EFFECTIVE CONVERGENCE IN PROBABILITY AND AN
ERGODIC THEOREM FOR INDIVIDUAL RANDOM SEQUENCES*

V. V. V'YUGIN†

(Translated by the author)

Abstract. An algorithmic analysis of the ergodic theorem for a measure-preserving transformation is given. We prove that the classical ergodic theorem is not algorithmically effective. We present a formulation and a proof of the ergodic theorem for individual random sequences based on A. N. Kolmogorov’s algorithmic approach to the substantiation of the theory of probability and information theory.

Key words. ergodic theorem, quasiergodic theorem, stationary measure, convergence in probability, convergence almost everywhere, algorithm, random sequence, algorithmical randomness

1. Introduction. If $P$ is a measure, $f$ an integrable function, and $T$ a measure-preserving transformation, then by the ergodic theorem the limit

$$\lim_{n \to \infty} \frac{f(\omega) + f(T\omega) + \cdots + f(T^{n-1}\omega)}{n}$$

exists for almost all $\omega$ (see [1]). The classical theory of probability cannot give a precise mathematical description of the class of $\omega$ for which (1.1) holds. This problem is difficult to solve since the ergodic theorem is in some definite sense not algorithmically effective. This follows from the fact that, in general, there is in (1.1) no algorithmically computable rate of convergence in probability. Analogously, the quasiergodic theorem on convergence of (1.1) in the mean in $L_2$ (and in $L_1$) (see [16]) and the ergodic theorems for sequences of random variables, which are stationary in the broad sense (see [7], [6]), are algorithmically noneffective.

Nevertheless, the ergodic theorem has an algorithmically satisfactory formulation as well as a proof in the framework of Kolmogorov’s approach to the foundation of probability and information theory [10], [11], [12], [19]. In the framework of this approach a definition of an individual random sequence was obtained on the basis of the theory of algorithms. An investigation of this problem was begun by von Mises [15]. The precise mathematical definition of a random sequence worked out by Kolmogorov’s pupil Martin-Löf [14] gives us an opportunity to formulate laws of probability theory for individual random sequences and to obtain new “algorithmic” formulations and proofs of these laws. For instance, Vovk [20] found an algorithmic proof of the law of the iterated logarithm for chaotic (random in Kolmogorov’s sense) sequences. Up to now it has been an open question whether an ergodic theorem holds for individual random sequences. The most known proofs of the ergodic theorem — [2], [8], [9], [18, Chapter 5, section 3] — are “nonconstructive” and cannot be adapted

---

*Received by the editors July 12, 1996. This research was partially supported by the International Science Foundation and the Government of the Russian Federation, grant MRS300 from and by INTAS, and project 93-0893.

†Institute for Information Transmission Problems, RAS, Bol’shoy Karetnyi per., 19, Moscow 101447, Russia.
in order to obtain the needed result. Van Lambalgen [13, p. 58] threw doubt on the possibility carrying over the ergodic theorem to individual random sequences.

We will show in this paper that for any computable measure \( P \), for any random function \( f \), and for any measure preserving transformation \( T \), the limit (1.1) exists for any infinite sequence \( \omega \) which is typical in the Kolmogorov–Uspenskii sense [12], i.e., Martin-Löf’s random. The combinatorial part of the proof is based on a little known “constructive” proof of the ergodic theorem proposed by Bishop [3].

The results of section 3 on algorithmic noneffectiveness, of convergence in probability give the reason for the difficulty of finding a “pointwise” version of the ergodic theorem.

2. Basic notions and notation. For simplicity of the exposition we consider the probability space \((\Omega, \mathcal{F}, P)\) as basis. Here \( \Omega \) is a set of all infinite sequences consisting of 0’s and 1’s. A Borel field \( \mathcal{F} \) is generated by the balls of the type \( \Gamma_x = \{ \omega \in \Omega \mid x \subseteq \omega \} \), where \( x = x_1 x_2 \cdots x_k \) is a finite sequence consisting of 0’s and 1’s, and \( x \subseteq \omega \) means that the sequence \( \omega \) is an extension of the finite sequence \( x \). We consider a measure \( P \) on the space \( \Omega \).

We will use the basic notions from [18]. An arbitrary transformation of a measurable function \( T: \Omega \rightarrow \Omega \) is called a transformation of the space \( \Omega \). The transformation \( T \) of the space \( \Omega \) preserves the measure if \( P(T^{-1}(A)) = P(A) \) for all measurable subsets \( A \) of the space \( \Omega \). A measurable subset \( A \subseteq \Omega \) is called invariant under the transformation \( T \) if \( T^{-1}(A) = A \). The transformation \( T \) is called ergodic if each set which is invariant under \( T \) has measure 0 or 1. The main example of the transformation preserving the measure is a shift defined by \((T \omega)_i = \omega_{i+1}\), where \( \alpha_i \) denotes the \( i \)th element of an infinite sequence \( \alpha \). If the shift \( T \) preserves a measure \( P \), then such \( P \) is stationary, i.e.,

\[
P\{\omega \mid \omega_1 = x_1, \ldots, \omega_{i+k-1} = x_k\} = \{\omega \mid \omega_1 = x_1, \ldots, \omega_k = x_k\}
\]

for all positive integers \( i, k \geq 1 \), and all numbers \( x_1, \ldots, x_k \) equal to 0 or 1.

For any function \( f \) on \( \Omega \) we denote by \( Tf \) a function \( Tf(\omega) = f(T \omega) \). Denote \( T^0 \omega = \omega \), \( T^{k+1} \omega = T(T^k \omega) \) for \( k \geq 0 \).

We shall need also some algorithmic notions. The theory of algorithms is systematically treated in, for example, [17]. To understand this paper an intuitive idea of algorithm will suffice.

In this paper we consider algorithms which transform finite objects into finite objects. Integral and rational (but not real) numbers are examples of such objects. Finite sequences of finite objects are again finite objects. Balls of the type \( \Gamma_x \) are also finite objects since a finite sequence \( x \) of 0’s and 1’s representing such a ball is a finite object. Let \( A \) be a set of all finite objects of some type (for instance, the set of all pairs of rational numbers) and \( B \) be another. In section 3 we shall need some effective way of describing all computable functions of the type \( A \rightarrow B \). More precisely, the following assertion — a theorem on the existence of a universal function — holds [17].

\textbf{Proposition 1.} There is a computable function \( K(i, x) \), where \( i \) is a positive integer and \( x \in A \), with values in \( B \) such that for any computable function \( m(x) \) from \( A \) to \( B \) there exists an \( i \) such that \( m(x) = K(i, x) \) for all \( x \) and both sides of this equality are simultaneously defined or undefined.

The range of some computable function is called recursively enumerable.

We will consider a set \( N \) of all positive integers, a set \( Q \) of all rational numbers, and a set \( R \) of all real numbers. Let \( X \) be a set of all finite sequences, consisting of 0’s and 1’s. A function \( f \) from \( \Omega \) to \( R \cup \{-\infty, +\infty\} \) is lower semicomputable if there
exist computable functions \( r(i) \) from \( \mathbb{N} \) to \( \mathbb{Q} \) and \( x(i) \) from \( \mathbb{N} \) to \( \Xi \) such that the inequality \( r < f(\omega) \) holds if and only if \( r = r(i) \) and \( x(i) \subseteq \omega \) for some \( i \). In other words, the lower semicomputability of \( f(\omega) \) means that if \( r < f(\omega) \), then this fact will sooner or later be established by some algorithm. Whereas, if \( r \geq f(\omega) \) we may be forever uncertain. A function \( f \) is lower semicomputable if the function \(-f\) is upper semicomputable.

A function \( f \) taking rational values as well as \(-\infty\) and \(+\infty\) is called simple if there is a partition of the space \( \Omega \) into a finite sequence of balls such that \( f(\omega) \) is constant on each of them. So, the simple function is a constructive (finite) object. The next assertion gives a useful characterization of lower semicomputable functions [21].

**Proposition 2.** For any lower semicomputable function \( f(\omega) \) there is a recursively enumerable nondecreasing sequence of simple functions \( f_n(\omega) \) such that

\[
\lim_n f_n(\omega) = f(\omega).
\]

**Proof.** Let \( x(i) \) and \( r(i) \) be from the definition of lower semicomputable function. Let us define

\[
f_n(\omega) = \sup \{ r \mid r = r(i), \ x(i) \subseteq \omega, \ i \leq n \} \quad \text{(put sup} \emptyset = -\infty).
\]

A function \( f \) from \( \Omega \) to \( \mathbb{R} \) is called computable if it is simultaneously lower and upper semicomputable. It is easy to verify that in this case there is an algorithm which uses as inputs a rational \( \varepsilon > 0 \) and a sequence \( \omega \in \Omega \). After a finite number of steps this algorithm puts out a rational number \( r \) such that \( |f(\omega) - r| < \varepsilon \). Note that this algorithm uses only a finite initial part of the sequence \( \omega \).

The notion of a computable function from \( \Xi \) to \( \mathbb{R} \) is defined analogously. A measure \( P \) on \( \Omega \) is computable if the function \( P(x) = P(\Gamma_x) \) is computable, where \( x = x_1 \cdots x_n \) is a finite sequence consisting of 0's and 1's.

We also use the concept of a computable operation on \( \Omega \cup \Xi \). Following [19] in order to define such an operation, we consider a recursively enumerable set \( F \) of ordered pairs of finite sequences satisfying the following properties:

1. \((x, \emptyset) \in F \) for all \( x \), where \( \emptyset \) is the empty sequence;
2. if \((x, y) \in F, x \subseteq x' \) and \( y' \subseteq y \), then \((x', y') \in F \);
3. if \((x, y) \in F \) and \((x, y') \in F \), then \( y \subseteq y' \) or \( y' \subseteq y \).

A computable operation \( f \) is defined by the formula

\[
f(\omega) = \sup \{ y \mid x \subseteq \omega \text{ and } (x, y) \in F \text{ for some } x \}.
\]

More precisely, the computable operation may be considered as some algorithm which when fed with an infinite or a finite sequence \( \omega \) takes it sequentially bit by bit, processes it, and produces an output sequence also sequentially bit by bit.

A transformation \( T \) of the space \( \Omega \) is called computable if it coincides on \( \Omega \) with some computable operation \( f \) from \( \Omega \cup \Xi \) into itself. A shift is the simplest example of a computable transformation.

**3. Convergence in probability.** Let \( f_n(\omega) \) be a sequence of random functions. The sequence \( f_n(\omega) \) converges in probability to a function \( f(\omega) \) if, for each \( \delta > 0 \),

\[
P \left\{ \omega : |f_n(\omega) - f(\omega)| > \delta \right\} \longrightarrow 0 \quad \text{as } n \rightarrow \infty.
\]

This convergence in probability is called algorithmically effective if there is a computable non-negative integer-valued function \( m(\delta, \varepsilon) \) such that, for all rational \( \delta > 0 \)
and \( \varepsilon > 0 \),

\[
P\left\{ \omega: \left| f_n(\omega) - f(\omega) \right| > \delta \right\} < \varepsilon
\]

for all \( n \geq m(\delta, \varepsilon) \). It is easy to see that this definition is equivalent to the requirement that, for all rational \( \varepsilon > 0 \) and \( \delta > 0 \),

\[
(3.1) \quad P\left\{ \omega: \left| f_n(\omega) - f_n'(\omega) \right| > \delta \right\} < \varepsilon
\]

for all \( n, n' > m(\delta, \varepsilon) \) and for some computable function \( m(\delta, \varepsilon) \).

As a rule, the convergence in probability for sums of random variables is algorithmically effective. The function \( m(\delta, \varepsilon) \) can be found by means of the Chebyshev inequality or analogous estimates. For instance, if \( X_i(\omega) \) is a sequence of independent random variables with mean \( \mu \) and finite variance \( \sigma^2 \), then

\[
(3.2) \quad P\left\{ \omega: \left| \frac{1}{n} \sum_{i=1}^{n} X_i(\omega) - \mu \right| > \delta \right\} \leq \frac{\sigma^2}{n\delta^2} \leq \varepsilon
\]

for all \( n \geq m(\delta, \varepsilon) = \sigma^2/(\varepsilon \delta^2) \) (for simplicity we assume that \( \sigma \) is a rational number).

If \( P \) is a stationary measure, then the convergence in probability (3.3) below holds for the corresponding function \( f \) [18, Chapter 5, section 3]. In Theorem 1 the sequence \( \omega \) belongs to \( \Omega \), and \( \omega_k \) stands for the \( k \)th element of \( \omega \).

**Theorem 1.** There exists a computable stationary measure \( P \) such that the convergence in probability

\[
(3.3) \quad P\left\{ \omega: \left| \frac{1}{n} \sum_{k=1}^{n} \omega_k - f(\omega) \right| > \delta \right\} \rightarrow 0
\]

is not algorithmically effective.

**Proof.** Let \( K(i, \delta, \varepsilon) \) be a computable universal function, \( \delta \) and \( \varepsilon \) non-negative rational numbers. For any computable function \( m(\delta, \varepsilon) \), taking non-negative rational values, there exists an \( i \) such that \( m(\delta, \varepsilon) = K(i, \delta, \varepsilon) \) for all rational \( \delta, \varepsilon > 0 \).

We give an algorithmically effective method of constructing, for any \( i \), a measure \( P_i \) generating the homogeneous Markov chain with two states 0 and 1.

For any \( i \) we define a computable real number \( \alpha_i \) with an infinite binary expansion \( 0, \alpha_{i1} \alpha_{i2} \ldots \alpha_{ij} \ldots \) as follows. For \( 1 \leq s < \infty \), put \( \alpha_{is} = 1 \), if \( K(i, \frac{1}{4}, 2^{-(i+1)}) \) terminates in \( \leq s \) steps of the computation, \( s > \max\{K(i, \frac{1}{4}, 2^{-(i+1)}), 10\} \) and \( \alpha_{is'} = 0 \) for all \( 1 \leq s' < s \). Put \( \alpha_{is} = 0 \), otherwise. Here we set \( \delta = \frac{1}{4}, \varepsilon = 2^{-(i+1)} \). Let \( k(i) \) be equal to \( s \) such that \( \alpha_{is} = 1 \) if such \( s \) exists, and \( k(i) = \infty \), otherwise. Thus, \( \alpha_i = 2^{-k(i)} \), if \( k(i) < \infty \) and \( \alpha_i = 0 \), otherwise.

Let us define a homogeneous Markov chain by putting \( P_i\{\omega_1 = 0\} = P_i\{\omega_1 = 1\} = \frac{1}{2} \) as initial probabilities and \( P_i\{\omega_{s+1} = 0 \mid \omega_s = 1\} = P_i\{\omega_{s+1} = 1 \mid \omega_s = 0\} = \alpha_i \) as transition probabilities, where \( s \) is arbitrary.

Let \( P_i \) be a corresponding probability distribution on the whole space \( \Omega \). Since the real number \( \alpha_i \) is computable with an arbitrary degree of accuracy, the measure \( P_i \) is computable. We define the measure \( P \) by

\[
P = \sum_{i=1}^{\infty} 2^{-i} P_i.
\]
Since each measure $P_i$ is stationary, $P$ is also a stationary measure. It is easy to see that the measure $P$ is also computable with an arbitrary degree of accuracy.

Let $m(\delta, \varepsilon)$ be an arbitrary computable function defined for all positive rational numbers $\delta$ and $\varepsilon$ and let $i$ be such that $m(\delta, \varepsilon) = K(i, \delta, \varepsilon)$ for these $\delta$ and $\varepsilon$. Then, $k(i) < \infty$ and $\alpha_i \neq 0$. As follows from the ergodic theorem for Markov chains, the stationary probability distribution for a Markov chain generated by $P_i$ is $\pi(0) = \pi(1) = \frac{1}{2}$. By the law of large numbers for a Markov chains [18],

\begin{equation}
\mathbb{P}_i \left\{ \omega : \left| \frac{1}{n} \sum_{j=1}^{n} \omega_j - \frac{1}{2} \right| < 0,1 \right\} \rightarrow 1 \quad \text{as} \quad n \to \infty.
\end{equation}

Note that, for a sequence consisting of $k(i)$ zeros and for a sequence consisting of $k(i)$ ones,

$$
\mathbb{P}_i(1 \cdots 1) = \mathbb{P}_i(0 \cdots 0) = \frac{1}{2} (1 - \alpha_i)^{k(i)-1} = \frac{1}{2} - \frac{1}{2} (k(i) - 1) 2^{-k(i)} + ck^2(i) 2^{-2k(i)} > \frac{2}{5}
$$

(since $k(i) > 10$), where $0 \leq c \leq 1$.

From this it follows that

$$
\mathbb{P}_i \left\{ \omega : \frac{1}{k(i)} \sum_{j=1}^{k(i)} \omega_j = 1 \quad \text{or} \quad 0 \right\} > \frac{4}{5}.
$$

Moreover, $k(i) > m(\frac{1}{4}, 2^{-(i+1)})$. From this and (3.4) it follows that there is a sufficiently large $n > m(\frac{1}{4}, 2^{-(i+1)})$ such that

$$
\mathbb{P}_i \left\{ \omega : \left| \frac{1}{k(i)} \sum_{j=1}^{k(i)} \omega_j - \frac{1}{n} \sum_{j=1}^{n} \omega_j \right| > \frac{1}{4} \right\} > \frac{1}{2}.
$$

Then, the $P$-measure of this set is larger then $2^{-i+1} = \varepsilon$, and so, the numbers $k(i), n,$ and the function $m(\delta, \varepsilon)$ do not satisfy condition (3.1). The theorem is proved.

The construction of this theorem is an algorithmic formalization of some heuristic arguments proposed by Bishop [3].

The sequence of functions $f_n(\omega)$ converges to a function $f(\omega)$ almost surely if $\mathbb{P}\{\omega : \lim_n f_n(\omega) = f(\omega)\} = 1$. Since the convergence almost surely is equivalent to

\begin{equation}
\mathbb{P} \left\{ \omega : \sup_{m \geq n} \left| \frac{1}{m} \sum_{k=1}^{m} f_k(\omega) - f(\omega) \right| > \delta \right\} \rightarrow 0
\end{equation}

as $n \to \infty$ for any $\delta > 0$, we obtain the following corollary.

**Corollary 1.** There exists a computable stationary measure $P$ such that the convergence almost surely (3.5) is not algorithmically effective.

The quasiergodic theorem in $L_2$ asserts that for any function $f$ from $L_2$ there exists a function $\hat{f}$ from $L_2$ which is invariant under the measure preserving transformation $T$ such that $A_n f \to \hat{f}$ in the sense of $L_2$, i.e., $\|A_n f - \hat{f}\|_2 \to 0$, where

$$
A_n f = \frac{1}{n} \sum_{k=0}^{n-1} T^k f \quad [1], [16].
$$

The quasiergodic theorem for the convergence in $L_1$ is considered analogously.
COROLLARY 2. There exist a computable stationary measure $\mathbf{P}$, a measure preserving transformation $T$, and a computable function $f \in L_2$ such that the convergence in the mean $\|A_n f - f\|_2 \to 0$ is not algorithmically effective. An analogous assertion holds for the convergence in $L_1$.

This corollary is proved analogously to Corollary 1.

Similarly, a theorem on the convergence of means

$$h_n = \frac{1}{n} \sum_{k=1}^{n} z_k$$

is, in general, not algorithmically effective, where the random variables $z_1, z_2, \ldots$ form a random process: $E(z_i) = a, E((z_i - a)^2) = b^2$ which is stationary in the broad sense and the correlation coefficient $R(z_i, z_j)$ is a function from $|i - k|$ (see [7]). By means of the construction of Theorem 1 it can be proved that, in general, the convergence $E\{(h_n - h_m)\} \to 0$ is not algorithmically effective.

4. Algorithmically random (typical) individual sequences. In this paper we use a measure-theoretic approach to the definition of the individual random sequence of outcomes. This approach was developed by Martin-Löf [14] in the framework of Kolmogorov’s algorithmic approach to the substantiation of the theory of probability [10], [11]. The measure-theoretic approach stresses the property which asserts that the random sequence must belong to each “reasonable majority” of sequences [12]. The accurate definition of majority is an algorithmic analogue of a set of measure 1. Since each set of measure 1 is a complement of a set of measure 0, it is sufficient to define the algorithmic concept of the set of measure 0. Let $\mathbf{P}$ be some computable measure on $\Omega$. A set $M \subseteq \Omega$ has $\mathbf{P}$-measure 0 if, for each rational $\epsilon > 0$, there is a sequence $x_1, x_2, \ldots$ of elements of $\Xi$ such that for $U_\epsilon = \bigcup_i \Gamma_{x(i)}$ we have $M \subseteq U_\epsilon$ and $\mathbf{P}(U_\epsilon) < \epsilon$. A $\mathbf{P}$-null set $M$ is called effectively null if there exists a computable function $x(\epsilon, i)$ such that $M \subseteq U_\epsilon = \bigcup_i \Gamma_{x(\epsilon, i)}$ and $\mathbf{P}(U_\epsilon) < \epsilon$ for each rational $\epsilon > 0$.

It can be proved that for any computable measure $\mathbf{P}$ there exists an effectively $\mathbf{P}$-null set which is the largest with respect to measure-theoretic inclusion [19], [14]. A complement of this largest effectively $\mathbf{P}$-null set is called the constructive support of the measure $\mathbf{P}$.

A sequence $\omega \in \Omega$ is called typical with respect to $\mathbf{P}$ (random in the sense of Martin-Löf) if it belongs to the constructive support of the measure $\mathbf{P}$. A more detailed consideration can be found in the papers mentioned in this section.

The following simple proposition shows that if the convergence almost surely is algorithmically effective, then this convergence holds for each typical sequence. We must only assume the computability of the measure $\mathbf{P}$ and the sequence $f_n$ (the last notion is defined analogously to the computability of functions).

PROPOSITION 3. Under the assumptions above, if the sequence $f_n(\omega)$ converges almost surely algorithmically effectively then $\lim_n f_n(\omega)$ exists for each $\omega$ which is typical with respect to the measure $\mathbf{P}$.

Proof. As follows from the definition of the effective convergence almost surely, there exists an integer-valued computable function $m(\delta, \epsilon)$ such that

$$\mathbf{P}\left\{ \omega: \sup_{k, k' \geq n} |f_k(\omega) - f_{k'}(\omega)| > \delta \right\} < \epsilon$$

for all $n \geq m(\delta, \epsilon)$ and all positive rational $\delta$ and $\epsilon$. Let us denote $W_{m(\delta, \epsilon)} = \{\omega: |f_k(\omega) - f_{k'}(\omega)| > \delta\}$. As follows from the computability of $f_k$ this set is equal
to the union of a recursively enumerable sequence of balls of the type $\Gamma_x$. Put also $V_i = \bigcup_{k,k' \geq n(i)} W_{k,k',1/i}$, where $n(i) = n(1/i, 2^{-(i+1)})$. Then $P(V_i) < 2^{-(i+1)}$ for all $i$.

Suppose that $\lim_n f_n(\omega)$ does not exist for some sequence $\omega$ which is typical with respect to a measure $P$. Then there exists an $i$ such that $|f_k(\omega) - f_{k'}(\omega)| > i^{-1}$ for infinitely many $k$ and $k'$. For any $j \geq i$ choose such $k, k' \geq n(j)$. Then $\omega \in W_{k,k',1/j} \subseteq V_j$. Put $V_s = \bigcup_{j \geq s} V_j$. We have $\omega \in V_s$ for all $s$.

By definition, $U_{s+1} \subseteq U_s$ and $P(U_s) < 2^{-s}$ for all $s$. The sequence $U_s$ defines an effectively $P$-null set $\cap_s U_s$ and a sequence $\omega$ belongs to this set. A contradiction with the typicalness of $\omega$ obtained here proves this proposition.

As a rule, a law of the probability theory, i.e., an assertion which holds almost surely, allows us to define an effectively $P$-null set of all $\omega$ for which this law fails. From this it follows that this law holds for all $\omega$ from the constructive support of the measure $P$. To define such an effectively $P$-null set we usually use algorithmically effective estimates of the convergence, as in Proposition 3. For instance, for a sequence of independent trials it is easy to deduce such estimates from the estimate (see [4])

$$P\{\omega: \sup_{k \geq n} \left| \frac{S_k}{k} - p \right| > \varepsilon \} \leq e^{-4} \frac{1}{n+1}.$$  

Nevertheless, as Theorem 1 shows, we can not always use Proposition 3. A convergence of a special type for sums of the random variables, which is proved to be algorithmically effective, will be used in the proof of the ergodic theorem.

In the framework of the measure-theoretic approach one can introduce a quantitative characteristic of nonrandomness of an outcome $\omega$. Following [21], we consider, for an arbitrary computable measure $P$, a function $p(\omega)$ from $\Omega$ to $\mathbb{R}^+ \cup \{\infty\}$ which characterizes the degree of difference between the measure $P$ and an outcome $\omega$ — the outcomes having large values of $p(\omega)$ are considered to be impossible from the standpoint of the measure $P$. Here $\mathbb{R}^+$ is a set of all non-negative real numbers. More precise, a function $p(\omega)$ is called a measure of impossibility of $\omega$ with respect to $P$ if

1. $p(\omega)$ is lower semicomputable;
2. $E_P(p) = \int p(\omega) dP \leq 1$.

The functions possessing properties (1) and (2) are sometimes called the effective tests of randomness. Requirement (2) appeared at first in [5].

The Chebyshev inequality for $p(\omega)$ follows from (2):

$$P\{\omega: p(\omega) > r \} < r^{-1}$$

for any $r > 0$.

Requirement (1) means that the condition of “falsification” $p(\omega) > r$, is positively decidable. Note that we do not require $p(\omega)$ to be a computable function. This corresponds to the natural asymmetry between the properties of “falsification” and “verification” ($p(\omega) \leq r$). The next proposition is proved analogously to the Kolmogorov theorem on the existence of an optimal mode of the description of finite objects.

Proposition 4. Let $P$ be an arbitrary computable measure. There is a (universal) measure of impossibility $p_0(\omega)$ such that, for any measure of impossibility $p(\omega)$, $c p_0(\omega) \geq p(\omega)$ for each $\omega \in \Omega$, where $c$ is a positive constant.

The proof of this proposition can be found in [21].

The function $p_0(\omega)$ is called a level of impossibility of an outcome $\omega$ with respect to a measure $P$. 

Proposition 5. Let $\mathbf{P}$ be an arbitrary computable measure. An infinite sequence $\omega$ consisting of 0’s and 1’s is typical with respect to the measure $\mathbf{P}$ if and only if $p_0(\omega) < \infty$.

Proof. Let $U_m = \{\omega: p_0(\omega) > 2^{-m}\}$. It follows from (2) that $\mathbf{P}(U_m) < 2^{-m}$ for each $m$. By (1), if $p_0(\omega) > r$, then there exists an initial fragment $\omega_1 \cdots \omega_n$ of the sequence $\omega$ such that $p_0(\omega) > r$ for any sequence $\omega'$, where $\omega'_1 = \omega_1 \cdots \omega'_n = \omega_n$. From this it follows that the set $U_m$ is open and, moreover, there exists an algorithm, which using $m$ enumerates the balls defining $U_m$. Hence, if $p_0(\omega) = \infty$, then $\omega$ belongs to the effectively $\mathbf{P}$-null set $\cap U_m$.

To prove the converse assertion, suppose that $\omega$ belongs to some effectively $\mathbf{P}$-null set. By definition there exists a sequence $U_m$ of open sets such that $U_{m+1} \subseteq U_m$ and $\mathbf{P}(U_m) < 2^{-m}$ for all positive integers $m$. Also, $\omega \in \cap U_m$. In addition, using $m$ we can effectively enumerate all the balls forming $U_m$. Let us define a lower semicomputable function

$$t(m, \omega) = \begin{cases} 1, & \text{if } \omega \in U_m, \\ 0, & \text{otherwise}. \end{cases}$$

Put $p(\omega) = \sum_{m=1}^{\infty} t(m, \omega)$. It is easy to see that the function $p(\omega)$ is lower semicomputable. Also,

$$\int p(\omega) \, d\mathbf{P} = \sum_{m=1}^{\infty} 2^{-m} \int t(m, \omega) \, d\mathbf{P} \leq 1.$$

Hence, the function $p(\omega)$ is a measure of impossibility of an outcome $\omega$ with respect to $\mathbf{P}$. From $\omega \in \cap U_m$ it follows that $p_0(\omega) = \infty$. The proposition is proved.

Many results of probability theory are of the form, “$A(\omega)$ holds $\mathbf{P}$-almost surely,” where $A(\omega)$ is some property of the type (5.1). One of the ways to strengthen such results is to prove its “pointwise” counterpart, “If $p(\omega) < \infty$, then $A(\omega)$,” where $p$ satisfies properties (1) and (2).

5. The ergodic theorem for typical (individual random) sequences. Let $(\Omega, \mathcal{F}, \mathbf{P})$ be a probability space, $T$ a transformation preserving the measure $\mathbf{P}$, and $f(\omega)$ an integrable random function.

The ergodic theorem, in its general form for a probability space $(\Omega, \mathcal{F}, \mathbf{P})$ and a transformation $T$ preserving the measure $\mathbf{P}$, asserts that $\mathbf{P}$-almost surely

$$\lim_{n \to \infty} \frac{1}{n} \sum_{k=0}^{n-1} f(T^k \omega) = \bar{f}(\omega)$$

for some integrable function $\bar{f}$, which is invariant with respect to $T$, such that $\mathbf{E}(\bar{f}) = \mathbf{E}(f)$, where $\mathbf{E}$ is an expectation (a function $\bar{f}$ is invariant if $f(T\omega) = \bar{f}(\omega)$ for almost all $\omega$). If the transformation $T$ is ergodic, then $\bar{f}(\omega) = \mathbf{E}(f)$ almost surely [1], [18].

If the measure $\mathbf{P}$ is stationary we set $f(\omega_1 \omega_2 \cdots) = \omega_1$, $T$ is the shift. Then we have

$$\frac{1}{n} \sum_{k=0}^{n-1} f(T^k \omega) = \frac{1}{n} \sum_{k=1}^{n} \omega_k$$

in (5.1).

Unlike many other assertions of probability theory, which hold almost surely, the well-known proofs of this theorem cannot effectively define the class of $\omega$ for which (5.1) holds. In this section we will use a more “constructive” proof of the ergodic
Theorem due to Bishop’s [3, p. 233] (in simplified form), which allows us to prove its “pointwise counterpart,” namely, to prove that (5.1) holds for any typical sequence. We must only additionally assume the computability of the measure P, transformation T, and function f.

**Theorem 2.** Let P be a computable measure, T a computable transformation preserving the measure P, and f a computable real function on Ω. Then there exists an integrable function ñ such that E(f) = E(f) and

$$\lim_{n \to \infty} \frac{1}{n} \sum_{k=0}^{n-1} f(T^k \omega) = ñ(\omega)$$

for any typical (random in the sense of Martin-Löf) with respect to the measure P sequence ω. In addition, ñ(Tω) = ñ(ω). If the transformation T is ergodic, then ñ(ω) = E(f) for this sequence ω.

**Proof.** Let us denote $f_j(\omega) = f(T^j \omega)$, where $j = 0, 1, \ldots$. It follows from the computability of f that f is continuous (since for any rational $r > 0$ both sets \{ω: $r < f(\omega)$\} and \{ω: $r > f(\omega)$\} are open in the topology on Ω generated by the balls $\Gamma_r$). Since Ω is compact, for some rational M we have $|f(\omega)| \leq M$ for all $\omega$. Let $\alpha$ and $\beta$ be rational numbers such that $-M \leq \alpha < \beta \leq M$. Let us define $a(u, f, \omega) = \sum_{j=0}^{u} (f_j(\omega) - \alpha)$ and $b(v, f, \omega) = \sum_{j=0}^{v} (f_j(\omega) - \beta)$, where u and v are non-negative integers. By definition,

$$\omega \in A(u, v) \quad \text{if and only if} \quad a(u, f, \omega) < b(v, f, \omega).$$

Let n be a non-negative integer. A sequence of integers $s = \{u_1, v_1, \ldots, u_N, v_N\}$ is called admissible if $-1 \leq u_1 < v_1 \leq u_2 < v_2 \leq \cdots < u_N < v_N \leq n$. We define $a(-1, f, \omega) = 0$. Denote $m_s = N$. Let us consider the Bishop function for the crossing of the boundaries $\alpha$ and $\beta$:

$$\sigma_n(\omega, \alpha, \beta) = \sup \left\{ N: \omega \in \bigcap_{j=1}^{N} A(u_j, v_j) \bigcap_{j=1}^{N-1} A(u_{j+1}, v_j) \right\}$$

(5.3) for some admissible $s = \{u_1, v_1, \ldots, u_N, v_N\}$.

By definition, $\sigma_n(\omega, \alpha, \beta)$ is lower semicomputable as a function of $n, \omega, \alpha, \beta$.

It is easy to see that if $\lim_{n \to \infty} \frac{1}{n} \sum_{j=0}^{n-1} f_j(\omega)$ does not exist, then there are rational numbers $\alpha$ and $\beta$ such that

$$-M \leq \liminf_{n \to \infty} \frac{1}{n} \sum_{j=0}^{n-1} f_j(\omega) < \alpha < \beta < \limsup_{n \to \infty} \frac{1}{n} \sum_{j=0}^{n-1} f_j(\omega) \leq M.$$

In this case, the value of $\sigma_n(\omega, \alpha, \beta)$ is unbounded as $n \to \infty$. Note that a converse assertion is also true.

For any admissible sequence $d = \{s_1, t_1, \ldots, s_N, t_N\}$, let

$$S(f, d, \omega) = \sum_{j=1}^{N} (b(t_j, f, \omega) - a(s_j, f, \omega)).$$

The proof of the theorem is based on the following combinatorial lemma.
LEMMA. For any admissible sequence $q$ there exists an admissible sequence $d$ such that $S(f,d,\omega) \geq S(f,q,\omega)$ and $m_d \geq \sigma_n(\omega, \alpha, \beta)$.

Proof. By definition of $\sigma_n$, there is an admissible sequence $-1 \leq u_1 < v_1 \leq u_2 < v_2 \leq \cdots \leq u_N < v_N \leq n$ such that (5.3) holds and $N = \sigma_n(\omega, \alpha, \beta)$.

To prove the lemma it is sufficient to prove that if for any admissible sequence $q$ we have $m_q < N$, then there is an admissible sequence $d$ such that $S(f,d,\omega) \geq S(f,q,\omega)$ and $m_d = m_q + 1$. Let $q$ be $-1 \leq s_1 < t_1 \leq s_2 < t_2 \leq \cdots \leq s_m < t_m \leq n$ and $m = m_q$. Put $s_{m+1} = n$. The value $v_{m+1}$ is defined since $m+1 \leq N$. In addition, $v_{m+1} \leq n = s_{m+1}$. Hence, there is a minimal $i$ such that $v_i \leq s_i$. If $i = 1$ we set $d = \{u_1, v_1, t_1, \ldots, s_m, t_m\}$. Let $i > 1$. Then $v_{i-1} > s_{i-1}$ and, hence, $s_{i-1} < v_{i-1} \leq u_i < v_i \leq s_i$. If $u_i < t_{i-1}$ put $d = \{s_1, t_1, \ldots, s_{i-1}, v_{i-1}, u_i, t_{i-1}, \ldots, t_m\}$.

If $u_i \geq t_{i-1}$ put $d = \{s_1, t_1, \ldots, t_{i-1}, u_i, v_i, s_i, t_i, \ldots, s_m, t_m\}$, if $i \leq m$ and $d = \{s_1, t_1, \ldots, t_m, u_{m+1}, v_{m+1}\}$ if $i = m + 1$. The sequence $d$ is admissible and $m_d = m_q + 1$. Also, we have

$$S(f,d,\omega) = S(f,q,\omega) + b(v_i, f, \omega) - a(u_i, f, \omega)$$

or

$$S(f,d,\omega) = S(f,q,\omega) + b(v_{i-1}, f, \omega) - a(u_{i-1}, f, \omega).$$

Hence we have $S(f,d,\omega) \geq S(f,q,\omega)$. The lemma is proved.

Note that $T_{f_i} = f_{i+1}$ for all $i$. Let $d = \{s_1, t_1, \ldots, s_m, t_m\}$ be an admissible sequence. By definition, if $s_j \geq 0$, then $a(s_j, f, \omega) = a(s_j - 1, f, T\omega) + f_0(\omega) - \alpha$ and $b(t_j, f, \omega) = b(t_j - 1, f, T\omega) + f_0(\omega) - \beta$.

A simple analysis of the corresponding sums shows that

$$S(f,d,\omega) = S(f,d', T\omega) + a - (\beta - \alpha) m_d,$$

where $a = 0$ if $s_1 \geq 0$ and $a = f_0(\omega) - \alpha$ if $s_1 = -1$. Moreover, $d' = \{s_1 - 1, t_1 - 1, \ldots, s_m - 1, t_m - 1\}$ if $s_1 \geq 0$ and $d' = \{-1, t_1 - 1, s_2 - 1, t_2 - 1, \ldots, s_m - 1, t_m - 1\}$ if $s_1 = -1$ and $t_1 > 0$. If $s_1 = -1$ and $t_1 = 0$, then $d' = \{s_2 - 1, t_2 - 1, \ldots, s_m - 1, t_m - 1\}$.

Let $\lambda_n(\omega) = \max \{S(f,d,\omega): d \text{ is admissible}\}$. Then

$$S(f,d,\omega) \leq \lambda_n(T\omega) + (f_0(\omega) - \alpha)^+ - (\beta - \alpha) m_d,$$

where $h^+ = \max(h,0)$ for any $h$. By the lemma, for any admissible $q$ there is an admissible $d$ such that $m_d \geq \sigma_n(\omega, \alpha, \beta)$ and $S(f,q,\omega) \leq S(f,d,\omega)$. Then,

$$S(f,q,\omega) \leq S(f,d,\omega) \leq \lambda_n(T\omega) + (f_0(\omega) - \alpha)^+ - (\beta - \alpha) \sigma_n(\omega, \alpha, \beta).$$

Taking the maximum with respect to $q$ we obtain

$$\lambda_n(\omega) \leq \lambda_n(T\omega) + (f_0(\omega) - \alpha)^+ - (\beta - \alpha) \sigma_n(\omega, \alpha, \beta).$$

Hence,

$$(\beta - \alpha) \sigma_n(\omega, \alpha, \beta) \leq (f_0(\omega) - \alpha)^+ + \lambda_n(T\omega) - \lambda_n(\omega).$$

Integrating with respect to $\omega$ we obtain

$$\int (\beta - \alpha) \sigma_n(\omega, \alpha, \beta) d\mathbf{P} \leq \int (f_0(\omega) - \alpha)^+ d\mathbf{P},$$

since $\int \lambda_n(T\omega) d\mathbf{P} = \int \lambda_n(\omega) d\mathbf{P}$. The last equality holds since the transformation $T$ preserves the measure $\mathbf{P}$. Since $f(\omega)$ is bounded, we have $\int (f(\omega) - \alpha)^+ d\mathbf{P} \leq 2M$.

Put

$$\sigma(\omega, \alpha, \beta) = \sup_n \sigma_n(\omega, \alpha, \beta).$$
Since $\sigma_{n+1}(\omega, \alpha, \beta) \geq \sigma_n(\omega, \alpha, \beta)$ for all $n$, $\sigma(\omega, \alpha, \beta)$ is integrable and
\[
\int (2M + 1)^{-1}(\beta - \alpha) \sigma(\omega, \alpha, \beta) d\mathbf{P} < 1.
\]

As follows from (5.3) the function $\sigma(\omega, \alpha, \beta)$ is lower semicomputable. Let us consider an expectation of this function by $\alpha$ and $\beta$. Let $\alpha(i)$ and $\beta(i)$ be computable functions with rational values such that $-M \leq \alpha(i) < \beta(i) \leq M$ for all $i$ and such that for any pair $(\alpha, \beta)$ of rational numbers if $-M \leq \alpha < \beta \leq M$, then $\alpha(i) = \alpha$ and $\beta(i) = \beta$ for some $i$. Such enumeration can be easily defined. Put
\[
p(\omega) = \frac{1}{2} (2M + 1)^{-1} \sum_{i=1}^{\infty} i^{-2}(\beta(i) - \alpha(i)) \sigma(\omega, \alpha(i), \beta(i)).
\]

It is easy to see that the function $p(\omega)$ is lower semicomputable and $\int p(\omega) d\mathbf{P} \leq 1$. So, $p(\omega)$ is a measure of impossibility of an outcome $\omega$ with respect to $\mathbf{P}$. As noted earlier, if $\lim_{n \to \infty} (1/n) \sum_{k=0}^{n-1} f(T^k \omega)$ does not exist, then there are rational numbers $\alpha$ and $\beta$ such that $\alpha < \beta$ and $\sigma(\omega, \alpha, \beta) = \infty$. Then $p(\omega) = \infty$. Hence, if $p(\omega) < \infty$, then $\lim_{n \to \infty} (1/n) \sum_{k=0}^{n-1} f(T^k \omega)$ exists. In particular, this limit exists if the sequence $\omega$ is typical with respect to $\mathbf{P}$.

As follows from (5.2), $f(T\omega) = 1$ for any random $\omega$. If the transformation $T$ is ergodic, then $f(\omega) = c (= E(f))$ for almost all $\omega$. Let us suppose that $f(\omega) = d \neq c$ for some $\omega$ which is typical with respect to $\mathbf{P}$. Choose rational numbers $r_1$ and $r_2$ such that $r_1 < d < r_2$ and $c < r_1$ or $c > r_2$. Put
\[
S_n = \left\{ \alpha \in \Omega: r_1 < \frac{1}{n} \sum_{k=0}^{n-1} f(T^k \alpha) < r_2 \right\},
\]
\[
\overline{S}_n = \left\{ \alpha \in \Omega: r_1 \leq \frac{1}{n} \sum_{k=0}^{n-1} f(T^k \alpha) \leq r_2 \right\}.
\]

Since the limit (5.2) is equal to $c$ almost surely, convergence in probability also holds and we have $\mathbf{P}(\overline{S}_n) \to 0$ as $n \to \infty$. A function $\mathbf{P}(\overline{S}_n)$ is upper semicomputable, since $r > \mathbf{P}(\overline{S}_n)$ if and only if
\[
1 - r < \mathbf{P}\left\{ \alpha \in \Omega: r_1 > \frac{1}{n} \sum_{k=0}^{n-1} f(T^k \alpha) \text{ or } r_2 < \frac{1}{n} \sum_{k=0}^{n-1} f(T^k \alpha) \right\}.
\]

Then for any $m$ we can effectively find an $n \geq m$ such that $\mathbf{P}(\overline{S}_n) < 2^{-(m+1)}$. By definition, $S_n = \bigcup_{j=0}^{r} \Gamma_{x(n,j)}$, where $x(n, j)$ is a computable function. Moreover, $\mathbf{P}(S_n) \leq \mathbf{P}(\overline{S}_n)$. Put $V_m = S_n$. Let us define a recursively enumerable sequence of the enclosure sets $U_i = \bigcup_{n \geq i} V_n$. By definition, $\mathbf{P}(U_i) < 2^{-i}$ for all $i$. The typical sequence $\omega$ belongs to each of these sets $U_i$. We arrive at a contradiction since $\cap_i U_i$ is an effectively $\mathbf{P}$-null set. The theorem is proved.

It is easy to see that we could take the expectation of the function $\sigma_n(\omega, \alpha, \beta)$ only with respect to $\alpha = (k/s) M$ and $\beta = (k + 1)/s M$, where $k = -s, -s + 1, \ldots, s - 1$, and $s = 1, 2, \ldots$. Thus we have
\[
p(\omega) = \frac{1}{4} (2M + 1)^{-1} \sup_{s = 1}^{n} \sum_{k=-s}^{s-1} \sum_{s=1}^{n} \sigma_n(\omega, \alpha, \beta).
\]
It is easy to verify that this function satisfies conditions (1) and (2) of the definition of a measure of impossibility.

The main statement of the ergodic theorem can be given in the following form:

$$\text{If } p(\omega) < \infty, \text{ then } \lim_{n \to \infty} \left( \frac{1}{n} \sum_{k=0}^{n-1} f(T^k \omega) \right) \text{ exists.}$$

The Chebyshev inequality for $p(\omega)$,

$$P\{\omega: p(\omega) > h\} < \frac{1}{h},$$

is an algorithmically effective analogue of the inequality of the type (3.2).

REFERENCES