Russian Math. Surveys **60**:6 1121–1146

Uspekhi Mat. Nauk **60**:6 115–138 DOI 10.1070/RM2005v060n06ABEH004284

Weighted averages, uniform distribution, and strict ergodicity

V. V. Kozlov

Abstract. A circle of problems related to the application of the Riesz and Voronoi summation methods in ergodic theory, number theory, and probability theory is considered. The first digit paradox is discussed, strengthenings of the classical result of Weyl on the uniform distribution of the fractional parts of the values of a polynomial are indicated, and the possibility of sharpening the Birkhoff–Khinchin ergodic theorem is considered. In conclusion, some unsolved problems are listed.

Contents

§ 1. First digit paradox

Let us begin with a discussion of the well-known problem on the frequency of occurrence of a digit g ($1 \leqslant g \leqslant 9$) in the sequence of first digits of the powers of 2:

$$
\begin{aligned}
&2,\ 4,\ 8,\ 1,\ 3,\ 6,\ 1,\ 2,\ 5,\ 1,\ 2,\ 4,\ 8,\ 1,\ 3,\ 6,\ 1,\ 2,\ 5,\ 1,\\
&2,\ 4,\ 8,\ 1,\ 3,\ 6,\ 1,\ 2,\ 5,\ 1,\ 2,\ 4,\ 8,\ 1,\ 3,\ 6,\ 1,\ 2,\ 5,\ 1,\\
&2,\ 4,\ 8,\ 1,\ 3,\ 7,\ \dots
\end{aligned}
\tag{1.1}
$$

We introduce the number sequence $x_n(g)$ (or briefly, x_n) which is equal to 1 if the decimal expansion of 2^n begins with the digit g and to 0 otherwise. The frequency of occurrence of the digit g is defined as the value

$$\lim_{n\to\infty} \frac{\nu_g(n)}{n},$$

AMS 2000 *Mathematics Subject Classification.* Primary 40G05, 11K06, 37A30; Secondary 60F15.

where $\nu_g(n) = \sum_1^n x_k$. As is well known, this limit exists and is equal to

$$\lg \frac{1+g}{g}.$$

In particular, 7 occurs in the sequence (1.1) more frequently than 8.

This result can be stated in another way,

$$x_n \to \lg \frac{1+g}{g} \quad (C), \tag{1.2}$$

where C stands for Cesàro convergence.

Let us recall the proof of the formula (1.2). Below we use the idea of this proof in another situation. It is clear that the decimal representation of a number 2^n begins with the digit g if

$$g \cdot 10^k \leqslant 2^n < (g+1) \cdot 10^k$$

for some integer $k \geqslant 0$. Taking the common logarithms in these inequalities and passing to the fractional parts, we obtain the inequalities

$$\lg g \leqslant \{n \lg 2\} < \lg(g+1). \tag{1.3}$$

Since the number $\lg 2$ is irrational, the sequence of fractional parts $\{n \lg 2\}$ is uniformly distributed on the interval $[0,1]$ in the sense of Weyl's classical definition. But this just means that, in the mean, the frequency of occurrence of the numbers $\{n \lg 2\}$ in the interval $[\lg g, \lg(g+1))$ is equal to the length of this interval.

We now simplify the problem by replacing 2^n by n. The question is: What is the frequency with which the positive integers begin with the digit g? In this case the inequality (1.3) must be replaced by the inequality

$$\lg g \leqslant \{\lg n\} < \lg(g+1).$$

As is well known, although the sequence of fractional parts of the logarithms is dense in the unit interval, it is not uniformly distributed. This result was established by J. Franel already in 1917. Hence, in this case the ratio $\nu_g(n)/n$ (the mean frequency of occurrence of the digit g) has no limit at all as $n \to \infty$.

Let us again take the number sequence x_n which is equal to 1 if the decimal expansion of n begins with the digit g and to 0 otherwise.

Theorem 1.

$$x_n \to \lg \frac{1+g}{g} \quad (R, 1/n).$$

Here $(R, 1/n)$ stands for the logarithmic summation method, that is, $s_n \to s$ $(R, 1/n)$ if

$$\lim_{n\to\infty} \frac{s_1 + s_2/2 + \dots + s_n/n}{1 + 1/2 + \dots + 1/n} = \lim_{n\to\infty} \frac{1}{\ln n} \sum_{k=1}^n \frac{s_k}{k} = s.$$

This method *includes* the Cesàro method: if $s_n \to s$ (C), then $s_n \to s$ $(R, 1/n)$.

Theorem 1 was proved by Duncan [1]. This theorem was preceded by an earlier result of Flehinger [2] who used iterations of the Cesàro method instead of the Riesz means. It should be noted that Theorem 1 is a quite special case of the results of an earlier paper [3] concerning properties of the uniform distribution when the Cesàro method is replaced by the Riesz summation methods (we speak of this replacement in § 2).

Theorem 1 justifies the first digit paradox: contrary to our expectations, a positive integer taken at random begins with 1 almost seven times more often than with 9. This paradox is usually associated with the name of F. Benford, an American physicist, who noticed once that the edges of the pages in a book containing a detailed table of logarithms were most soiled at the beginning of the book [4]. This means that people most frequently seek logarithms of numbers beginning with 1 and most infrequently logarithms of numbers beginning with 9. V. I. Arnol'd also noticed that the distribution of the first digits in tables of population and areas of countries of the world also satisfies Benford's law (see the comments in the book [5]).

It should be noted that Benford's law was discovered and published sixty years earlier by the American astronomer Simon Newcomb [6]. Newcomb presented the following table of frequencies of occurrence of diverse digits at the first and second place in the decimal representation of positive integers.

digit		first digit	second digit
0	...	...	0.1197
1	...	0.3010	0.1139
2	...	0.1761	0.1088
3	...	0.1249	0.1043
4	...	0.0969	0.1003
5	...	0.0792	0.0967
6	...	0.0669	0.0934
7	...	0.0580	0.0904
8	...	0.0512	0.0876
9	...	0.0458	0.0850

Newcomb *computed* these frequencies using the assumption that the sequence $\{\lg n\}$ is uniformly distributed on the interval $[0,1]$. As we saw above, the very definition of a uniformly distributed sequence needs to be made precise. Newcomb gave no exact definitions. However, one must keep in mind that his pioneering work was published 35 years earlier than the classical work of Hermann Weyl on the uniform distribution modulo 1.

It is clear from Newcomb's table that the second digits in the decimal representation of positive integers are also distributed non-uniformly: 0 occurs 1.4 times more often than 9. The frequency of occurrence of a digit g $(0 \leqslant g \leqslant 9)$ at the second place is computed in the same way. The decimal expansions of these numbers are of the form

$$1g\ldots,\ 2g\ldots,\ \ldots,\ 9g\ldots.$$

One must compute the frequencies of occurrence of these numbers (for instance, in the sense of logarithmic convergence) and then add these frequencies.

The frequency of occurrence of the numbers with decimal representation of the form $kg\dots$ is equal to

$$\lg\frac{10k+g+1}{10k+g}\,.$$

Hence, the frequencies of occurrence of the digits $0,1,\dots,9$ at the second place are equal to

$$\lg\frac{11}{10}\cdot\frac{21}{20}\cdots\frac{91}{90},\dots,\lg\frac{20}{19}\cdot\frac{30}{29}\cdots\frac{100}{99}\,.$$

The problem of computing the frequencies of occurrence of the digits $0,1,\dots,9$ at the third place can be solved in a similar way:

$$\lg\frac{101}{100}\cdot\frac{111}{110}\cdot\frac{121}{120}\cdots\frac{991}{990},\dots,\lg\frac{110}{109}\cdot\frac{120}{119}\cdot\frac{130}{129}\cdots\frac{1000}{999}\,.$$

These numbers differ from one another by even less.

Theorem 2. *The frequency of occurrence of any digit at the nth place tends to* $1/10$ *as* $n\to\infty$.

An assertion having a similar meaning was noted earlier in [7]. Let $D(n)$ be the number of all digits of all positive integers not exceeding n and let $D_g(n)$ be the number of occurrences of the digit g in the decimal representation of all these numbers. As was shown in [7],

$$D_g(n)/D(n)\to 1/10. \tag{1.4}$$

Formally, Theorem 2 does not depend on this result, because the frequency of occurrence of a digit g is defined in (1.4) by using Cesàro convergence, whereas Theorem 2 deals with frequencies whose definition uses stronger summation methods.

The relation (1.4) was in fact known before the paper [7]. For example, for binary expansions this assertion is contained in the textbook [8] as an exercise in the appendix to Chap. V.

Let us prove Theorem 2. To be definite, we consider the frequency of occurrence of 0,

$$\lg\biggl(1+\frac{1}{10^n}\biggr)\biggl(1+\frac{1}{10^n+10}\biggr)\cdots\biggl(1+\frac{1}{10^{n+1}-10}\biggr). \tag{1.5}$$

For convenience of notation, we add to (1.5) the term

$$\lg\biggl(1+\frac{1}{10^{n+1}}\biggr), \tag{1.6}$$

which tends to 0 exponentially rapidly (like 10^{-n}) as $n\to\infty$. As a result, the sum of (1.5) and (1.6) becomes

$$s_n=\sum_{j=0}^{9\cdot 10^{n-1}}\lg\biggl(1+\frac{1}{10^n+10j}\biggr).$$

We use the following elementary inequality: if $x > 0$, then

$$0 < x - \ln(1+x) < x^2/2.$$

Hence,

$$0 < u_n - s_n < v_n, \tag{1.7}$$

where

$$u_n = \frac{1}{10\ln 10}\sum_{j=0}^{9\cdot 10^{n-1}} \frac{1}{10^{n-1}+j}, \qquad v_n = \frac{1}{2\ln 10}\sum_{j=0}^{9\cdot 10^{n-1}} \frac{1}{(10^n+10j)^2}.$$

Further,

$$\int_0^{9\cdot 10^{n-1}} \frac{dx}{10^{n-1}+x} < \sum_{j=0}^{9\cdot 10^{n-1}} \frac{1}{10^{n-1}+j} < \int_0^{9\cdot 10^{n-1}} \frac{dx}{x-1+10^{n-1}}.$$

The integrals on the left- and right-hand side are equal to

$$\ln 10 \quad \text{and} \quad \ln\frac{10^n-1}{10^{n-1}-1} = \ln\biggl(10 + \frac{9}{10^{n-1}-1}\biggr),$$

respectively. Hence, $u_n \to 1/10$ as $n \to \infty$.

Since

$$\sum_{j=0}^{9\cdot 10^{n-1}} \frac{1}{(10^n+10j)^2} < \frac{9\cdot 10^{n-1}+1}{10^{2n}} < \frac{1}{10^n},$$

it follows that $v_n \to 0$ as $n \to \infty$. In this case it follows from (1.7) that $s_n \to 1/10$, as was to be proved.

It follows from the proof that the difference between (1.5) and 1/10 decreases like 10^{-n} (as n increases). Thus, the nth digit paradox is practically imperceptible for large n.

For a discussion of diverse aspects of the first digit paradox, see [7] and [9].

§ 2. Weighted averages and types of uniform distribution

The considerations in the previous section that are related to the replacement of the convergence of arithmetic means (Cesàro convergence) by other summation methods lead us to a natural generalization of uniformly distributed sequences.

Let x_n $(n \geqslant 1)$ be a sequence of points in the unit interval, let $L \subset [0,1]$ be an arbitrary interval of length l, and let f be the characteristic function of this interval. Let S be a linear and regular summation method.[1]

We say that a sequence of points x_n is *S-uniformly distributed modulo* 1 (or, briefly, is *S-u.d.*) if

$$f(x_n) \to \operatorname{meas} L = l \quad (S) \tag{2.1}$$

[1] *Russian Editors' note*: A summation method is said to be *regular* if it is applicable to any convergent series and, when applied to a convergent series, gives the result equal to the sum of the series.

for any interval L. If S is the Cesàro method, then we obtain Weyl's classical definition.

One can readily prove that the relation (2.1) holds for all Riemann integrable functions f, where one must of course replace the length l by the integral of f over the interval $0 \leqslant x \leqslant 1$. As in the classical case, this result admits the following converse assertion: if

$$e^{2\pi i m x_n} \to 0 \quad (S)$$

for any integer $m \neq 0$, then the sequence x_n is S-u.d. It is clear that every S-u.d. sequence is dense in the unit interval.

In fact, sequences uniformly distributed in a generalized sense were considered by Weyl himself in the classical paper "Über die Gleichverteilung von Zahlen mod. Eins" (see [5][2]). Weyl takes the summation method S to be the Riesz methods (R, p_n) whose weight coefficients p_n are either monotone decreasing with $\sum p_n = \infty$ or monotone increasing with

$$(n+1)p_n = O\biggl(\sum_1^n p_j\biggr). \tag{2.2}$$

Weyl proves the following assertion (see [5], Theorem 10, p. 74). Suppose that $P(x)$ is a polynomial with at least one irrational coefficient (aside from the constant term). Then the relation

$$f(P(n)) \to \int_0^1 f(x)\,dx \quad (R, p_n) \tag{2.3}$$

holds for any Riemann integrable function f under the above assumptions about the coefficients p_n.

For a linear form $P(x)$ and a monotone decreasing sequence p_n this assertion is also contained in Problem 173 of the book [10] by Pólya and Szegő.

However, these assertions contain nothing new as compared with the result on the uniform distribution modulo 1 of the sequence $\{P(n)\}$ (this is Theorem 9 in [5], p. 69). The fact is that the following property holds for $p_{n+1} \leqslant p_n$: if $s_n \to s$ (C), then $s_n \to s$ (R, p_n). This follows from the Cesàro theorem ([11], Theorem 14), established in 1888 (25 years before Weyl's paper). In other words, under the above assumptions the Riesz method *includes* the Cesàro method. On the other hand, if $p_{n+1} \geqslant p_n$, then, conversely, the Cesàro method always includes the Riesz method ([11], Theorem 14). Moreover, if the condition (2.2) holds, then these methods are equivalent. Thus, the limit relation (2.3) is meaningful only for Riesz methods such that the numbers p_n increase but the property (2.2) fails.

We note that if (2.2) holds, then p_n can be increasing like a power of n. On the other hand, if the p_n are increasing exponentially rapidly, then the Riesz method loses its power and becomes equivalent to ordinary convergence ([11], Theorem 15). Therefore, the real interest in this generalization involves cases in which, for example, $p_n \sim \exp n^\gamma$, where $0 < \gamma < 1$.

We indicate sufficient conditions for R-uniform distribution.

[2] *Russian Editors' note*: *Math. Ann.* **77** (1916), 313–352; reprinted in *Gesammelte Abhandlungen*, Band I, Springer-Verlag, Berlin 1968, pp. 563–599, and in *Selecta Hermann Weyl*, Birkhäuser, Basel 1956, pp. 111–147.

Theorem 3. *Let the continuously differentiable functions $f(x)$ and $g(x)$ $(x \geqslant 1)$ satisfy the following conditions*:

1) $f' \neq 0$ *and* $f'(x) \to 0$ *as* $x \to \infty$,
2) *the function* $g > 0$ *is either non-decreasing and*

$$\frac{g(x)}{\displaystyle\int_1^x g(t)\,dt} \to 0 \quad as \quad x \to \infty,$$

or non-increasing and

$$\int_1^\infty g(x)\,dx = \infty, \tag{2.4}$$

3) *the ratio* g/f' *is either non-increasing, or non-decreasing and*

$$\frac{g(x)}{f'(x)\displaystyle\int_1^x g(t)\,dt} \to 0 \quad as \quad x \to \infty.$$

Then the sequence $\{f(n)\}$ *is* $(R, g(n))$*-uniformly distributed.*

This theorem was proved in [3] under somewhat different assumptions ($f(x) \nearrow \infty$ and $f'(x) \searrow 0$). The above version of the assertion on R-uniform distribution was indicated in [12]. The proof of Theorem 3 uses the Euler–Maclaurin summation formula.

This theorem implies a series of useful corollaries.

Corollary 1. *Under the assumptions of Theorem* 3 *the sequence* $\{f(n)\}$ *is dense in the unit interval.*

Corollary 2 (Fejér theorem [10], [13]). *If* $f'(x)$ *tends monotonically to* 0 *and* $x|f'(x)| \to \infty$ *as* $x \to \infty$, *then the sequence* $\{f(n)\}$ *is Weyl uniformly distributed.*

Corollary 3. *Let* $f' > 0$, $f'(x) \searrow 0$, *and* $f(x) \to \infty$ *as* $x \to \infty$. *Then* $\{f(n)\}$ *is an* $(R, f'(n))$*-uniformly distributed sequence.*

One must set $g = f'$ and use Theorem 3. In particular, the sequence $\{c \ln n\}$, $c \neq 0$, is $(R, 1/n)$-u.d.

§ 3. Uniform distribution and Voronoi convergence

Along with Riesz convergence, we also use Voronoi convergence. Let $p_1 > 0$ and $p_n \geqslant 0$ again. For a sequence s_n $(n \geqslant 1)$ we set

$$u_n = \frac{p_n s_1 + p_{n-1} s_2 + \cdots + p_1 s_n}{p_1 + p_2 + \cdots + p_n}\,.$$

It is clear that if $s_i = s_1$, then $u_n = s_1$. If $u_n \to s$, then we write $s_n \to s \ \ (W, p_n)$. The theory of Voronoi summation is presented in detail in [11].

In the Western mathematical literature, Voronoi's method is attributed to Nörlund, who considered this method 18 years after the publication of G. F. Voronoi's note in 1901.

The regularity criterion for the W-method is as follows:

$$\frac{p_n}{p_1+\cdots+p_n}\to 0.$$

It turns out that any two regular Voronoi methods are *compatible*: if $s_n\to s$ (W) and $s_n\to s'$ (W'), then $s=s'$. We stress that the Riesz methods fail to have this important property.

If a W-method is regular and the sequence of weight coefficients p_n is non-increasing, then the method (W,p_n) includes the Cesàro method C. Conditions for the converse inclusion are of great interest. Let $p_1=1$ and $p_n>0$. If the W-method is regular, p_n is non-increasing, and

$$\frac{p_{n+1}}{p_n}\geqslant\frac{p_n}{p_{n-1}}$$

for all $n>2$, then C includes (W,p_n).

There is an analogue, noted in [12], of Theorem 3 for the uniform distribution in the sense of Voronoi means.

Theorem 4. *Let the continuously differentiable functions $f(x)$ and $g(x)$ $(x\geqslant 1)$ satisfy the following conditions*:

1) $f'\neq 0$ *and* $f'(x)\to 0$ *as* $x\to\infty$,
2) *the function g is positive and monotone*, *and*

$$\int_1^\infty g(x)\,dx=\infty,$$

$$\frac{g(x)}{\displaystyle\int_1^x g(t)\,dt}\to 0 \quad as \quad x\to\infty, \tag{3.1}$$

3) *the ratio $g(n-x)/f'(x)$ is non-decreasing or non-increasing on the interval* $1\leqslant x\leqslant n$,
4) $|f'(x)|\displaystyle\int_1^x g(t)\,dt\to\infty$ *as* $x\to\infty$.

Then the sequence $\{f(n)\}$ is $(W,g(n))$-u.d. on the unit interval.

The condition (3.1) is the regularity condition for the Voronoi method. Taking $g(x)=1$, we obtain the well-known Fejér theorem on Weyl's uniform distribution.

Corollary. *Let $f'>0$, let $f'(x)\searrow 0$, let $g(x)>0$ be non-increasing*, *and let*

$$f'(x)\int_1^x g(t)\,dt\to\infty \tag{3.2}$$

as $x\to\infty$. Then $\{f(n)\}$ is $(W,g(n))$-u.d.

For example, we can set $g(x)=1/x$. Then (3.2) becomes

$$f'(x)\ln x\to\infty.$$

In this case the sequence $\{f(n)\}$ is $(W,1/n)$-u.d.

The case in which the function g is increasing is of special interest. In this case the $(W, g(n))$-method includes the Cesàro method. However, here (as a rule) the condition 3) of Theorem 4 can fail. We indicate a possible modification of Theorem 4.

In what follows we assume that f is twice continuously differentiable. We write $\varphi(x) = 1/f'(x)$. Differentiating the product $\varphi(x)g(n-x)$ with respect to x and dividing the result by this product, we obtain

$$\frac{\varphi'(x)}{\varphi(x)} - \frac{g'(n-x)}{g(n-x)}. \tag{3.3}$$

Suppose that the functions φ'/φ and g'/g are *monotonically* decreasing to 0 as $x \to \infty$. In particular, we have $f''/f' \to 0$ monotonically. Then for sufficiently large n the function (3.3) has exactly one zero in the interval $[1, n]$, and we denote this zero by x_n.

Theorem 5 [14]. *Let the following conditions hold*:

1) $f' > 0$, $f'(x) \to 0$, *and* $f''(x)/f'(x)$ *tends monotonically to* 0 *as* $x \to \infty$,
2) $g > 0$, g *is monotonically increasing, and* $g'(x)/g(x) \to 0$ *monotonically as* $x \to \infty$,
3) $$\frac{g(n)}{f'(x_n)\displaystyle\int_1^n g(t)\,dt} \to 0, \qquad \frac{g(x_n)}{f'(n)\displaystyle\int_1^n g(t)\,dt} \to 0 \tag{3.4}$$
as $n \to \infty$.

Then the sequence $\{f(n)\}$ *is* $(W, g(n))$*-u.d.*

Corollary. *Let* $f' > 0$, *let the functions* $f'(x)$ *and* $\dfrac{f''(x)}{f'(x)}$ *tend monotonically to* 0 *as* $x \to \infty$, *and let*

$$\frac{f''(x)}{f'^2(x)} \to 0. \tag{3.5}$$

Then $\{f(n)\}$ *is* $(W, 1/f'(n))$*-u.d.*

One must set $g(x) = 1/f'(x)$. Then it is easy to show that all the conditions of Theorem 5 are satisfied in this case, and $x_n = (n-1)/2$. Here the conditions (3.4) follow from (3.5).

Example. We set $f(x) = \ln^\alpha x$, where $\alpha > 0$. The condition (3.5) holds only for $\alpha > 1$. In this case the sequence $\{\ln^\alpha n\}$ is W-u.d. for an appropriate choice of a regular W-method (for W one can take, for instance, the Cesàro method). However, this property fails for $\alpha = 1$.

Theorem 6 [14]. *Let* $a \geqslant 2$ *be an integer and let* W *be a regular Voronoi summation method. Then the sequence*

$$\left\{\frac{1}{2}\log_a n\right\}$$

is not W*-u.d.*

Very recently, G. A. Kalyabin used the method of [14] to prove that this result holds for any sequence $\{\alpha \ln n\}$, $\alpha \neq 0$.

§ 4. Uniform distribution on the torus

Let $\mathbb{T}^n$ be the n-dimensional torus with the angular coordinates $x = (x_1, \dots, x_n)$ varying $\operatorname{mod} 2\pi$. A *motion* on $\mathbb{T}^n$ is a continuous map $t \mapsto x(t)$, $t \in \mathbb{R}$. It is said to be (Weyl) *uniformly distributed* on $\mathbb{T}^n$ if

$$\lim_{T\to\infty} \frac{\nu_D(T)}{T} = \frac{\operatorname{meas} D}{\operatorname{meas} \mathbb{T}^n}$$

for any Jordan measurable domain $D \subset \mathbb{T}^n$, where $\nu_D(T)$ is the sum of the lengths of the intervals on $[0, T]$ on which $x(t) \in D$, and $\operatorname{meas} \mathbb{T}^n = (2\pi)^n$.

Weyl gave the following criterion for the uniform distribution:

$$\lim_{T\to\infty} \frac{1}{T} \int_0^T e^{i(m,x(t))}\, dt = 0$$

for any integral vector $m \neq 0$. He also noted the following result (Theorem 8 in [5], p. 69): if $x_1(t), \dots, x_n(t)$ are n polynomials in t such that

$$\sum m_j x_j(t) \neq \text{const}$$

for any integers m_j which do not vanish simultaneously, then the motion $x_j = x_j(t)$ $(1 \leqslant j \leqslant n)$ is uniformly distributed on $\mathbb{T}^n$.

Let $t \mapsto \lambda(t)$ be a positive continuous function with

$$\int_0^\infty \lambda(t)\, dt = \infty. \tag{4.1}$$

We say that $f(t) \to \overline{f}\ \ (R, \lambda)$ if

$$\int_0^T \lambda(t) f(t)\, dt \Big/ \int_0^T \lambda(t)\, dt \to \overline{f}$$

as $T \to \infty$. This is a continuous analogue of the Riesz summation method. We indicate the main properties of the (R, λ)-method without trying to be exhaustive and complete.

First, the (R, λ)-method is linear and regular. The latter property is an immediate consequence of l'Hôpital's rule.

Further, let $\mu(t) > 0$ be another function satisfying (4.1). We say that (R, λ) *includes* (R, μ) if the convergence $f(t) \to \overline{f}\ \ (R, \mu)$ implies $f(t) \to \overline{f}\ \ (R, \lambda)$. The following assertion gives a sufficient condition for the inclusion.

Theorem 7. *Let*

$$\lambda(\tau) \int_0^\tau \mu\, dt \Big/ \mu(\tau) \int_0^\tau \lambda\, dt \to 0 \quad \text{as} \quad \tau \to \infty \tag{4.2}$$

and let the function

$$\lambda(\tau) \int_0^\tau \mu\, dt \Big/ \mu(\tau) \tag{4.3}$$

be monotone for $\tau \geqslant \tau_0$. Then (R, λ) includes (R, μ).

Suppose that the functions λ and μ belong to the same Hardy field and are of order $+\infty$ with respect to t:

$$\frac{\ln \lambda(t)}{\ln t} \to +\infty \quad \text{and} \quad \frac{\ln \mu(t)}{\ln t} \to +\infty$$

as $t \to +\infty$. Then the condition (4.2) can be represented as follows:

$$\frac{\dot{\lambda}}{\lambda} \Big/ \frac{\dot{\mu}}{\mu} \to 0. \tag{4.4}$$

If λ and μ belong to the same Hardy field, then the condition that the function (4.3) is monotone for sufficiently large values of τ is certainly satisfied.

Remark. Let $(R, \lambda(n))$ and $(R, \mu(n))$ be two ordinary (discrete) Riesz methods and let $\sum \lambda(n) = \sum \mu(n) = \infty$. By the Cesàro theorem, if

$$\frac{\lambda(n+1)}{\lambda(n)} \leqslant \frac{\mu(n+1)}{\mu(n)}, \tag{4.5}$$

then $(R, \lambda(n))$ includes $(R, \mu(n))$. If one represents (4.5) in the equivalent form

$$\frac{\lambda(n+1) - \lambda(n)}{\lambda(n)} \leqslant \frac{\mu(n+1) - \mu(n)}{\mu(n)},$$

then it becomes clear (cf. (4.4)) that Theorem 7 is a continuous analogue of the Cesàro theorem.

Example. Let $R_\alpha = (R, \exp t^\alpha)$, $\alpha \geqslant 0$. It is clear that R_0 is the classical Cesàro method. By Theorem 7, R_α includes R_β if $\beta \geqslant \alpha$. As will be proved below, for $\alpha \geqslant 1$ the methods R_α are equivalent to ordinary convergence.

Theorem 8. *Suppose that*

$$\lambda(\tau) \Big/ \int_0^\tau \lambda \, dt \geqslant c = \text{const} > 0 \tag{4.6}$$

and both f and $\dot{f}$ are bounded. Then it follows from the condition $f(t) \to \overline{f}$ (R, λ) that $f(t) \to \overline{f}$ (in the usual sense).

For functions λ of infinite order with respect to t the condition (4.6) is equivalent to the condition

$$\dot{\lambda}/\lambda \geqslant c.$$

In particular, all the methods $R_\alpha = (R, \exp t^\alpha)$ with $\alpha \geqslant 1$ are equivalent to ordinary convergence.

In the well-known Wiener Tauberian theorems it is assumed that the function f is bounded and slowly oscillating. The latter property certainly holds if the derivative is bounded. For the proofs of Theorems 7 and 8, see [15].

A continuous motion $x\colon \mathbb{R} \to \mathbb{T}^n$ is said to be (R,λ)-*uniformly distributed* (briefly, (R,λ)-*u.d.*) if

$$\lim_{\tau\to\infty} \int_0^\tau \lambda(t) f(x(t))\,dt \Big/ \int_0^\tau \lambda(t)\,dt = \frac{1}{(2\pi)^n} \int_{\mathbb{T}^n} f(x)\,d^n(x)$$

for *any continuous* function $f\colon \mathbb{T}^n \to \mathbb{R}$. This relation remains valid for any Riemann integrable function f provided that the function $t \mapsto f(x(t))$ is Riemann integrable on any finite interval. For $\lambda(t) = 1$ we obtain Weyl's definition of uniform distribution.

To prove that a motion x is (R,λ)-u.d., it suffices to verify the condition

$$\int_0^\tau \lambda(t) e^{i(m,x(t))}\,dt \Big/ \int_0^\tau \lambda(t)\,dt \to 0$$

as $\tau \to \infty$ for any $m \in \mathbb{Z}^n \setminus \{0\}$ (generalized Weyl criterion).

Theorem 9. *Let $x_1(t), \dots, x_n(t)$ be n polynomials such that no non-trivial integral combination of them is constant. If*

$$\lambda(\tau) \Big/ \int_0^\tau \lambda(t)\,dt \to 0$$

and the functions λ and t belong to the same Hardy field, then the motion

$$x_j = x_j(t), \qquad 1 \leqslant j \leqslant n, \tag{4.7}$$

is (R,λ)-u.d. on $\mathbb{T}^n$.

The conditions of this assertion certainly hold if, for example, $\lambda = \exp t^\alpha$ for $\alpha < 1$. Thus, Theorem 9 in [15] is a sharpening of Weyl's result (Theorem 8 in [5], p. 69) on the uniform distribution of the motion (4.7) on $\mathbb{T}^n$ in the ordinary sense (for $\lambda = 1$).

Let us now discuss the problem of the distribution of logarithms. We assume first that $n = 1$. The motion $x = \ln t$, $t \geqslant a > 0$, is not Weyl u.d. on the circle. Indeed,

$$\frac{1}{\tau} \int_a^\tau e^{i\ln t}\,dt = \frac{e^{i\ln\tau}}{1+i} + \frac{\text{const}}{\tau}\,,$$

and the right-hand side oscillates and does not tend to 0 as $\tau \to \infty$. On the contrary, for any $\varepsilon > 0$ the motion $x = \ln^{1+\varepsilon} t$ is Weyl u.d. on the circle $\mathbb{T} = \{x \bmod 2\pi\}$.

On the other hand, since

$$\int_a^\tau \frac{e^{im\ln t}}{t}\,dt = \frac{e^{i\ln\tau}}{im} + \text{const} = O(1)$$

for any integer $m \neq 0$, it follows that the motion $x = \ln t$ is $(R, 1/t)$-u.d. Of course, the $(R, 1/t)$-method is stronger than the Cesàro method.

These remarks can be generalized. The following theorem holds.

Theorem 10. *If the function $t\lambda(t)$ is monotone for $t \geqslant t_0$ and*

$$\tau\lambda(\tau) \Big/ \int_0^\tau \lambda(t)\,dt \to 0,$$

then the motion $x_j = \omega_j \ln t$ with rationally independent frequencies $\omega_1, \dots, \omega_n$ is (R, λ)-u.d. on $\mathbb{T}^n$.

A similar assertion holds for iterated logarithms. For instance, the function $\ln\ln t$ is $(R, (t\ln t)^{-1})$-u.d. on the circle $\{x \bmod 2\pi\}$.

Let $f\colon \mathbb{T}^n \to \mathbb{R}$ be a continuous function with zero mean value and let $\omega_1, \dots, \omega_n$ be a family of constant frequencies (which is not necessarily non-resonant). Consider the integral

$$I(\tau, x_0) = \int_0^\tau f(\omega_1 t + x_1^0, \dots, \omega_n t + x_n^0)\,dt.$$

As was proved in [16], under these assumptions one can always find a family of initial phases $x_0 = (x_1^0, \dots, x_n^0)$, $f(x_0) = 0$, such that $I(\tau, x_0) \geqslant 0$ for any $\tau \in \mathbb{R}$. A similar conclusion holds for the inequality $I(\tau, x_0) \leqslant 0$, of course. This result admits a generalization [15].

Theorem 11. *Let $\lambda(t) > 0$ be a non-increasing function. Then under the above assumptions*

$$\int_0^\tau \lambda(t) f(\omega t + x_0)\,dt \geqslant 0 \quad (\leqslant 0)$$

for any τ and for some initial phases x_0 such that $f(x_0) = 0$.

Corollary 1. *Let f be a continuous function on $\mathbb{T}^n$ with zero mean and let $p \geqslant 1$. Then there is an $x_0 \in \mathbb{T}^n$ such that*

$$\int_0^\tau f(\omega t^p + x_0)\,dt \geqslant 0 \quad (\leqslant 0)$$

for any τ.

It suffices to make a change of variable by the formula $t^p = z$.

Corollary 2. *Let f be a continuous function with zero mean and let the product $e^z\lambda(e^z)$ be non-increasing. Then there is an $x_0 \in \mathbb{T}^n$ such that*

$$\int_1^\tau \lambda(t) f(\omega \ln t + x_0)\,dt \geqslant 0 \quad (\leqslant 0)$$

for all τ.

It suffices to pass to the new variable $z = \ln t$. In particular, the conclusion of Corollary 2 holds for $\lambda = 1/t$. If no condition is imposed on the function λ, then the sign need no longer be definite. For example, for any x_0 the integral

$$\int_1^\tau \cos(\ln t + x_0)\,dt$$

changes sign infinitely many times as $\tau \to \infty$.

We assume now that $f\colon \mathbb{T}^n \to \mathbb{R}$ is a non-constant *infinitely differentiable* function with zero mean value and the frequencies $\omega_1, \dots, \omega_n$ are rationally independent. As was proved in [17], if $f(x_0) \neq 0$, then the integral

$$\int_0^\tau f(\omega t + x_0)\, dt$$

changes sign infinitely many times as $\tau \to \infty$. For $n = 1$ this is obvious, and for $n = 2$ it is proved in [16].

Theorem 12. *Let $\lambda(t) > 0$ be a non-decreasing function. If all zeros of the function*

$$\tau \mapsto \int_0^\tau \lambda(t) f(\omega t + x_0)\, dt \tag{4.8}$$

are simple, then there are infinitely many of them.

If the function λ is decreasing, then the integral (4.8) can have only finitely many zeros. For a simple example one can take the classical Fresnel integral

$$\int_0^\tau \cos t^2\, dt. \tag{4.9}$$

By the change of variable $t^2 = z$ this integral is reduced to the form (4.8), where λ tends monotonically to 0. As is well known, the function (4.9) has a single simple zero $\tau = 0$.

§ 5. Strict ergodicity

Let M be a compact metric space and let T be a homeomorphism of M. By the Krylov–Bogolyubov theorem, T preserves some Borel measure μ on M. If a normalized ($\mu(M) = 1$) Borel measure *invariant* with respect to T is unique, then the continuous transformation T is said to be *strictly ergodic*.

It is clear that every strictly ergodic homeomorphism T is ergodic with respect to a unique invariant Borel measure μ. However, a continuous ergodic transformation need not be strictly ergodic. For example, any strictly ergodic transformation has no periodic points. Otherwise there is an additional invariant Borel measure concentrated on the periodic trajectory of the homeomorphism T.

We indicate a characteristic property of a strictly ergodic transformation (*Oxtoby theorem* [18]):

$$f(T^n x) \rightrightarrows \int_M f\, d\mu \quad (C) \tag{5.1}$$

for any continuous function $f\colon M \to \mathbb{R}$. The symbol $\rightrightarrows$ stands here for uniform convergence. Since T is invertible, the relation (5.1) remains valid for $n \to -\infty$.

Example. Let T be an orientation-preserving homeomorphism of the circle $M = \{x \bmod 2\pi\}$. It is clear that $Tx = x + f(x)$, where f is a continuous 2π-periodic function. As was shown by Poincaré, for any x we have

$$f(T^n x) \to 2\pi\lambda \quad (C). \tag{5.2}$$

The number λ is called the *rotation number* of the homeomorphism T. If T has no periodic points, then T is strictly ergodic (see, for instance, [19]).

The relation (5.1) can be extended to a broader class of functions. Let us say that a function $f\colon M \to \mathbb{R}$ is $\mathcal{R}$-*integrable* (with respect to a normalized Borel measure μ on M) if for any $\varepsilon > 0$ there are two continuous functions f_1 and f_2 such that

$$f_1(x) \leqslant f(x) \leqslant f_2(x) \quad \text{for any} \quad x \in M,$$
$$\int_M (f_2 - f_1)\,d\mu < \varepsilon. \tag{5.3}$$

It is clear that $\mathcal{R}$-integrable functions are bounded and all continuous functions are $\mathcal{R}$-integrable.

The integral of an $\mathcal{R}$-integrable function is defined in the following natural way. Consider a sequence ε_n tending to 0, and let $f_1^{(n)}$ and $f_2^{(n)}$ be corresponding sequences of continuous functions satisfying the conditions (5.3). One can readily show that the limits as $n \to \infty$ of the sequences of integrals of $f_1^{(n)}$ and $f_2^{(n)}$ with respect to μ exist and coincide; we call this common value the $\mathcal{R}$-*integral of* f with respect to the Borel measure μ and write

$$\int_M f\,d\mu.$$

It is easy to see that the $\mathcal{R}$-integral is well defined (the definition does not depend on the choice of the sequences ε_n, $f_1^{(n)}$, and $f_2^{(n)}$).

For example, let M be the multidimensional torus $\mathbb{T}^k$ and let μ be the standard measure on $\mathbb{T}^k$. In this case the class of $\mathcal{R}$-integrable functions coincides with the class of Riemann integrable functions.

One can somewhat generalize the Oxtoby theorem [20].

Theorem 13. *A transformation T is strictly ergodic if and only if*

$$f(T^n x) \rightrightarrows \int_M f\,d\mu \quad (C)$$

for any $\mathcal{R}$-integrable function f.

A subset $D \subset M$ is said to be $\mathcal{R}$-*measurable* if the characteristic function $f = 1_D$ of D is $\mathcal{R}$-integrable. We set

$$\operatorname{meas} D = \int_M 1_D\,d\mu.$$

It is clear that $\operatorname{meas} M = 1$ by the assumption that the measure μ is normalized.

For the function f in Theorem 1 one can take the characteristic function of *any* $\mathcal{R}$-measurable domain D. We introduce the sequence s_n $(n \geqslant 0)$ by the following rule: $s_k = 1$ if $x_k = T^k x \in D$ and $s_k = 0$ otherwise. Let $\nu(n) = \sum_0^{n-1} s_k$. In this case we get that for any strictly ergodic transformation

$$\lim_{n\to\infty} \frac{\nu(n)}{n} = \operatorname{meas} D.$$

Thus, the time spent in the domain D by the trajectory x_k, $k \geqslant 0$, of *any* point $x \in M$ is proportional in the mean to the measure of this domain. This is the *general* definition of a Weyl uniformly distributed sequence (see, for instance, [21]).

If the sequence s_k just introduced converges to $\operatorname{meas} D$ from the point of view of a linear regular method S for any choice of an $\mathcal{R}$-measurable domain D, then the sequence of points $x_k \in M$, $k \geqslant 0$, is naturally said to be *S-uniformly distributed* (briefly, *S-u.d.*).

Let us show, following [20], how to refine the Oxtoby theorem by using the Riesz and Voronoi summation methods which are *included* in the Cesàro method.

Theorem 14. *Let T be a strictly ergodic homeomorphism of a compact metric space M and let f be an $\mathcal{R}$-integrable function on M. If*

$$\frac{p_0 + |p_1 - p_0| + \dots + |p_n - p_{n-1}| + p_n}{p_0 + p_1 + \dots + p_n} \to 0, \tag{5.4}$$

then

$$f(T^n x) \rightrightarrows \int_M f\,d\mu \quad (R, p_n) \quad \text{and} \quad (W, p_n). \tag{5.5}$$

The condition (5.4) *certainly holds if*

a) $p_{n+1} \leqslant p_n$ *and* $\sum p_n = \infty$, *or*
b) $p_{n+1} \geqslant p_n$ *and*

$$\frac{p_n}{p_0 + \dots + p_n} \to 0. \tag{5.6}$$

Example. Let T be an orientation-preserving homeomorphism of the circle and let T have no periodic points. In this case, the limit relation (5.2) can be sharpened by replacing the Cesàro method by arbitrarily weak Riesz and Voronoi methods. In [22] this fact was noted in the general case in which the homeomorphism T can have periodic points.

For another example we consider a *complex skew shift transformation* T on the k-dimensional torus $\mathbb{T}^k = \{x_1, \dots, x_k \bmod 1\}$, where T is given by the formula

$$\begin{aligned} Tx = \bigl((x_1 + \alpha) \bmod 1,\ (x_2 + p_{2,1}x_1) \bmod 1,\ \dots,& \\ (x_k + p_{k,1}x_1 + \dots + p_{k,k-1}x_{k-1}) \bmod 1\bigr)& \end{aligned} \tag{5.7}$$

and α and $p_{i,j}$ are real numbers. This transformation obviously preserves the standard measure on $\mathbb{T}^k$. As was shown in [19], if α is irrational and $p_{j,j-1} \neq 0$ for any $2 \leqslant j \leqslant k$, then the transformation (5.7) is strictly ergodic. Hence, the trajectory of any point $x \in \mathbb{T}^k$ (the sequence of points $T^n x$, $n \geqslant 0$) is Weyl uniformly distributed.

Let $\pi_j \colon \mathbb{T}^k \to \{x_j \bmod 1\}$ be the natural projection of the k-dimensional torus onto the circle,

$$\pi_j(x_1, \dots, x_k) = x_j.$$

It is clear that, under the projection π_j, every trajectory $T^n x$, $n \geqslant 0$, is taken into a sequence of points uniformly distributed modulo 1. To show this, is suffices to take for f a function periodically dependent on only a single variable x_j.

We now set $x = 0$. One can prove (see [19], [23]) that

$$\pi_k(T^n x)\big|_{x=0} = \{P(n)\}, \tag{5.8}$$

where $P(z)$ is a polynomial in z of degree k, and the coefficient of z^k is irrational if α is. The converse assertion also holds: for any polynomial P of degree k with an irrational coefficient of the highest power there is a complex skew shift transformation of the k-dimensional torus of the form (5.7) for which the number α is irrational, $p_{j,j-1} \neq 0$, and the formula (5.8) holds. This implies (by Furstenberg [19]) Weyl's remarkable result on the uniform distribution of the fractional parts of any polynomial with an irrational leading coefficient. The general case in which there is an irrational coefficient of z^r, $r \geqslant 1$, can easilly be reduced to this case. Weyl's original proof is based on quite different ideas (see [5]).

Using Theorem 14, Furstenberg's reduction, and the strict ergodicity condition for the skew shift (5.7), we arrive at the following sharpening of Weyl's theorem [20].

Theorem 15. *Let $P(z) = a_0 z^k + a_1 z^{k-1} + \dots + a_k$ be a polynomial with one of its coefficients $a_0, a_1, \dots, a_{k-1}$ irrational. If $q_{n+1} \leqslant q_n$ and $\sum q_n = \infty$, then the sequence $\{P(n)\}$ is (W, q_n)-u.d.; if $p_{n+1} \geqslant p_n$ and the relation* (5.6) *holds, then the sequence $\{P(n)\}$ is (R, p_n)-u.d.*

For first-degree polynomials Theorem 15 was proved in [24] and [25].

In the conclusion of this section we supplement the Oxtoby theorem by another assertion about the behaviour of the sum

$$\sigma_n(x) = f(x) + f(Tx) + \dots + f(T^{n-1}x), \qquad x \in M. \tag{5.9}$$

Theorem 16. *Let T be a strictly ergodic homeomorphism of a compact metric space M and let f be a continuous function on M. Then there is a point x_+ (x_-) on M such that*

$$\sigma_n(x_+) - n\int_M f\,d\mu \geqslant 0 \qquad \left(\sigma_n(x_-) - n\int_M f\,d\mu \leqslant 0\right)$$

for any integer n.

This assertion, which was noted in [20], can readily be extended to Riesz (Voronoi) means with non-increasing (non-decreasing) weights. In this form it is a generalization of Theorem 11.

On the other hand, if a continuous function f is not constant, then for almost all points $x \in M$ the difference

$$\sigma_n(x) - n\int_M f\,d\mu \tag{5.10}$$

changes sign infinitely many times as $n \to \infty$. This is a consequence of a general result established in [26] for general ergodic transformations. It is of interest to note that the set of points $x \in M$ for which the difference (5.10) changes sign only finitely many times can be dense in M. A corresponding example for an ergodic rotation of the circle is indicated in [16]. In this example f is a continuous but nowhere differentiable function. The theorem on the oscillations of the difference (5.10) was generalized in [27] for Riesz (Voronoi) means with non-decreasing (non-increasing) weights.

§ 6. Weighted averages and the strong law of large numbers

To express the law of large numbers in probability theory one usually considers Cesàro convergence. However, the use of weighted means makes this circle of problems more flexible. On the one hand, the use of Riesz and Voronoi summation methods *included* in the Cesàro method enables one to sharpen classical results. On the other hand, the use of methods *including* the Cesàro method enables one to extend the area of applicability of the law of large numbers.

Let us consider a sequence of random variables $\xi_1, \xi_2, \dots$ with zero expectations and finite variances $\sigma_1^2, \sigma_2^2, \dots$. These variables are not assumed to be independent, and therefore their mutual covariances $R_{i,j} = \operatorname{cov}(\xi_i, \xi_j) = \mathsf{E}(\xi_i \xi_j)$, are non-zero in general. The classical Bernstein theorem [28] claims that if

1) the variances σ_n^2 are bounded,
2) $|R_{i,j}| \leqslant \varphi(|i-j|)$ and $\varphi(m) \to 0$ as $m \to \infty$,

then
$$\mathsf{P}\left\{\left|\frac{\xi_1 + \dots + \xi_n}{n}\right| \geqslant \varepsilon\right\} \to 0 \tag{6.1}$$
as $n \to \infty$ for any $\varepsilon > 0$.

The conditions 1) and 2) can be somewhat weakened by replacing them by the following conditions:
$$\sigma_1^2 + \dots + \sigma_n^2 = o(n^2), \tag{6.2}$$
$$\varphi(n) \to 0 \quad (C). \tag{6.3}$$

Thus, (6.2) and (6.3) imply (6.1). It is this assertion we are going to generalize [29].

Theorem 17. *If*
$$\frac{p_1^2\sigma_1^2 + \dots + p_n^2\sigma_n^2}{(p_1 + \dots + p_n)^2} \to 0 \tag{6.4}$$
and
$$\frac{p_1 R_{1,n+1} + \dots + p_n R_{n,n+1}}{p_1 + \dots + p_n} \to 0, \tag{6.5}$$
then
$$\mathsf{P}\left\{\left|\frac{p_1\xi_1 + \dots + p_n\xi_n}{p_1 + \dots + p_n}\right| \geqslant \varepsilon\right\} \to 0 \tag{6.6}$$
as $n \to \infty$ for any $\varepsilon > 0$.

The condition (6.4) can be replaced by the simpler condition
$$\frac{\sigma_n^2 p_n}{p_1 + \dots + p_n} \to 0. \tag{6.7}$$

More precisely, using Stolz's theorem and the assumption
$$p_n \Big/ \sum_1^n p_j \to 0, \tag{6.8}$$

one can readily prove that (6.7) implies (6.4).

On the other hand, assuming that $|R_{i,j}| \leqslant \varphi(|i-j|)$, one can represent the condition (6.5) in a simpler and more universal form:

$$\varphi(n) \to 0 \quad (W, p_n). \tag{6.9}$$

The condition (6.8) is the regularity criterion for the (W, p_n)-method. In particular, if $\varphi(n) \to 0$, then the condition (6.9) is certainly satisfied, and hence the same holds for (6.5).

What is new in Theorem 17? The classical condition (6.2) means that the variances σ_n^2 can increase slower than n; more precisely, if $\sigma_n^2 = o(n)$, then (6.2) holds. If $\sigma_n^2 = cn$ ($c = \text{const} > 0$), then the condition (6.2) fails, and it is well known that in this case the classical law of large numbers (6.1) fails in general even for independent random variables (for this topic, see [30], Chap. X).

Let us set $p_n = 1/n$. Then the relation (6.7) becomes $\sigma_n^2 = o(n \ln n)$. In particular, the variances can grow linearly with n, and we are dealing with the law of large numbers in the form (6.6), provided, of course, that the covariances $R_{i,j}$ tend to 0 uniformly with respect to $|i-j|$ (or, more generally, provided that the condition (6.9) holds with the weight coefficients $p_n = 1/n$).

One can use another approach to the circle of problems under consideration. Let us consider the case when $\sigma_1^2 = \sigma_2^2 = \cdots$ (for example, the random variables $\xi_1, \xi_2, \dots$ are identically distributed but dependent). Then the condition (6.7) certainly holds by the assumption (6.8). One can consider the case in which $p_n = \exp n^\alpha$, where $0 \leqslant \alpha < 1$. As α increases, the 'power' of the corresponding (W, p_n)-methods also increases, and therefore we obtain weaker and weaker conditions about the decrease of the covariances $\varphi(n)$. On the other hand, the power of the $(R, \exp n^\alpha)$-methods decreases, and as $\alpha \to 1$ these methods approach ordinary convergence in the limit. However, in this case the convergence in probability to 0 of the Riesz means for the random variables $\xi_1, \xi_2, \dots$ becomes a more subtle (and hence deeper) fact.

The idea of generalizing the law of large numbers by using summation methods other than the Cesàro method is certainly not new. We mention the papers [31] and [32] in which relations of the form (6.6) are proved for matrix summation methods of more general form, but the random variables $\xi_1, \xi_2, \dots$ are assumed there to be independent.

Following [33], let us now discuss the possibility of generalizing the *strong* law of large numbers.

Theorem 18. *Let $\xi_1, \xi_2, \dots$ be identically distributed independent random variables with zero means, let the variables have finite moments of order $\leqslant 2k$, and let the numbers $p_1, p_2, \dots$ be such that*

$$\sum_{n=1}^{\infty} \left[\frac{p_1^2 + \cdots + p_n^2}{(p_1 + \cdots + p_n)^2} \right]^k < \infty.$$

Then

$$\mathsf{P}\{\xi_n \to 0 \ (R, p_n)\} = 1.$$

Corollary. *Let ξ_n have finite moments of all orders. Then*

$$\xi_n \to 0 \quad (R, \exp n^\alpha) \quad a.s.$$

for any $0 \leqslant \alpha < 1$.

Theorem 19. *Let $\xi_1, \xi_2, \dots$ be independent random variables with zero mean values and with variances $\sigma_1^2, \sigma_2^2, \dots$. If $p_j > 0$, $\sum p_j = \infty$, and*

$$\sum \frac{p_j^2 \sigma_j^2}{(p_1 + \dots + p_j)^2} < \infty, \tag{6.10}$$

then

$$\xi_n \to 0 \quad (R, p_n) \quad a.s. \tag{6.11}$$

It is clear that if $p_1 = p_2 = \cdots$, then the condition (6.10) passes into the classical Kolmogorov condition $\sum \sigma_n^2/n^2 < \infty$.

We apply Theorem 19 to the case of identically distributed random variables. In this case the condition (6.10) becomes

$$\sum \frac{p_n^2}{(p_1 + \dots + p_n)^2} < \infty. \tag{6.12}$$

For instance, let $p_n = \exp n^\alpha$, $\alpha \geqslant 0$. Then the series (6.12) is convergent only for $\alpha < 1/2$, which is a weaker result than Theorem 18. However, it should be noted that in Theorem 19 one imposes weaker conditions on the random variables. We also note that for $\alpha = 1/2$ one can present an example of a sequence of *unbounded* identically distributed independent random variables $\xi_1, \xi_2, \dots$ for which the relation (6.11) fails.

Theorem 20. *Let $\xi_1, \xi_2, \dots$ be independent random variables with zero expectations and equal variances. If $|\xi_n| \leqslant M$ a.s., $M = \mathrm{const}$, and the numbers p_n are such that*

$$\frac{p_n^2}{\sum_1^n p_k^2} = \frac{\psi(n)}{\ln \ln \sum_1^n p_k^2},$$

where $\psi(n) \to 0$ as $n \to \infty$, then

$$\xi_n \to 0 \quad (R, p_n) \quad a.s.$$

This assertion is derived by using the Khinchin–Kolmogorov law of the iterated logarithm.

In particular, if $p_n = \exp n^\alpha$, where $\alpha < 1$, then the function ψ is decreasing like

$$\frac{\ln n}{n^{1-\alpha}}.$$

We now set

$$p_n = \exp \frac{n}{\ln^\gamma n}, \qquad \gamma > 0. \tag{6.13}$$

Using the Euler–Maclaurin summation formula, one can readily see that the fractions

$$\frac{\sum p_k^2}{\left(\sum p_k\right)^2}$$

decrease asymptotically like $1/\ln^\gamma n$ as n increases. Hence, the series in Theorem 18 are divergent for any k. However, in this case one has

$$\psi(n) \sim \ln^{1-\gamma} n.$$

Thus, if $\gamma > 1$, then (by Theorem 20) the coefficients of the form (6.13) satisfy the strong law of large numbers.

We should probably stress that the conditions on the random variables ξ_n in Theorem 20 are stronger than those in Theorem 18.

§ 7. Individual ergodic theorem

Let M be a finite measure space with measure μ ($\mu(M) < \infty$), and let T be a measure-preserving transformation (not necessarily invertible): $\mu(T^{-1}A) = \mu(A)$ for any measurable domain A. By the *Birkhoff–Khinchin theorem*, for any $f \in L_1$ and almost any $x \in M$ the sequence $f(T^n x)$ is Cesàro convergent to a T-invariant integrable function $\overline{f}(x)$, and

$$\int_M f\,d\mu = \int_M \overline{f}\,d\mu.$$

This remarkable result was preceded by the *von Neumann mean ergodic theorem* asserting the convergence in the mean

$$f(T^n x) \to \overline{f}(x) \quad (C).$$

Let L_2 be the linear space of square-summable complex-valued functions f on M. To any function f one can assign the function $g(x) = f(Tx)$, $x \in M$. As is well known, the correspondence $f \mapsto g = Uf$ defines an isometric operator U on L_2. The von Neumann theorem asserts that

$$U^n f \to Pf \quad (C),$$

where P stands for the operator projecting the space L_2 onto the subspace of all U-invariant vectors.

Theorem 21. *If U is an isometric operator on a complex Hilbert space and the weight coefficients p_n satisfy the condition* (5.4), *then*

$$U^n f \to Pf \quad (R, p_n) \quad \textit{and} \quad (W, p_n).$$

For $p_n = 1$ we obtain the classical von Neumann theorem. Theorem 21 claims that as $n \to \infty$ the functions $f(T^n x)$ converge in the mean to an invariant function of the transformation T in a stronger sense than is claimed in von Neumann's theorem.

Theorem 21 remains meaningful if one replaces the Hilbert space by the complex plane. In this case any isometric operator is represented by a complex number z

whose modulus is equal to 1. If $z = 1$, then the Voronoi means are obviously equal to 1. If $z \neq 1$, then Theorem 21 asserts that

$$z^n \to 0 \quad (W, q_n).$$

For example, let us consider the case $q_n = 1/n$. It is clear that

$$\frac{z^n + z^{n-1}/2 + \cdots + 1/n}{1 + 1/2 + \cdots + 1/n} = \frac{z^{n+1}(z^{-1} + z^{-2}/2 + \cdots + z^{-n-1}/n)}{\ln n + O(1)}. \tag{7.1}$$

Since $z^{-1} \neq 1$, the expression in parentheses tends to $\ln(1 - z^{-1})$ as $n \to \infty$. This is a univalent continuous function on the circle with a point deleted: $\{z : |z| = 1$ and $z \neq 1\}$. Hence, the expression (7.1) tends to 0 like $1/\ln n$.

In [34] more general summation methods defined by an infinite matrix of non-negative numbers $a_{n,k}$ $(n, k = 0, 1, 2, \dots)$ with

$$\sum_{k=0}^{\infty} a_{n,k} = 1$$

for any n were considered. There one can find a criterion for the means

$$\sum_{k=0}^{\infty} a_{n,k} U^k f$$

to converge in the mean to Pf:

(1) $\lim_{n\to\infty} \sum_{k=0}^{\infty} a_{n,k\alpha+j} = 1/\alpha$ for any $0 \leqslant j \leqslant \alpha$ and any $\alpha = 2, 3, 4, \dots$,
(2) the sequence $\{n\gamma\}$ is uniformly distributed with respect to this summation method for any irrational γ.

We note that the condition (2) is not constructive. The fact is that the description of matrix methods for which the sequence $\{n\gamma\}$, $\gamma \notin \mathbb{Q}$, is uniformly distributed on the interval $[0, 1]$ is still an open problem (for this topic, see [21]).

In [34] one can also find sufficient conditions on the coefficients $a_{n,k}$ under which the mean ergodic theorem holds for weakly and strongly mixing transformations. For results of another nature related to the generalization of von Neumann's theorem for weighted averages, see [35].

The following question seems to be substantially more difficult: Is it possible to sharpen the Birkhoff–Khinchin individual ergodic theorem by replacing Cesàro summation by a weaker Riesz or Voronoi summation method? As we have seen in § 5, the answer is positive for strictly ergodic transformations (Theorem 14). However, one cannot exclude the possibility that the answer could be negative for pointwise convergence in typical situations.

A similar problem was treated in papers of Baxter ([36], [37]; see also [38]) concerning the Riesz method (R, p_n) for which the weight coefficients are determined from the recursion relation

$$p_n = f_1 p_{n-1} + \cdots + f_n p_0, \qquad p_0 = 1,$$

where f_k $(k \geqslant 1)$ is a given sequence of non-negative numbers such that $\sum f_k = 1$. It is clear that the numbers p_n belong to the interval $[0, 1]$ and tend to a finite limit as $n \to \infty$. Under some additional assumptions, an individual ergodic theorem for the (R, p_n)-method was proved in [21] and [22]. In particular, if $f_1 = 1$ and $f_k = 0$ for $k \neq 1$, then we obtain the usual Birkhoff–Khinchin theorem.

If the numbers p_n tend to a finite limit, then the (R, p_n)-method can turn out to be equivalent to the ordinary Cesàro method (Baxter himself does not exclude this possibility). Therefore, the most interesting case (according to [22]) is the case when $p_n \to 0$. However, if the numbers p_n here are monotone decreasing, then the (R, p_n)-method includes the Cesàro method, and therefore the generalized individual ergodic theorem follows immediately from the usual Birkhoff–Khinchin theorem. We also mention the paper [39], in which the convergence of Riesz means of a sequence of random variables with non-increasing weights p_n was considered.

To better understand the complications arising here, let us consider in more detail the case in which M is the unit interval $[0, 1]$ and T is the *Bernoulli transformation*

$$x \mapsto 2x \bmod 1.$$

This transformation is not invertible but preserves the usual measure on $[0, 1]$, namely, $\operatorname{meas} L = \operatorname{meas}(T^{-1}L)$, where L is an arbitrary measurable set on $[0, 1]$ and $T^{-1}L$ is the full pre-image of L. Since the Bernoulli transformation is ergodic, it follows that

$$f(T^n x) \to \int_0^1 f(x)\,dx \quad (C) \tag{7.2}$$

for any integrable function f and for almost all points x.

We introduce a piecewise-constant function f as follows: $f(x) = 1$ for $0 \leqslant x < 1/2$ and $f(x) = -1$ for $1/2 \leqslant x \leqslant 1$. The functions $r_{n+1}(x) = f(T^n x)$, $n \geqslant 0$, are called the *Rademacher functions*. They form an orthonormal system. In particular, their mean values vanish and the variances are equal to 1. The relation (7.2) becomes the strong law of large numbers (by Borel; see [40]):

$$r_n(x) \to 0 \quad (C) \tag{7.3}$$

for almost all x.

Hence, the limit relation (7.3) can be replaced by the stronger relation

$$r_n(x) \to 0 \quad (R, p_n) \tag{7.4}$$

if the weight coefficients p_n satisfy the condition of Theorem 20. In particular, (7.4) holds for $p_n = \exp n^\alpha$, $\alpha < 1$. However, the problem of the validity of (7.4) for $p_n = \exp(n/\ln n)$ reduces to a more subtle study of the law of the iterated logarithm.

§ 8. Some unsolved problems

1°. What regular matrix summation methods S admit at least one S-u.d. sequence? For example, by [41] a criterion for the existence of an (R, p_n)-u.d. sequence reduces to the condition (6.8). A simpler problem is as follows: what are conditions on the weight coefficients p_n under which there is a (W, p_n)-u.d. sequence?

These problems are motivated by the following generalization of von Neumann's theorem (1925) established in [41]: if at least one W-u.d. sequence exists, then *every* sequence dense in $[0,1]$ can be made W-u.d. after an appropriate reordering. This result apparently holds for all linear regular summation methods.

2°. The following has been regarded as an open problem up to now: describe all S-methods for which the sequences $\{n\alpha\}$, where α is an arbitrary irrational number, are S-u.d. sequences. This problem for regular R- and W-summation methods looks simpler (but is also meaningful). The relation (5.4) is a *sufficient* condition.

3°. Let

$$P(z) = \alpha z^k + a_{k-1}z^{k-1} + \dots + a_1 z + a_0$$

be a polynomial of degree $k \geqslant 1$, let α be irrational, and let $x \mapsto f(x)$ be a continuous periodic function with unit period and zero mean. We set

$$\sigma_n(a) = \sum_{j=1}^{n} f(\{P(j)\}), \qquad a = (a_0, \dots, a_{k-1}). \tag{8.1}$$

There are many publications in which non-trivial estimates of such sums have been obtained in the cases $f(x) = \sin 2\pi x$ and $f(x) = \cos 2\pi x$ (see [42]).

From the results presented in § 5 one can derive the following properties of the sum (8.1):

1) there exist a_+ and a_- such that $\sigma_n(a_+) \geqslant 0$ and $\sigma_n(a_-) \leqslant 0$ for all n,
2) for almost all $a \in \mathbb{R}^k$ the sum $\sigma_n(a)$ changes sign infinitely many times as $n \to \infty$.

The latter property means that σ_n cannot take only positive or only negative values for $n \geqslant n_0$.

Is it true that the sum (8.1) has the following *recurrence* property: for any $\varepsilon > 0$ and any N there is an $n > N$ such that $|\sigma_n(a)| < \varepsilon$? For $k = 1$ this was proved in [43] and [44] under the assumption that the function f is absolutely continuous. Moreover, the recurrence property holds in this case uniformly with respect to the parameter a. A counterexample for continuous functions is indicated in [16]. Therefore, for $k > 1$ we should restrict ourselves to smooth functions.

4°. It is reasonable to consider the analogous problem for sums of the form

$$\sum_{j=1}^{n} g(j)F(\{f(j)\})$$

or

$$\sum_{j=1}^{n} g(n-j)F(\{f(j)\}),$$

where F is a smooth periodic function with unit period and zero mean value and f and g are the functions in Theorem 3 and Theorems 4–5, respectively.

5°. Is it true that if $\lambda(t)$ is a non-decreasing function and $f(x_0) \neq 0$, then the function (4.8) has infinitely many zeros?

6°. By a theorem of I. M. Vinogradov, the sequence $\{\alpha p\}$, where α is irrational and p takes the values of the consecutive primes, is uniformly distributed in the sense of Weyl's classical definition. Is it possible to sharpen this result by replacing the Cesàro method by a weaker Riesz or Voronoi method? A similar question applies for a sequence $\{F(p)\}$, where F is a polynomial for which at least one coefficient (aside from the constant term) is irrational.

7°. Finally, let us again formulate the main problem which was in fact the very reason for which this paper was written: Is it possible to replace Cesàro convergence in the Birkhoff–Khinchin ergodic theorem by a weaker Riesz or Voronoi convergence? Here is a possible refinement of this problem: Is it true that $f(T^n x) \to \overline{f}(x)$ (R, p_n), where $p_n = \exp n^\alpha$, $0 < \alpha < 1/2$?

Bibliography

[1] R. L. Duncan, "Note on the initial digit problem", *Fibonacci Quart.* **7** (1969), 474–475.
[2] B. J. Flehinger, "On the probability that a random number has initial digit A", *Amer. Math. Monthly* **73** (1966), 1056–1061.
[3] M. Tsuji, "On the uniform distribution of numbers mod 1", *J. Math. Soc. Japan* **4** (1952), 313–322.
[4] F. Benford, "The law of anomalous numbers", *Proc. Amer. Philos. Soc.* **78** (1938), 551–572.
[5] H. Weyl, *Selected works*, Nauka, Moscow 1984; the above references to theorems are to theorems in the paper cited in footnote #2. (Russian)
[6] S. Newcomb, "Note on the frequency of use of the different digits in natural numbers", *Amer. J. Math.* **4** (1881), 39–40.
[7] P. Diaconis, "The distribution of leading digits and uniform distribution mod 1", *Ann. Probab.* **5**:1 (1977), 72–81.
[8] N. Bourbaki, *Éléments de mathématique. Fonctions d'une variable réelle. Théorie élémentaire*, 2nd rev. ed., Hermann, Actualités Sci. Indust., no. 1074, Paris 1958; Russian transl., Nauka, Moscow 1965; English transl. of 3rd rev. ed., Springer-Verlag, Berlin 2004.
[9] R. A. Raimi, "The first digit problem", *Amer. Math. Monthly* **83**:7 (1976), 521–538.
[10] G. Pólya and G. Szegö, *Problems and theorems in analysis*, Part 1, reprint of 1978 English transl., Springer-Verlag, Berlin 1998; Russian transl. of orig. German, Nauka, Moscow 1978.
[11] G. H. Hardy, *Divergent series*, reprint of 1963 rev. ed., Jacques Gabay, Sceaux 1951; Russian transl. of orig. 1949 ed., Inostr. Lit., Moscow 1951.
[12] V. V. Kozlov, "On uniform distribution", *Izv. Vuzov Severn.-Kavkaz. Region. Ser. Estestv. Nauki* (2001), 96–99. (Russian)
[13] S. B. Gashkov and V. N. Chubarikov, *Arithmetic. Algorithms. Complexity of computing*, Nauka, Moscow 1996. (Russian)
[14] V. V. Kozlov and T. Madsen, "Uniform distribution and Voronoi convergence", *Mat. Sb.* **196**:10 (2005), 103–110; English transl., *Sb. Math.* **196**:10 (2005), 1495–1502.
[15] V. V. Kozlov, "On a uniform distribution on a torus", *Vestnik Moskov. Univ. Ser.* I *Mat. Mekh.* **2004**:2, 22–29; English transl., *Moscow Univ. Math. Bull.* **59**:2 (2004), 23–31.
[16] V. V. Kozlov, "The integrals of quasiperiodic functions", *Vestnik Moskov. Univ. Ser.* I *Mat. Mekh.* **1978**:1, 106–115; English transl., *Moscow Univ. Math. Bull.* **33**:1-2 (1978), 31–38.
[17] N. G. Moshchevitin, "On the recurrence of an integral of a smooth three-frequency conditionally periodic function", *Mat. Zametki* **58**:5 (1995), 723–735; English transl., *Math. Notes* **58**:5-6 (1996), 1187–1196.
[18] J. C. Oxtoby, "Ergodic sets", *Bull. Amer. Math. Soc.* **58**:2 (1952), 116–136.
[19] H. Furstenberg, "Strict ergodicity and transformation of the torus", *Amer. J. Math.* **83**:4 (1961), 573–601.
[20] V. V. Kozlov, "Weighted means, strict ergodicity, and uniform distributions", *Mat. Zametki* **78**:3 (2005), 358–367; English transl., *Math. Notes* **78**:3-4 (2005), 329–337.
[21] L. Kuipers and H. Niederreiter, *Uniform distribution of sequences*, Wiley, New York 1974; Russian transl., Nauka, Moscow 1985.
[22] V. V. Kozlov and T. Madsen, "Poincaré rotation numbers and Riesz and Voronoi means", *Mat. Zametki* **74**:2 (2003), 314–315; English transl., *Math. Notes* **74**:1-2 (2003), 299–301.

[23] I. P. Kornfel'd [Cornfeld], S. V. Fomin, and Ya. G. Sinai, *Ergodic theory*, Nauka, Moscow 1980; English transl., Springer-Verlag, New York 1982.
[24] J. Cigler, "Methods of summability and uniform distribution", *Compositio Math.* **16** (1964), 44–51.
[25] A. F. Dowidar, "A note on the generalized uniform distribution (mod 1)", *J. Natur. Sci. Math.* **11** (1971), 185–189.
[26] G. Halász, "Remarks on the remainder in Birkhoff's ergodic theorem", *Acta Math. Acad. Sci. Hungar.* **28**:3–4 (1976), 389–395.
[27] A. A. Sorokin, "On oscillations of Riesz and Voronoi means", *Vestnik Moskov. Univ. Ser.* I *Mat. Mekh.* **2005**:2, 13–17; English transl. in *Moscow Univ. Math. Bull.* **60**:2 (2005) (to appear).
[28] S. N. Bernshtein, "On the law of large numbers", *Soobshch. Khar'kov. Mat. Obshch. Ser.* II **16** (1918), 82–87. (Russian)
[29] V. V. Kozlov, T. Madsen, and A. A. Sorokin, "On weighted mean values of weakly dependent random variables", *Vestnik Moskov. Univ. Ser.* I *Mat. Mekh.* **2004**:5, 34–37; English transl., *Moscow Univ. Math. Bull.* **59**:5 (2005), 36–39.
[30] W. Feller, *An introduction to probability theory and its applications*, vol. I, 3rd rev. ed., Wiley, New York 1968; Russian transl., Mir, Moscow 1984.
[31] W. E. Franck and D. L. Hanson, "Some results giving rates of convergence in the law of large numbers for weighted sums of independent random variables", *Trans. Amer. Math. Soc.* **124**:2 (1966), 347–359.
[32] D. L. Hanson and F. T. Wright, "Some convergence results for weighted sums of independent random variables", *Z. Wahrscheinlichkeitstheor. verw. Geb.* **19** (1971), 81–89.
[33] V. V. Kozlov, "Summation of divergent series and ergodic theorems", *Trudy Sem. Petrovsk.* **22** (2002), 142–168; English transl., *J. Math. Sci.* (N.Y.) **114**:4 (2003), 1473–1490.
[34] D. L. Hanson and G. Pledger, "On the mean ergodic theorem for weighted averages", *Z. Wahrscheinlichkeitstheor. verw. Geb.* **13**:2 (1969), 141–149.
[35] L. W. Cohen, "On the mean ergodic theorem", *Ann. of Math.* (2) **41** (1940), 505–509.
[36] G. Baxter, "An ergodic theorem with weighted averages", *J. Math. Mech.* **13**:3 (1964), 481–488.
[37] G. Baxter, "A general ergodic theorem with weighted averages", *J. Math. Mech.* **14**:2 (1965), 277–288.
[38] R. V. Chacon, "Ordinary means imply recurrent means", *Bull. Amer. Math. Soc.* **70** (1964), 796–797.
[39] V. F. Gaposhkin, "On the summation of stationary sequences by Riesz methods", *Mat. Zametki* **57**:5 (1995), 653–662; English transl., *Math. Notes* **57**:5-6 (1995), 450–456.
[40] M. Kac, *Statistical independence in probability, analysis and number theory*, Wiley, New York 1959; Russian transl., Inostr. Lit., Moscow 1963.
[41] E. V. Gordelii, "On the von Neumann theorem on perturbations of everywhere dense sequences", *Vestnik Moskov. Univ. Ser.* I *Mat. Mekh.* **2004**:6, 18–24; English transl., *Moscow Univ. Math. Bull.* **59**:6 (2005), 17–23.
[42] Loo-Keng Hua, *Die Abschätzung von Exponentialsummen und ihre Anwendung in der Zahlentheorie*, Teubner, Leipzig 1959; Russian transl., Mir, Moscow 1964.
[43] V. V. Kozlov, "On a problem of Poincaré", *Prikl. Mat. Mekh.* **40**:2 (1976), 352–355; English transl., *J. Appl. Math. Mech.* **40** (1976), 326-329.
[44] E. A. Sidorov, "Conditions for uniform Poisson stability of cylindrical systems", *Uspekhi Mat. Nauk* **34**:6 (1979), 184–188; English transl., *Russian Math. Surveys* **34**:6 (1979), 220–224.

Steklov Institute of Mathematics,
Russian Academy of Sciences

Received 17/AUG/05
Translated by A. I. SHTERN

Typeset by $\mathcal{A}\mathcal{M}\mathcal{S}$-TEX