square in 1890. This Peano Curve is not a topologocally homeomorphisme but these results led us to define rigorously the concept of dimension.

Poincaré was probably the first to point out the necessity of the mathematical definition of dimension in 1912 and defined the dimension inductively by using intuitive geometrical concepts. Brower followed on the Poincaré's line and Urysohn and Menger established the concept of dimension which is topologically invariant in any compact metric space. There are many subsequent results on dimension in an abstract topological space, such as Hurewicz, Tumarkin, Fréchet, Kunugi, Morita, Nagami and so on.

The dimension stated just now is topological invariant, hence called topological dimension. But in order to classify the set of nonnormal numbers, the topological dimension is not at all a good device. What we need is fractional dimension taking non-integral values or Hausdorff dimension. In 1919 Hausdorff introduced the outer measure and his dimension and calculated the Hausdorff dimension of the ternary Cantor set which is equal to log 2/log 3 .

Every linear set of positive Lebesgue measure has Hausdorff dimension one. Further more generally every set of n-dimensional positive Lebesgue measure has Hausdorff dimension n. Every subset of n-dimensional Euclidean space is of Hausdorff dimension less than or equal to n .

The set of all normal numbers is a linear set of full Lebesgue measure, then its Hausdorff dimension is equal to one. The set of nonnormal numbers is henceforth a null set but we can classify the subsets of nonnormal numbers by their digit properties by means of Hausdorff dimension.

The set of all numbers which are not simply normal is a null set but of Hausdorff dimension one. This was proved by Eggleston in 1949 and demonstrated by using A.Beyer's result by Mendès France (3) in 1963 and by Billingsley in 1965 with his powerful technique to calculate the Hausdorff dimension. The set of simply normal numbers which are not normal is of measure zero but of Hausdorff dimension one, proved by myself in 1971.

Denote the set of all normal numbers to the base r by $B(r)$. If we translate $B(r)$ by adding some integer, but it still remains the same set $B(r)$, because the normality depends only on the decimal expansion aftter the point and not on the integral part. If we replace adding integer by a rational, the $B(r)$ translated still remains the same obviously. But the translation by the multiplication of non-zero rational $\lambda \in Q - \{0\}$ is not so evident. Maxfield proved that if $x \in B(r)$ and $\lambda \in Q - \{0\}$, $\mu \in Q$

then $\lambda x + \mu \varepsilon B(r)$ in 1953. Batut and myself in 1980 gave an extremely elementary proof of this result without any knowledge on Fourier analysis and got also an ergodic proof.

A big problem on normal numbers is the change of base. Cassels' paper said that Steinhauss posed the following problem in the New Scottish Book as the problem 144, which we can't find in the new edition by Birkhäuser 1981: How far the property of being normal with respect to different bases is independent?

J.W.S.Cassels and W.M.Schmidt replied to this qestion independently. If two positive integers r and s are algebraically independent, that is $\log r/\log s \varepsilon Q$, then $B(r) \neq B(s)$. The Hausdorff dimension of the difference $B(r) - B(s)$ is equal to one proved by myself in 1978. Related problems to base change are posed by Mendès France in Problems Session.

Let P be the set of real polynomials and let $E(P)$ be the set of real numbers whose n-th binary digit from a certain point on is 0 or 1 according as $[\phi(n)]$ is even or odd for some $\phi \varepsilon P$. Mendès France (4) proved that no number in $E(P)$ is normal to the base 2 and that $E(P)$ has Hausdorff dimension zero. This results led him to study the spectrum of arithmetical sequences; a part of Fourier analysis on arithmetical sequences.