Ergodic theorems for algorithmically random points

Vladimir V. V'yugin [1]

Institute for Information Transmission Problems, Russian Academy of Sciences, Bol'shoi Karetnyi per. 19, Moscow GSP-4, 127994, Russia. e-mail vyugin@iitp.ru

Abstract

This paper is a survey of applications of the theory of algorithmic randomness to ergodic theory. We establish various degrees of constructivity for asymptotic laws of probability theory. In the framework of the Kolmogorovs approach to the substantiation of the probability theory and information theory on the base of the theory of algorithms, we formulate probabilistic laws, i.e. statements which hold almost surely, in "a pointwise" form, i.e., for Martin-Löf random points. It is shown in this paper that the main statement of ergodic theory – Birkhoff's ergodic theorem, is non-constructive in the strong (classical) sense, but it is constructive in some weaker sense – in terms of Martin-Löf randomness.

1 Introduction

In the framework of the Kolmogorovs approach to the substantiation of the probability theory and information theory (see Kolmogorov [10]–[13]) on the basis of the theory of algorithms, probabilistic laws, i.e. statements of the form $P\{\omega : A(\omega)\} = 1$, where $A(\omega)$ is some asymptotic formula (of a probabilistic law), are presented in "a pointwise" form: "ω is random $\Rightarrow A(\omega)$".

Most proofs of such laws, like the strong law of large numbers or law of the iterated logarithm, stand up to constructive analysis and can be directly translated into the algorithmic form. An exception is the Birkhoff's ergodic theorem (see Bilingsly [3], Krengel [14]).

In Section 3 we analyse the main statement of ergodic theory – Birkhoff's ergodic theorem, in terms of algorithmic information theory and Martin-Löf

[1] This research was partially supported by Russian foundation for fundamental research: 20-01-00203-a.

randomness (see Li and Vitanyi [16]). Here the key point is the presence or absence of computable estimates for the rate of convergence almost surely for time averages.

In Section 3.3 we show that the classical ergodic theorem for ergodic transformations stands up to constructive analysis, since there is a computable estimate for the rate of convergence almost surely of time averages in the maximum ergodic theorem. The classical proof of this theorem for ergodic measure preserving transformations is directly translated into the algorithmic form.

This is not the case in the general position: we prove that the Birkhoff's ergodic theorem is indeed in some strong sense "nonconstructive".

In Section 3.4 we show that in the case of a not necessarily ergodic measure preserving transformation there is no computable estimate for the rate of convergence of time averages. Nevertheless, in Section 4.2 we show that in the general case, for arbitrary measure preserving transformation (not necessary ergodic) a little-known Bishop's [4] proof of the ergodic theorem can be used to obtain the algorithmic version of this theorem: time-averaged values of any computable function defined on the prefixes of the trajectory of an arbitrary Martin-Löf random point converges but there is no computable estimate for the rate of this convergence.

2 Preliminaries

Let $\{0,1\}^* = \cup_n \{0,1\}^n$ be the set of all finite binary sequences, and $\Omega = \{0,1\}^\infty$ be the set of all infinite binary sequences. In what follows by a sequence (finite or infinite) we mean the binary sequence, i.e., the sequence $\omega_1\omega_2\ldots$, where $\omega_i \in \{0,1\}$ for $i = 1, 2, \ldots$. For any finite or infinite $\omega = \omega_1 \ldots \omega_n \ldots$, we denote its prefix (initial fragment) of length n as $\omega^n = \omega_1 \ldots \omega_n$. We write $x \subseteq y$ if a sequence y is an extension of a sequence x, $l(x)$ is the length of x, λ is the empty sequence.

Let $\mathcal{R}$ be the set of all real numbers, $\mathcal{R}_+$ be the set of all nonnegative real numbers, $\mathcal{N}$ and $\mathcal{Q}$ – be the sets of all positive integer numbers and of all rational numbers.

For the basics of computability and algorithmic randomness theory, see for instance, Rogers [17] and Li and Vitanyi [16].

We fix the model of computation. Algorithms may be regarded as Turing machines and so the notion of a program and time of computation will be well-defined. Any Turing machine using a program can calculate the values of a possibly partially defined function f of the type $f : \{0,1\}^* \to \{0,1\}^*$. Such a function f is called computable or, according to historical tradition, partial recursive. This means that Turing machine when fed with an input – a finite sequence $x \in \{0,1\}^*$, transforms it according instructions of some program to another finite sequence y, stops and outputs the result $y = f(x)$ or never stops and outputs no result. In the last case, we say that the result of computation on the input x is undefined or that the value of $f(x)$ is undefined.

Any algorithm transforms finite objects into finite objects. Integer and rational numbers (but no reals) are examples of finite objects. Finite sequences of finite objects are again finite objects. The main property of finite objects that we use is that they can be enumerated with positive integers, and therefore, they can be arguments and values of computable (partial recursive) functions and algorithms.

A function f is computable (or partial recursive) if there is an algorithm (Turing machine) computing values of f. For any input x, the corresponding Turing machine when fed with x stops after several steps and outputs the result $f(x)$ if $f(x)$ is defined and never stops otherwise. We call a function f total if $f(x)$ is defined for every x.

A set of finite objects is called recursively enumerable if it is the domain of some computable function. It can be proved that a nonempty set A is recursively enumerable if and only if it is the range of some total recursive (computable) function.

Let A be a set of all finite objects of certain type. A function $f\colon A \to \mathcal{R} \cup \{+\infty\}$ is called lower semicomputable if there is a sequence of total functions $h_n : \{0,1\}^* \to \mathcal{Q}$ such that (i) $h_n(x) \le h_{n+1}(x)$ for all n and for all x, (ii) the function $h(n,x) = h_n(x)$ is computable, (iii) $f(x) = \lim_{n\to\infty} h_n(x)$ for each x.

This definition is equivalent to the following one. A function $f\colon A \to \mathcal{R} \cup \{+\infty\}$ is lower semicomputable if and only if the set

$$\{(r,x) : x \in \{0,1\}^*,\ r \in \mathcal{Q}, r < f(x)\}$$

is recursively enumerable. This means that there is an algorithm which when fed with a rational number r and a finite object x eventually stops if $r < f(x)$ and never stops otherwise. In other words, the semicomputability of f means that if $f(x) > r$ this fact will sooner or later be learned, whereas if $f(x) \le r$ we may be for ever uncertain.

A function $f\colon A \to \mathcal{R} \cup \{-\infty\}$ is called upper semicomputable if there is a sequence of total functions $q_n : \{0,1\}^* \to \mathcal{Q}$ such that (i) $q_n(x) \ge q_{n+1}(x)$ for all n and for all x, (ii) the function $q(n,x) = q_n(x)$ is computable, (iii)

$f(x) = \lim_{n\to\infty} q_n(x)$ for each x. This definition is equivalent to the following. A function f is upper semicomputable if and only if the set

$$\{(r,x) : x \in \{0,1\}^*,\ r \in \mathcal{Q}, r > f(x)\}$$

is recursively enumerable.

A function $f : \{0,1\}^* \to \mathcal{R}$ is called computable if it is lower semicomputable and upper semicomputable. It can be proved that there exists an algorithm which, given a finite sequence x and a rational number $\epsilon > 0$, computes a rational approximation of the number $f(x)$ with accuracy ϵ: given x and a rational $\epsilon > 0$ this algorithm finds an n such that $q_n(x) - h_n(x) < \epsilon$ and outputs $h_n(x)$ (or $q_n(x)$) as the result.

The topology on Ω is generated by intervals $\Gamma_x = \{\omega \in \Omega : x \subset \omega\}$, where x is a finite binary sequence. The Borel subsets of Ω can be defined using these intervals and set theoretic operations.

A probability measure P on Ω can be defined by the values $P(x) = P(\Gamma_x)$, where $x \in \{0,1\}^*$. Also, (i) $P(\lambda) = 1$ and (ii) $P(x) = P(x0) + P(x1)$ for every x. This function is further extended to all Borel subsets of Ω.

A measure P is computable if the function $x \to P(x)$ is computable. An example of computable probability measure is the uniform Bernoulli measure L, where $L(\Gamma_x) = 2^{-l(x)}$ for any finite binary sequence x.

An open subset U of Ω is called effectively open if it can be represented as a union of a computable sequence of intervals: $U = \bigcup_{i=1}^{\infty} \Gamma_{x_i}$, where $f(i) = x_i$ is a computable function. A sequence of effectively open sets U_n, $n = 1, 2, \ldots$, is called uniformly effectively open if each set U_n can be represented as $U_n = \bigcup_{i=1}^{\infty} \Gamma_{x_{n,i}}$, where $f(n,i) = x_{n,i}$ is a computable function from n and i.

Let P be a computable measure. Martin-Löf test of randomness with respect to P is an uniformly effectively open sequence U_n, $n = 1, 2, \ldots$, of effectively open sets such that $P(U_n) \le 2^{-n}$ for every n. It can be added the requirement $U_{n+1} \subseteq U_n$ for all n.[2]

An infinite binary sequence ω passes the test U_n, $n = 1, 2, \ldots$, if $\omega \notin \bigcap U_n$. Otherwise, it is rejected by this test. A sequence (point) ω is Martin-Löf random with respect to a computable measure P if it passes each Martin-Löf test of randomness.

[2] It is easy to that any test $\{U_n\}$ can be redefined as $U'_n = \cup_{i>n} U_i$ such that $U'_n) \le 2^{-n}$ and $U'_{n+1} \subseteq U'_n$ for all n.

3 Algorithmic ergodic theory

We confine our attention to the Cantor probability space $(\Omega, \mathcal{F}, P)$, where Ω is the set of all infinite binary sequences, $\mathcal{F}$ is the collection of all Borel subsets of Ω generated by intervals $\Gamma_x = \{\omega \in \Omega : x \subset \omega\}$ and P is a computable measure on Ω. [3]

Recall some basic notions of ergodic theory. An arbitrary measurable mapping of a probability space into itself is called transformation. A transformation $T : \Omega \to \Omega$ preserves a measure P on Ω if $P(T^{-1}(A)) = P(A)$ for all measurable subsets A of the space Ω. A subset A is called invariant with respect to T if $T^{-1}A = A$ up to a set of measure 0. A transformation T is called ergodic with respect to P if each subset A invariant with respect to T has measure 0 or 1.

An example of a transformation on Ω is the (left) shift

$$T(\omega_1\omega_2\omega_3\dots) = \omega_2\omega_3\dots$$

If the shift preserves a measure P then this measure is called stationary. A measure P is called ergodic if the shift is ergodic with respect to P.

The uniform measure L is an example of computable stationary and ergodic measure.

3.1 Poincare's recurrence theorem

Now, we give an example of the algorithmic analysis in terms of Martin-Löf randomness of the well-known statement of ergodic theory – the Poincare recurrence theorem, which states as follows:

Let T be a measure P preserving transformation and E be a measurable subset of Ω. Then for all $n > 0$ the set of all $\omega \in E$ such that $T^n\omega \notin E$ has measure 0. Equivalently, for almost all $\omega \in E$ it will be $T^n\omega \in E$ for some $n > 0$, i.e. trajectory of the point ω will visit E again. Moreover, it happens infinitely many times. [4]

We will formulate an algorithmically efficient analogue of this statement for the uniform measure L on the space Ω of all infinite binary sequences: $L(\Gamma_x) =$

---

[3] Hoyrup and Rojas [8] showed that any computable probability space is isomorphic to the Cantor space in both the computable and measure-theoretic senses. Therefore, there is no loss of generality in restricting to this case.

[4] This statement is nontrivial when the measure of the set E is positive.

$2^{-l(x)}$, and for the shift T. A set E is called effectively closed if it is the complement of some effectively open set.

An algorithmic version of Poincare's recurrence theorem is presented in the following theorem

Theorem 1 *Let T be a shift on Ω and E be an effectively closed set of positive measure. Then for any Martin-Löf random sequence $\omega \in E$ there will be $T^n\omega \in E$ for some $n > 0$. Moreover, this is true for infinitely many n.*

Theorem 1 will be a direct consequence of the following statement, which is attributed to Kuchera [15]. See also, Bienvenu et al. [2]. [5]

Proposition 1 *Let U be an effectively open set of uniform measure such that $L(U) < 1$. Then for any infinite sequence ω Martin-Löf random with respect to the measure L and for an arbitrary number N there is a number $n \geq N$ such that $T^n\omega \notin U$.*

Proof. Let N be an arbitrary positive integer number. By U^* denote the set of all $\omega \in \Omega$ such that $T^n\omega \in U$ for all $n \geq N$.

Let $L(U) < r$ for some rational number r such that $r < 1$. We represent an effectively open set U as a union of a computable sequence of pairwise disjoint intervals

$$U = \cup_i \Gamma_{x_i},$$

where the finite sequences x_i and x_j do not extend each other. Denote $U_1 = U$ and define

$$U_2 = \cup_{i,j}\Gamma_{x_i x_j},$$
$$U_3 = \cup_{i,j,s}\Gamma_{x_i x_j x_s}$$

etc. Here $x_i x_j$ is the concatenation of the strings x_i and x_j. $x_i x_j x_s$ is understood in a similar way. We have

$$L(U_2) = \sum_{i,j} L(\Gamma_{x_i x_j}) = \sum_{i,j} 2^{-l(x_i)-l(x_j)} =$$
$$= \sum_i 2^{-l(x_i)} \sum_j 2^{-l(x_j)} < r^2$$

etc. Similarly, we have $L(U_n) < r^n$ for all n.

[5] Bienvenu et al. [2] showed that in any computable probability space, a point is Martin-Löf random if and only if it is satisfied to the statement of the Poincare's recurrence theorem for each computable ergodic transformations with respect to effectively closed sets.

Let $\omega \in U^*$. Denote $\omega' = \omega_{N+1}\omega_{N+2}\ldots$. Since $\omega' = T^N\omega \in U$, $\omega' = x_i\omega''$ for some i and $\omega'' \in \Omega$. Since $\omega'' = T^{N+l(x_i)}\omega \in U$, we have $\omega'' = x_j\omega'''$ for some j and $\omega''' \in \Omega$. Now, $\omega' = x_ix_j\omega'''$, and then $\omega' \in U_2$. Similarly, we obtain $\omega' \in U_3$ and etc.

The sequence of the effectively open sets $\{U_m : m = 1, 2, \ldots\}$ defines a Martin-Löf test of randomness. It was proved that $\omega' \in \cap_m U_m$, i.e. ω' is not Martin-Löf random. It is easy to see that in this case the original sequence ω is also not random. In particular, $U^* \subseteq \cap_m U_m$. The proposition is proved. □

To prove Theorem 1, one should take $U = \Omega \setminus E$ in Proposition 1. Since $L(E) > 0$, the inequality $L(U) < 1$ holds. By Proposition 1 $T^n\omega \notin U$ (equivalently, $T^n\omega \in E$) for infinitely many n. □

3.2 Algorithmically effective convergence

First, define the notion of computable function of type $f : \Omega \to \mathcal{R}$.

A function $h : \Omega \to \mathcal{R}^+$ is called simple if its domain Ω can be represented as a union of finite set of intervals: $\Omega = \cup_{i=1}^k \Gamma_{x_i}$, where $x_i \in \Xi$, and it takes constant rational values on each such interval: $h(\omega) = r_i \in \mathcal{Q}$ for each $\omega \in \Gamma_{x_i}$ for $1 \le i \le k$. Any simple function is a finite (constructive) object, and so, we can effectively identify all simple functions and positive integer numbers. [6]

A function $f : \Omega \to \mathcal{R} \cup \{+\infty\}$ is called lower semicomputable if there exists a recursively enumerable non-decreasing sequence h_n of simple functions: $h_n(\omega) \le h_{n+1}(\omega)$ for all n, such that $f(\omega) = \lim_{n\to\infty} h_n(\omega)$ for each $\omega \in \Omega$. Similarly, a function $f : \Omega \to \mathcal{R}$ is called upper semicomputable if there exists an recursively enumerable non-increasing sequence h_n of simple functions: $h_n(\omega) \ge h_{n+1}(\omega)$ for all n, such that $f(\omega) = \lim_{n\to\infty} h_n(\omega)$ for each $\omega \in \Omega$.

A function $f : \Omega \to \mathcal{R}$ is called computable if it is lower semicomptable and upper semicomptable. It is easy to see that in this case there is an algorithm which given an infinite sequence ω and a rational ϵ computes a rational approximation of $f(\omega)$ up to ϵ using some prefix of ω.

A sequence of functions $f_i : \Omega \to \mathcal{R} \cup \{+\infty\}$, $i = 1, 2, \ldots$, is called uniformly lower semicomputable if there exists a recursively enumerable non-decreasing by n sequence $h_{i,n}$ of simple functions: $h_{i,n}(\omega) \le h_{i,n+1}(\omega)$ for all i and n such that for any i, $f_i(\omega) = \lim_{n\to\infty} h_{i,n}(\omega)$ for each $\omega \in \Omega$. The definition of uniformly upper semicomputable sequence of function $f_i : \Omega \to \mathcal{R}$ is similar.

[6] This means that there is an algorithm which given any number n can reconstruct the sets $\{x_1, \ldots, x_k\}$ and $\{r_1, \ldots, r_k\}$. Also, such correspondence is one-to-one.

A sequence of computable functions $f_n : \Omega \to \mathcal{R}$, $n = 1, 2, \ldots$, is called uniformly computable it is uniformly lower semicomputable and uniformly upper semicomputable.

The algorithmic effective version of convergence in probability and almost surely of functions f_n of type $\Omega \to \mathcal{R}$ was considered by V'yugin [19].

Let P be a probability measure. A sequence of functions f_n converges to a function f in probability if for each $\delta > 0$

$$P\{\omega \in \Omega : |f_n(\omega) - f(\omega)| > \delta\} \to 0$$

as $n \to \infty$. This is equivalent to the fact that there is a function $m(\delta, \epsilon)$ such that

$$P\{\omega : |f_n(\omega) - f(\omega)| > \delta\} < \epsilon \tag{1}$$

for all $n \geq m(\delta, \epsilon)$ for each positive numbers δ and ϵ. The function $m(\delta, \epsilon)$ is called regulator of convergence.

A sequence of functions f_n effectively converges in probability to a function f if there exists a computable regulator of this convergence.

A sequence of functions f_n converges to a function f almost surely if $\lim_{n\to\infty} f_n(\omega) = f(\omega)$ for almost every $\omega \in \Omega$ (see [18]).

This definition is equivalent to the following. A sequence of functions f_n converges to a function f almost surely if a function $m(\delta, \epsilon)$ (regulator of convergence) exists such that

$$P\{\omega : \sup_{n \geq m(\delta,\epsilon)} |f_n(\omega) - f(\omega)| > \delta\} < \epsilon \tag{2}$$

for every positive rational numbers δ and ϵ.

A sequence of functions f_n efectively converges a function f almost surely if there exists a computable regulator of this convergence.

The following simple proposition was proved by V'yugin [19].

Proposition 2 *Let P be a computable measure and a uniformly computable sequence of functions $f_n : \Omega \to \mathcal{R}$ effectively converges almost surely to some function f. Then*

$$\lim_{n\to\infty} f_n(\omega) = f(\omega)$$

for each sequence ω Martin-Löf random with respect to P.

Proof. By (2) we have $P\{\omega : \sup\limits_{n,n'\geq m(\delta/2,\epsilon)} |f_n(\omega) - f_{n'}(\omega)| > \delta\} < \epsilon$ for every positive rational numbers δ and ϵ, where $m(\delta,\epsilon)$ is a computable function.

Denote $W_{n,n',j} = \{\omega : |f_n(\omega) - f_{n'}(\omega)| > \frac{1}{j}\}$. Since the sequence f_n is uniformly computable, this set is effectively open. Define

$$V_j = \bigcup_{n,n'\geq m(\frac{1}{2j},2^{-j})} W_{n,n',j}$$

for all j. Define also, $U_i = \bigcup_{j>i} V_j$. Then $P(U_i) \leq 2^{-i}$ for all i. Therefore, $\{U_i\}$ is the Martin-Löf test of randomness.

Assume that $\lim\limits_{n\to\infty} f_n(\omega)$ does not exist for some ω. Then a number i exists such that $|f_n(\omega) - f_{n'}(\omega)| > 1/i$ for infinitely many n and n'. For any $j > i$ the numbers $n, n' \geq m(\frac{1}{2j}, 2^{-j})$ exist such that $\omega \in W_{n,n',j} \subseteq V_j \subseteq U_i$. Since i is arbitrary, $\omega \in U_i$ for every i. Hence, the sequence ω is rejected by the test $\{U_i\}$. □

Now we show details of how Proposition 2 can be applied to prove that the strong law of large numbers holds for any sequence Martin-Löf random with respect to the uniform measure L.

Hoeffding's [7] inequality

$$L\left\{\omega \in \Omega : \left|\frac{1}{n}\sum_{i=1}^{n}\omega_i - \frac{1}{2}\right| \geq \delta\right\} \leq 2e^{-2n\delta^2}$$

implies

$$L\left\{\omega \in \Omega : \sup_{n\geq m(\epsilon,\delta)}\left|\frac{1}{n}\sum_{i=1}^{n}\omega_i - \frac{1}{2}\right| \geq \delta\right\} \leq 2e^{-2n\delta^2 + \ln\frac{1}{\delta}}.$$

Define $m(\epsilon,\delta) = \left\lfloor \frac{\ln\frac{2}{\epsilon\delta}}{2\delta^2} \right\rfloor$ for any rational $\epsilon > 0$ and $\delta > 0$. Then we obtain the algorithmically efficient almost sure convergence

$$L\left\{\omega \in \Omega : \sup_{n\geq m(\epsilon,\delta)}\left|\frac{1}{n}\sum_{i=1}^{n}\omega_i - \frac{1}{2}\right| \geq \delta\right\} \leq \epsilon$$

for all $\epsilon > 0$ and $\delta > 0$.

By Proposition 2 $\lim\limits_{n\to\infty}\frac{1}{n}\sum\limits_{i=1}^{n}\omega_i = \frac{1}{2}$ holds for each Martin-Löf random sequence $\omega = \omega_1\omega_2\dots$.

3.3 Ergodic theorem for Martin-Löf random sequences for ergodic transformations

Let us define the notion of a computable transformation of binary sequences. A computable representation of a transformation is a set $\hat{T} \subseteq \{0,1\}^* \times \{0,1\}^*$ such that

- (i) the set $\hat{T}$ is recursively enumerable;
- (ii) for any $(x,y),(x',y') \in \hat{T}$, if $x \subseteq x'$, then $y \subseteq y'$ or $y' \subseteq y$;
- (iii) if $(x,y) \in \hat{T}$, then $(x,y') \in \hat{T}$ for all $y' \subseteq y$;
- (iv) $(x,\lambda) \in \hat{T}$ for every x.

A transformation T of the set Ω is computable if a computable representation $\hat{T}$ exists such that (i)-(iv) hold and

$$T(\omega) = \sup\{y : x \subseteq \omega \& (x,y) \in \hat{T})\}$$

for all infinite sequence $\omega \in \Omega$, where sup is with respect to the partial ordering $x \subseteq x'$.

Denote $T^0\omega = \omega$, $T^{i+1}\omega = T(T^i\omega)$, so, any point $\omega \in \Omega$ generates the infinite trajectory $\omega, T\omega, T^2\omega, \dots$.

Let P be a computable measure, T be a computable ergodic transformation preserving the measure P. and $f \in L^1$ be a computable function of type $\Omega \to \mathcal{R}$ (observable). By $\|f\|$ denote the norm in L^1. Assume that $\sup_\omega |f(\omega)| < \infty$.

Consider the sequence of ergodic time averages S_n^f, $n = 1, 2, \dots$, where

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

Galatolo et al. [9] and Avigad et al. [1] showed that if the measure preserving transformation T is ergodic then the time averages $\{S_n^f\}$ effectively converge to a computable real number $c = \int f(\omega)dP$ almost surely as $n \to \infty$. We present details of this result for completeness of exposition.

Proposition 3 *Let P be a computable measure and T be a computable ergodic transformation preserving the measure P. Then the sequence of time averages $\{S_n^f\}$ effectively converges P-almost surely as $n \to \infty$.*

Proof. We assume without loss of generality that $\int f dP = 0$.[7] The sequence $\|S_n^f\|$ converges to 0 by the classical ergodic theorem.

[7] Replace f with $f - \int f(\omega)dP$ otherwise.

The maximal ergodic theorem (see Bilingsly [3]) says that

$$P\{\omega : \sup_n |S_n^f(\omega)| > \delta\} \leq \frac{1}{\delta}\|f\|$$

for any measure P and for any measure preserving ergodic transformation T.

Given $\epsilon, \delta > 0$ compute a $p = p(\delta, \epsilon)$ such that $\|S_p^f\| \leq \delta\epsilon/2$. By the maximal ergodic theorem for $g = S_p^f$ we have

$$P\{\omega : \sup_n |S_n^g(\omega)| > \delta/2\} \leq \frac{2}{\delta}\|S_p^f\| \leq \epsilon.$$

Now we check that $S_n^g(\omega)$ is not too far from $S_n^f(\omega)$. Expanding $S_n^g(\omega)$, one can check that

$$S_n^g(\omega) = \frac{1}{n}\sum_{k=0}^{n-1} g(T^k\omega) = \frac{1}{np}\sum_{k=0}^{p-1}\sum_{s=0}^{n-1} f(T^{k+s}\omega) = \frac{1}{np}\left(p\sum_{k=0}^{n-1} f(T^k\omega)\right) +$$

$$+\frac{1}{np}\left(\sum_{k=1}^{p-1}(p-k)f(T^{k+n-1}\omega) - \sum_{k=1}^{p-1}(p-k)f(T^{k-1}\omega)\right).$$

This implies that

$$\sup_\omega |S_n^g(\omega) - S_n^f(\omega)| \leq$$

$$\frac{2}{np}\sum_{k=1}^{p-1}(p-k)\sup_\omega |f(\omega)| < \frac{p-1}{n}\sup_\omega |f(\omega)| \leq \delta/2$$

for all $n \geq m(\delta, \epsilon)$, where

$$m(\delta, \epsilon) = 2(p(\delta, \epsilon) - 1)r/\delta$$

and r is the rational number such that $r > \sup_\omega |f(\omega)|$.

If $|S_n^f(\omega)| > \delta$ for some $n \geq m(\delta, \epsilon)$ then $|S_n^g(\omega)| > \delta/2$. Hence,

$$P\{\omega : \sup_{n \geq m(\delta,\epsilon)} |S_n^f(\omega)| > \delta\} \leq \epsilon,$$

where $m(\delta, \epsilon)$ is the computable function. The proposition is proved. □

Corollary 1 *Let P be a computable measure and T be a computable ergodic transformation preserving the measure P. Then for any sequence ω Martin-Löf random with respect to P the time average $S_n^f(\omega)$ converges to the number $\int f(\alpha)dP$ as $n \to \infty$.*

Proof. This corollary follows directly from Propositions 2 and 3. □

3.4 Lack of a computable rate of convergence in the ergodic theorem for stationary non-ergodic measure

In this section, we show that in general case there is no computable rate of convergence of time averages in the ergodic theorem. We give an example of a computable measure and a measure preserving transformation for which the convergence of averages S_n^f in probability and almost surely in Birkhoff's theorem is not algorithmically efficient.

Theorem 2 *There is a computable measure P and a measure preserving transformation for which there is no computable rate of the convergence in probability of time averages S_n^f as $n \to \infty$, where $f(\omega) = \omega_1$.*[8]

Proof. Let T be the shift on Ω, $f(\omega) = \omega_1$ for $\omega \in \Omega$. Then $S_n^f(\omega) = S_n(\omega) = \frac{1}{n}\sum_{i=1}^{n} \omega_i$.

We will construct a computable stationary measure P as a mixture of homogeneous stationary Markov measures P_i, $i = 1, 2, \ldots$. Each measure P_i will be computable and will contain information about the stopping problem of the universal algorithm.

Let $U(i, \delta, \epsilon)$ be a computable function universal for all computable functions of two arguments $m(\delta, \epsilon)$. By universality property of $U(i, \delta, \epsilon)$ for any computable function $m(\delta, \epsilon)$ a number i exists such that $m(\delta, \epsilon) = U(i, \delta, \epsilon)$ for all δ and ϵ(see Rogers [17]).

We will construct an example of such a measure P using the diagonal argument. Let

$$U^s(i, \delta, \epsilon) = \begin{cases} U(i, \delta, \epsilon) \text{ if the process of computation terminaes in} \\ \leq s \text{ steps,} \\ \text{ undefined, otherwise.} \end{cases}$$

For any i. Define the real number α_i by setting the bits of its binary expansion:

$$\alpha_i = 0.\alpha_{i1}\alpha_{i2}\ldots,$$

[8] This limit exists P-almost surely by the classical Birkhoff's ergodic theorem.

where

$$\alpha_{i,s} = \begin{cases} 1, \text{ if } u = U^s(i, \frac{1}{4}, 2^{-(i+1)}) \text{ is defined end } s > u, \\ 0 \text{ otherwise.} \end{cases}$$

It is easy to see that the value of each bit of $\alpha_{i,s}$ is computable by i and s. Besides, $\alpha_i > 0$ if and only if the value of $U(i, \frac{1}{4}, 2^{-(i+1)})$ is defined. Thus, the real number α_i is an indicator of the stopping problem at the inputs $\delta = \frac{1}{4}$ and $\epsilon = 2^{-(i+1)}$.

By definition if $\alpha_i > 0$ then binary decomposition of the number α_i consists of a block of zeros followed by ones. In this case $\alpha_i = 2^{-k(i)}$, where $k(i)$ is length of the block of zeros.

Let us define a homogeneous Markov chain for an arbitrary i by specifying the initial probabilities:

$$P_i\{\omega_1 = 0\} = P_i\{\omega_1 = 1\} = \frac{1}{2}$$

and the transition probabilities:

$$P_i\{\omega_{s+1} = 0|\omega_s = 1\} = P_i\{\omega_{s+1} = 1|\omega_s = 0\} = \alpha_i$$

for each $s = 1, 2, \ldots$. It is easy to show that the probability measure P_i generated by the given initial and transition probabilities is stationary. Moreover, it is computable.

According to the theory of Markov chains (see Shiryaev [18]), for $\alpha_i > 0$ this measure is also ergodic. If $\alpha_i = 0$, the measure P_i is concentrated on only two infinite sequences: $P_i(0^\infty) = P_i(1^\infty) = \frac{1}{2}$. The sets $\{0^\infty\}$ and $\{1^\infty\}$ are shift-invariant, then if $\alpha_i = 0$ the measure P_i is ergodic.

Each measure P_i is computable, and, moreover, there is an algorithm that computes the value of $P_i(x)$ uniformly in i and x. Define the measure

$$P(x) = \sum_{i=1}^{\infty} 2^{-i} P_i(x).$$

It is easy to prove that the measure P is computable. Since each measure P_i is stationary, the measure P is also stationary. It is clear from the definition that this measure is not ergodic.

By definition $S_n^f(\omega) = S_n(\omega) = \frac{1}{n}\sum_{i=1}^n \omega_i$. By Birkhoff's ergodic theorem applied for the shift, for P-almost all ω there exists the limit $\lim\limits_{n\to\infty} S_n(\omega)$ as $n \to \infty$.

Let $m(\delta, \epsilon)$ be an arbitrary computable function, which is a candidate for the regulator of convergence in probability of the time averages for the measure P. By universality of the function U, an i exists such that $m(\delta, \epsilon) = U(i, \delta, \epsilon)$ for every δ and ϵ. In this case $\alpha_i > 0$.

By the ergodic theorem for Markov processes, the stationary distribution for the Markov process generated by the measure P_i for $\alpha_i > 0$, is $\pi_0 = \frac{1}{2}$ $\pi_1 = \frac{1}{2}$. By the law of large numbers

$$P_i\{\omega : |S_n(\omega) - 1/2| < 0.01\} \to 1 \qquad (3)$$

as $n \to \infty$.

By definition, the number $k(i)$ is equal to the position number of the last zero in the binary representation of the number α_i, after which there are ones in this representation. It is easy to see that $\alpha_i = 2^{-k(i)}$.

Let us estimate the probabilities $P_i(0^{k(i)})$ and $P_i(1^{k(i)})$. By definition

$$P_i(0^{k(i)}) = P_i(1^{k(i)}) = \frac{1}{2}(1 - \alpha_i)^{k(i)-1} > \frac{2}{5}$$

for all sufficiently large $k(i)$. [9] Hence,

$$P_i\{\omega : S_{k(i)}(\omega) = 0 \text{ or } 1\} > \frac{4}{5}.$$

By definition $k(i) > m(\frac{1}{4}, 2^{-(i+1)})$. From here and from (3) it follows that there is an $n > m(\frac{1}{4}, 2^{-(i+1)})$ large enough such that

$$P_i\{\omega : |S_{k(i)}(\omega) - S_n(\omega)| > \frac{1}{4}\} > \frac{1}{2}.$$

Then P-measure of this set is more than $2^{-i} \cdot \frac{1}{2} = 2^{-(i+1)} = \epsilon$, i.e. the numbers $k(i)$ and n do not satisfy the requirement (1) for the convergence rate regulator.

The resulting contradiction proves the theorem. □

Since algorithmically efficient convergence almost surely implies algorithmically efficient convergence in probability, we obtain the following corollary from the theorem 2.

[9] Without loss of generality, we can assume that all steps s, at which some value of the universal function was first defined, are greater than some fixed value s_0.

Corollary 2 *There is a computable stationary measure P for which there is no computable regulator for the convergence almost surely for the time averages S^f_n.*

4 Birghoff's ergodic theorem for Martin-Löf random sequences

Let P be a measure, T be a measure preserving transformation, and f be an integrable function (observable). The classical Birkhoff's ergodic theorem says that for P-almost every points ω

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

where $\tilde{f}$ is an integrated and invariant with respect to T function such that $\int \tilde{f}(\omega)dP = \int f(\omega)dP$. [10]

Using Bishop's [4] analysis, V'yugin [19], [20] presented an algorithmic version of Birkhoff's pointwise ergodic theorem. Later this result was extended to a more general spaces by Hoyrup and Rojas [8], Galatolo et al. [9], Gacs et al. [6]

We will present the proof of this statement for Martin-Löf random points in Section 4.2. This proof is based on the Bishop's analysis of the Birghoff's theorem and on a notion of integral test of randomness, which will be defined in Section 4.1.

4.1 Integral tests of randomness

For further presentation, we need one more type of a test of randomness – integral tests.

Let P be a computable measure on Ω. A lower semicomputable function $f : \Omega \to \mathcal{R}_+ \cup \{+\infty\}$ is called integral test of randomness with respect to a computable measure P (integral P-test) if

$$E_P[f] = \int f(\omega)dP \leq 1.$$

Here E_P denotes the mathematical expectation with respect to P.

[10] Also, $\hat{f}(\omega) = E_P[f]$ for almost every ω if the transformation T is ergodic.

From the definition, for any integral test, the Markov inequality holds:

$$P\{\omega : f(\omega) > r\} < \frac{1}{r}$$

for each r. In particular, $f(\omega) < \infty$ for almost every ω.

Integral tests can be used to give an equivalent definition of the Martin-Löf random sequence.

Theorem 3 *Let P be a computable measure. An infinite binary sequence ω is Martin-Löf random with respect to the measure P if and only if $p(\omega) < \infty$ for each martingal test of randomness $p(\omega)$.*

Proof. Given a martingal test of randomness $p(\omega)$ define the Martin-Löf test of randomness

$$U_m = \{\omega : p(\omega) > 2^m\}$$

for each m.

By Markov inequality $P(U_m) \leq 2^{-m}$ for each m. Since the lower semicomputable function $p(\omega)$ can be represented as $p(\omega) = \lim_{n\to\infty} f_n(\omega)$, where f_n is a recursively enumerable non-decreasing sequence of simple functions, it holds

$$U_m = \{\omega \in \Omega : \exists n(f_n(\omega) > 2^m)\}.$$

Hence, the set U_m is (uniformly) effectively open.

If $p(\omega) = \infty$ then $\omega \in \cap_n U_n$, i.e., the sequence ω is not Martin-Löf random.

Let us prove the converse statement. Let $\{U_m\}$ be a Martin-Löf test of randomness. Define a sequence of characteristic functions

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

The sequence of functions $p_m(\omega)$ is uniformly lower semicomputable, since $\{U_m\}$ is uniformly effectively open. Define

$$p(\omega) = \sum_{m=1}^{\infty} p_m(\omega)$$

for any $\omega \in \Omega$. The function $p(\omega)$ is lower semicomputable and

$$\int p(\omega)dP = \sum_{m=1}^{\infty} \int p_m(\omega)dP = \sum_{m=1}^{\infty} P(U_m) \leq \sum_{m=1}^{\infty} 2^{-m} = 1.$$

Therefore, the function $p(\omega)$ is the integral P-test of randomness.

If $\omega \in \cap_m U_m$ then $p(\omega) = \infty$. Theorem is proved. $\square$

4.2 Effective version of the Birghoff's ergodic theorem

The effective version of Birkhoff's ergodic theorem will be considered for computable measure, measure preserving computable transformation, and for computable observable.

The formulation of Birkhoff's ergodic theorem for Martin-Löf random sequences is obtained from the original formulation by replacing the expression "for P -almost every ω" by "for each sequence ω Martin-Löf random with respect to the measure P".

Theorem 4 *Let P be a computable measure on Ω and f be an arbitrary computable integrable function of the type $\Omega \to \mathcal{R}_+$. Then for any measure preserving computable transformation T*

$$\lim_{n\to\infty} \frac{1}{n}\sum_{k=0}^{n-1} f(T^k\omega) = \tilde{f}(\omega) \tag{4}$$

for each Martin-Löf random sequence ω, where $\tilde{f}$ is an integrable invariant with respect to T function such that $\int \tilde{f}(\omega)dP = \int f(\omega)dP$.

Moreover, if the transformation T is ergodic then $\tilde{f}(\omega) = \int f(\alpha)dP$ for each random ω.

Proof. For an arbitrary infinite sequence ω, consider the time average

$$s_m(\omega) = S^f_{m+1}(\omega) = \frac{1}{m+1}\sum_{k=0}^{m} f(T^k\omega).$$

For convenience, in the further reasoning, we assume that $s_{-1}(\omega) = 0$.

Assume that $\int |f(\omega)|dP \le M$, where M is a positive integer number.

If the limit $\lim\limits_{m\to\infty} s_m(\omega)$ does not exist then there exist the rational numbers $\alpha < \beta$ such that $-M < \alpha < \beta < M$ and

$$\liminf_{m\to\infty} s_m(\omega) < \alpha < \beta < \limsup_{m\to\infty} s_m(\omega).$$

The converse is also true.

Let α and β be rational numbers such that $-M < \alpha < \beta < M$. Define the boundary crossing function $\sigma_n(\omega|\alpha,\beta)$ as follows. The value $\sigma_n(\omega|\alpha,\beta)$ is equal to the number of upward intersections of the interval (α,β) by the sequence $s_0(\omega), s_1(\omega), \ \dots, \ s_n(\omega)$. More precisely, we define

$$\begin{aligned} u_0 &= 0,\\ u_1 &= \min\{m : m \geq u_0, s_m(\omega) < \alpha\},\\ v_1 &= \min\{m : m > u_1, s_m(\omega) > \beta\},\\ &\dots\\ u_i &= \min\{m : m > v_{i-1}, s_m(\omega) < \alpha\},\\ v_i &= \min\{m : m > u_i, s_m(\omega) > \beta\},\\ &\dots\\ u_k &= \min\{m : m > v_{k-1}, s_m(\omega) < \alpha\},\\ v_k &= \min\{m : m > u_k, s_m(\omega) > \beta\}. \end{aligned}$$

Define the function

$$\sigma_n(\omega|\alpha,\beta) = \begin{cases} 0 \text{ if } v_1 > n,\\ \max\{k : v_k \leq n\} \text{ if } v_1 \leq n. \end{cases}$$

The value $\sigma_n(\omega|\alpha,\beta)$ is equal to the maximum number of upward crossings of the interval (α,β) by the average $s_m(\omega)$ for $m = 0, 1, \dots n$. The function $\sigma_n(\omega|\alpha,\beta)$ is uniformly lower semicomputable with respect to the arguments n, α, β.

It is easy to see that the limit $\lim\limits_{m\to\infty} s_m(\omega)$ does not exist if and only if $\sup_n \sigma_n(\omega|\alpha,\beta) = \infty$ for some α and β such that $\alpha < \beta$.

We temporarily fix the infinite sequence ω, as well as the positive integer number n and the rational numbers α and β such that $\alpha < \beta$.

Consider the non-relative deviations

$$\begin{aligned} a(u,\omega) &= \sum_{s=0}^{u}(f(T^s\omega) - \alpha),\\ b(v,\omega) &= \sum_{s=0}^{v}(f(T^s\omega) - \beta). \end{aligned}$$

It will be convenient for us to assume that $a(-1,\omega) = 0$.

Oscillation of relative frequencies entails oscillation of non-relative deviations.

A sequence $d = \{u_1, v_1, \ \ldots, \ u_k, v_k\}$ of integer numbers is called admissible if

$$-1 \leq u_1 < v_1 \leq u_2 < v_2 \leq \ldots \leq u_k < v_k \leq n.$$

The number of pairs in the admissible sequence d will be denoted by m_d ($m_d = k$) and will be called its length.

For any admissible sequence

$$d = \{s_1, t_1, \ldots, s_k, t_k\},$$

consider the cumulative sum of the differences of non-relative deviations:

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

The key role in the proof of the theorem is played by the following combinatorial lemma on lengthening an admissible sequence without decreasing the cumulative sum.

Lemma 1 *For each admissible sequence q, there exists an admissible sequence d such that $m_d \geq \sigma_n(\omega|\alpha, \beta)$ and $S(d, \omega) \geq S(q, \omega)$.*

Proof. Denote by $N = \sigma_n(\omega|\alpha, \beta)$ the maximum number of intersections of the interval (α, β) by a sequence of averages $s_0(\omega), \ldots \ , s_n(\omega)$. Let

$$p = \{-1 < u_1 < v_1 < u_2 < v_2 < \ldots < u_N < v_N \leq n\}$$

be that admissible sequence of length N by which the value $\sigma_n(\omega|\alpha, \beta)$ was determined: $N = \sigma_n(\omega|\alpha, \beta)$.

It suffices to prove that for any admissible sequence q of length $m_q < N$ an admissible sequence d exists such that $m_d = m_q + 1$ and $S(d, \omega) \geq S(q, \omega)$. We will use some elements of the sequence p to construct such d.

Let an admissible sequence q be given:

$$-1 \leq s_1 < t_1 \leq s_2 < t_2 \leq \ldots \leq s_m < t_m \leq n,$$

where $m = m_q < N$. We expand it by one pair of elements. Consider an auxiliary element $s_{m+1} = n$. Since $m + 1 \leq N$, the element v_{m+1} is presented in the sequence p.

Besides, $v_{m+1} \leq n = s_{m+1}$. Therefore, there is the smallest i such that $v_i \leq s_i$. If $i = 1$ then define

$$d = \{u_1, v_1, s_1, t_1, \ \ldots, \ s_m, t_m\}. \tag{5}$$

The length of the admissible sequence q has increased by one.

Consider the case $i > 1$. Then $v_{i-1} > s_{i-1}$ and the inequality

$$s_{i-1} < v_{i-1} < u_i < v_i \leq s_i$$

is valid. If $u_i < t_{i-1}$ define

$$d = \{s_1, t_1, \ \ldots, \ s_{i-1}, v_{i-1}, u_i, t_{i-1}, \ \ldots, \ s_m, t_m\}. \tag{6}$$

If $u_i \geq t_{i-1}$ define for $i \leq m$

$$d = \{s_1, t_1, \ldots \ , s_{i-1}, t_{i-1}, u_i, v_i, s_i, t_i, \ldots \ , s_m, t_m\}, \tag{7}$$

for $i = m + 1$ define

$$d = \{s_1, t_1, \ \ldots, \ s_m, t_m, s_{m+1}, t_{m+1}\}. \tag{8}$$

The sequence d is admissible and its length has increased by one: $m_d = m_q + 1$. It remains to check how the cumulative sums have changed for all variants of this definition.

By definitions (5), (7) and (8)

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

and the added term is positive. If d was defined by (6) then

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

By the definition of the sequence $\{u_1, v_1, \ \ldots, \ u_N, v_N\}$ $s_{v_{i-1}}(\omega) > \beta$ and $s_{u_i}(\omega) < \alpha$. Then $b(v_{i-1}, \omega) > 0 > a(u_i, \omega)$ and the added term is also positive.

Therefore, in both cases the cumulative sum increases: $S(\omega, d) > S(\omega, q)$. Lemma is proved. $\square$

Let $d = \{s_1, t_1, \ \ldots, \ s_m, t_m\}$ be an admissible sequence of length $m_d = m$ and $S(\omega, d)$ be the corresponding cumulative sum.

Let us apply the transformation T to the sequence ω and show how the cumulative sum changes. First, when $s_i \geq 0$, the following changes occur:

$$a(s_i, \omega) = a(s_i - 1, T\omega) + f(\omega) - \alpha,$$
$$b(t_i, \omega) = b(t_i - 1, T\omega) + f(\omega) - \beta.$$

From this and from the definition of the cumulative sum, we obtain

$$S(\omega, d) = S(T\omega, d') + a - (\beta - \alpha)m_d, \tag{9}$$

where

$$d' = \{s_1 - 1, t_1 - 1, \ \ldots, \ s_m - 1, t_m - 1\}$$

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

$$d' = \{s_2 - 1, t_2 - 1, \ \ldots, \ s_m - 1, t_m - 1\}.$$

In the sum (9), $a = 0$ if $s_1 \geq 0$ and $a = f(\omega) - \alpha$ if $s_1 = -1$.

Let us introduce the lower semicomputable function

$$\lambda_n(\omega) = \sup\{S(\omega, d) : d - \text{admissible sequence}\}.$$

Then by (9) we obtain

$$S(\omega, d) \leq \lambda_n(T\omega) + (f(\omega) - \alpha)^+ - (\beta - \alpha)m_d, \tag{10}$$

where $h^+ = \max\{h, 0\}$.

By Lemma 1 for any admissible sequence q, an admissible sequence d exists such that $m_d \geq \sigma_n(\omega|\alpha, \beta)$ and $S(\omega, q) < S(\omega, d)$. Then by (10)

$$\begin{aligned} & S(\omega, q) < S(\omega, d) \leq \\ & \leq \lambda_n(T\omega) + (f(\omega) - \alpha)^+ - (\beta - \alpha)\sigma_n(\omega|\alpha, \beta), \end{aligned} \tag{11}$$

We take in (11) maximum by q and get

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

Therefore,

$$(\beta - \alpha)\sigma_n(\omega|\alpha, \beta) \leq (f(\omega) - \alpha)^+ + \lambda_n(T\omega) - \lambda_n(\omega). \tag{12}$$

Integrating the inequality (12), we obtain

$$\int (\beta - \alpha)\sigma_n(\omega|\alpha, \beta)dP \leq \int (f(\omega) - \alpha)^+ dP. \tag{13}$$

Here, we use the assumption that the transformation T preserves the measure P. This assumption implies that

$$\int \lambda_n(T\omega)dP = \int \lambda_n(\omega)dP.$$

Since the integral of the function $|f(\omega)|$ is bounded by the number M,

$$\int (f(\omega) - \alpha)^+ dP \leq 2M.$$

Define

$$\sigma(\omega|\alpha, \beta) = \sup_n \sigma_n(\omega|\alpha, \beta).$$

It is easy to see that this function is lower semicomputable. In addition, since

$$\sigma_n(\omega|\alpha, \beta) \leq \sigma_{n+1}(\omega|\alpha, \beta)$$

for all n, this function is integrable and by (13)

$$\int (2M)^{-1}(\beta - \alpha)\sigma(\omega|\alpha, \beta)dP \leq 1$$

for each α and β such that $\alpha < \beta$.

By averaging the quantity $\sigma(\omega|\alpha, \beta)$ we can define an integral test of randomness as follows. Let the computable functions $\alpha(i)$ and $\beta(i)$ enumerate the set of all pairs of rational numbers $\{(\alpha, \beta) : -M < \alpha < \beta < M\}$. Define

$$p(\omega) = \frac{1}{2M} \sum_{i=1}^{\infty} \frac{1}{i(i+1)} (\beta(i) - \alpha(i))\sigma(\omega|\alpha(i), \beta(i)).$$

By definition the function $p(\omega)$ is lower semicomputable and

$$\int p(\omega)dP \leq 1,$$

i.e., it is an integral test of randomness with respect to the measure P. In addition, as previously noted, if the limit of averages $\lim\limits_{n\to\infty} s_n(\omega)$ does not exist then rational numbers α and β exist such that $\alpha < \beta$ and $\sigma(\omega|\alpha, \beta) = \infty$. In this case $p(\omega) = \infty$.

Therefore, for any infinite binary sequence ω the implication is true:

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

The main part of the theorem is proved.

Denote by $\tilde{f}(\omega)$ the limit of averages (4). It is easy to see that $\tilde{f}(\omega)$ is defined and $\tilde{f}(T\omega) = \tilde{f}(\omega)$ for almost every ω.

If the transformation T is ergodic then $\tilde{f}(\omega) = c$ for P-almost all ω, where $c = \int f(\omega)dP$ is the constant.

We need to prove the following statement.

Lemma 2 $\tilde{f}(\omega) = \int f(\omega)dP$ *for each sequence* ω *random with respect to the measure* P.

Proof. Assume that this assertion is violated. Then a random sequence ω exists such that $\tilde{f}(\omega) = d \neq c$. Take rational numbers r_1 and r_2 such that $r_1 < d < r_2$ and $c \leq r_1$ or $c \geq r_2$ and define

$$S_n = \{\alpha : r_1 < s_n(\alpha) < r_2\},$$
$$\bar{S}_n = \{\alpha : r_1 \leq s_n(\alpha) \leq r_2\}.$$

Since the limit (4) is equal to c almost surely, $P(\bar{S}_n) \to 0$ as $n \to \infty$. The function $P(\bar{S}_n)$ is upper semicomputable (by n), since

$$r > P(\bar{S}_n) \Leftrightarrow 1 - r < P\{\alpha : r_1 > s_n(\alpha) \text{ or } r_1 < s_n(\alpha)\}.$$

Therefore, using any m, we can efficiently find an $n \geq m$ such that $P(\bar{S}_n) < 2^{-m}$.

By definition the set S_n is effectively open. We have $P(S_n) \leq P(\bar{S}_n) < 2^{-m}$. Define $U_m = S_n$ for such n. The family $\{U_m\}$ of effectively open sets defines the Martin-Löf test of randomness with respect to the measure P.

It holds $\omega \in \cap_{m=1}^{\infty} U_m$, i.e., ω is not Martin-Löf random.

The resulting statement proves the lemma and Theorem 4. $\square$

Later a converse result was obtained by Franklin and Towsner [5]. Using the cutting and stacking method, they showed that for every infinite sequence ω, which is not Martin-Löf random with respect to the uniform measure L the measure preserving transformation T can be constructed such that the limit (4) does not exist.

References

[1] Avigad, J., Gerhardy, P., Towsner H.: Local stability of ergodic averages. Transactions of the American Mathematical Society **362** 1 (2010) 261288

[2] Bienvenu, L., Day, A. Mezhirov, I., Shen, A. Ergodic-type characterizations of Martin-Löf randomness. In 6th Conference on Computability in Europe (CiE 2010), volume 6158 of Lecture Notes in Comput. Sci., pages 49–58. Springer, Berlin, 2010

[3] Billingsley, P,: Ergodic Theory and Information. Wiley, 1956.

[4] Bishop, E. Foundation of Constructive Analysis. New York: McGraw-Hill, 1967

[5] Franklin, J.N.Y., Towsner, H.: Randomness and non-ergodic systems. Mosc. Math. J., 2014, **14** 4, 711–744

[6] Gacs, P., Hoyrup, M., Rojas, C.: Randomness on Computable Probability Spaces-A Dynamical Point of View. Theory of Computing Systems **48** 3 (2011) 465–485

[7] Hoeffding, W.: Probability inequalities for sums of bounded random variables. Journal of the American Statistical Association **58** 301 (1963) 13-30

[8] Hoyrup, M., Rojas, C.: Computability of probability measures and Martin-Löf randomness over metric spaces. Information and Computation **207** 7 (2009) 830-847

[9] Galatolo, S., Hoyrup, M., Rojas, C.: Computing the speed of convergence of ergodic averages and pseudorandom points in computable dynamical systems Computability and Complexity in Analysis (CCA 2010) EPTCS **24** (2010) 7-18 doi:10.4204/EPTCS.24.6

[10] Kolmogorov, A.N. Three approaches to the quantitative definition of information. *Problems Inform. Transmission*, **1**(1), 1965, 1–7.

[11] Kolmogorov. A.N. On the logical foundations of information theory and probability theory. *Problems Inform. Transmission*, **5**(3), 1969, 1–4.

[12] Kolmogorov, A.N. Combinatorial foundations of information theory and the calculus of probabilities. *Russian Math. Surveys*, **38**(4), 1983, 29–40.

[13] Kolmogorov, A.N. On logical foundations of probability. *Lecture Notes in Mathematics*, **1021**, 1983, 1–5.

[14] Krengel, U. Ergodic Theorems, Berlin, New York: de Cruyter 1984.

[15] Kucera A. Measure, Π^0_1 classes, and complete extensions of PA. Lecture Notes in Mathematics. **1141** 1985, 245–259.

[16] Li, M., Vitányi, P. An Introduction to Kolmogorov Complexity and Its Applications, Springer-Verlag. New York 1997.

[17] Rogers, H. Theory of Recursive Functions and Effective Computability, New York: McGraw-Hill 1967.

[18] Shiryaev, A.N. Probability, Berlin: Springer 1980.

[19] V'yugin, V.V.: Effective Convergence in Probability and an Ergodic Theorem for Individual Random Sequences. Theory Probab. Appl. **42** (1) 1998, 39-50.

[20] V'yugin, V.V.: Ergodic theorems for individual random sequences. Theoretical Computer Science **207** (4) 1998, 343–361.