On Instability of the Ergodic Limit Theorems with Respect to Small Violations of Algorithmic Randomness

Vladimir V'yugin

Institute for Information Transmission Problems

Bol’shoy Karetyni per. 19, Moscow GSP-4, 127994, Russia, Email:vyugin@iitp.ru

Abstract—An instability property of the Birkhoff’s ergodic theorem and related asymptotic laws with respect to small violations of algorithmic randomness is studied. The Shannon–McMillan–Breiman theorem and all universal compression schemes are also among them.

I. INTRODUCTION

A lack of effective convergence bounds is typical for asymptotic results of ergodic theory like Birkhoff’s ergodic theorem – Krengel [2], Shannon–McMillan–Breiman theorem and universal compressing schemes – Ryabko [3].

In this paper we present new impossibility results based on analysis of ergodic theory in algorithmic randomness framework – instability under small violations of algorithmic randomness.

It is well known that main laws of probability theory are valid not only almost surely but for any individual Martin-Löf random sequence. An infinite binary sequence \( \omega_1\omega_2\ldots \) is Martin-Löf random with respect to uniform or 1/2-Bernoulli measure if and only if \( K(\omega^n) \geq n - O(1) \) for all \( n \), where \( K(\omega^n) \) is a monotone Kolmogorov complexity of a binary string \( \omega^n = \omega_1\ldots\omega_n \).

A stability property under small violations of algorithmic randomness of the main limit probability laws was discovered by Schnorr [4] and Vovk [7]. They shown that the law of large numbers for uniform Bernoulli measure holds for a binary sequence \( \omega_1\omega_2\ldots \) if \( K(\omega^n) \geq n - \sigma(n) \), where \( \sigma(n) \) is an arbitrary computable function such that \( \sigma(n) = o(n) \) as \( n \to \infty \), and the law of iterated logarithm holds if \( K(\omega^n) \geq n - \sigma(n) \), where \( \sigma(n) \) is an arbitrary computable function such that \( \sigma(n) = o(\log \log n) \) as \( n \to \infty \).

In that follows \( l(\alpha) \) denotes the length of a finite sequence \( \alpha \), \( \Omega = \{0,1\}^\infty \). A probability measure \( P \) on \( \Omega \) is defined as follows: define \( P(\Gamma_\alpha) \) for each finite sequence \( \alpha \), where \( \Gamma_\alpha = \{ \omega \in \Omega : \alpha \subset \omega \} \), and extend it for all Borel subsets of \( \Omega \).

Some notions from ergodic theory are given in Section V-B. Let \( T \) be a measure preserving transformation, and \( f \) be a computable function. V’yugin [8] proved an algorithmic version of the ergodic theorem: for any infinite binary sequence \( \omega \) the following implication is valid:

\[
K(\omega^n) \geq n - O(1) \implies \lim_{n \to \infty} \frac{1}{n-1} \sum_{i=0}^{n-1} f(T^i \omega) = \hat{f}(\omega) \quad (1)
\]

for some \( \hat{f}(\omega) = E(f) \) for ergodic \( T \).

We show in this paper that a constant \( O(1) \) in (1) cannot be replaced on an arbitrary slow increasing variable \( \sigma(n) \).

II. MAIN RESULTS

Let \( d_P(\omega^n) = -\log P(\omega^n) - K(\omega^n) \) be a universal test of randomness with respect to a measure \( P \). For uniform measure \( d(\omega^n) = n - K(\omega^n) \).

In the following proposition some sufficient condition of stability of a probability law is given.

**Proposition 1:** Assume there exists a computable double sequence \( \{U_{k,n}\} \) of sets of finite binary sequences such that \( P(\cup_{i \in \mathbb{N}} U_{i,n}) \leq \rho_k(n) c_{k,n} \), where \( c_{k,n} \) is a computable sequence of positive numbers and \( \rho_k(n) \) is a sequence of positive functions (not necessary computable) such that \( \sum_{n=1}^{\infty} c_{k,n} < \infty \) and \( \rho_k(n) \to 0 \) as \( n \to \infty \) for each \( k \).

Then for any \( \alpha, d_P(\omega^n) = o(\log(1/\rho_k(n))) \) as \( n \to \infty \) implies \( \omega^n \notin U_{n,k} \) for almost all \( n \).

**Proof.** Since \( \sum_{n=1}^{\infty} c_{k,n} < \infty \), we can define a corresponding prefix-free code, and so, \( K(\omega^n) \leq -\log c_{k,n} + O(1) \) (all logarithms are on the base 2). Assume \( \omega^n \in U_{k,n} \) for infinitely many \( n \). Then \( P(\cup_{i \in \mathbb{N}} U_{i,n}) \geq P(\omega^n) \) for these \( n \). Then we have for any \( k, o(\log(1/\rho_k(n))) = d_P(\omega^n) = -\log P(\omega^n) - K(\omega^n) \geq -\log \rho_k(n) - \log c_{k,n} + O(1) = -\log \rho_k(n) + O(1) \) for infinitely many \( n \). This contradiction proves the proposition. \( \triangle \)

Using Proposition 1 we can prove stability property for main probability laws. Let \( \lambda \) be the uniform measure on \( \Omega \), where \( \lambda(\Gamma) = 2^{-l(\alpha)} \). By Chernoff inequality \( \lambda(U_{k,n}) \leq 2e^{-2\varepsilon_k^2 n} \), where \( \varepsilon_k \) is such that \( \varepsilon_k = o(k) \) as \( k \to \infty \) and \( U_{k,n} = \{ x : l(x) = n, \frac{1}{n} \sum_{i=1}^{n} x_i - \frac{1}{2} \geq \varepsilon_k \} \).

Suppose \( d(\omega^n) \leq a(n) \) for all \( n \), where \( a(n) \) is a computable function such that \( a(n) = o(n) \) as \( n \to \infty \).

We can write \( \lambda(U_{k,n}) \leq 2^{-a(n) c_{k,n}} \), where \( c_{k,n} = 2e^{-2\varepsilon_k^2 + a(n)} \). For any \( k \), the series \( \sum_{n=1}^{\infty} c_{k,n} \) convergent since
\( \alpha(n) = o(n) \). By Proposition \( \text{I} \) for each \( k, \omega^n \notin U_{n,k} \) for almost all \( n \), i.e., the law of large numbers holds.

An analogous construction can be developed for the law of iterated logarithm for uniform Bernoulli measure and for law of large numbers for stationary ergodic Markov chains.

We present “a measure free” (comparing with \( \text{[9]} \)) instability property of the ergodic theorem. In Theorems \( \text{I} \) and \( \text{II} \) the uniform measure on \( \Omega \) is considered.

**Theorem 1:** Let \( \sigma(n) \) be a nondecreasing unbounded computable function. Then there exist a computable ergodic measure preserving transformation \( T \) and a sequence \( \omega \in \Omega \) such that the inequality \( Km(\omega^n) \geq n - \sigma(n) \) holds for all \( n \) and the limit \( \text{II} \) does not exist for some computable indicator function \( f \) (with the range \( \{0, 1\} \)).

The construction of the transformation \( T \) is given in Section \( \text{III} \).

**Theorem 2:** A measure preserving transformation \( T \) can be constructed such that for any nondecreasing unbounded computable function \( \sigma(n) \) a sequence \( \omega \in \Omega \) exists such that \( Km(\omega^n) \geq n - \sigma(n) \) holds for all \( n \) and the limit \( \text{I} \) does not exist for some computable indicator function \( f \).

Does an ergodic transformation \( T \) exists satisfying Theorem \( \text{II} \)? is an open question.

Recall some notions of symbolic dynamics. A partition is a sequence pairwise of disjoint subsets \( \pi = (\pi_1, \ldots, \pi_k) \) of the interval \([0, 1] \) whose union is equal to this interval. A transformation \( T \) defines a measure on the set of all finite and infinite words in the alphabet \( A = \{0, 1, \ldots, k - 1\} \) as follows

\[
P(a_1a_2 \ldots a_n) = \lambda\{\omega : T^i(\omega) \in \pi_{a_i}, i = 1, 2, \ldots, n\}, \quad (2)
\]

where \( a_1a_2 \ldots a_n \) is a sequence of letters from \( A \). The measure \( P \) can be extended on all Borel subsets of \( A^\infty \) by the Kolmogorov’s extension theorem. The measure \( P \) defined by \( \text{II} \) is stationary and ergodic with respect to the left shift if and only if the measure preserving transformation \( T \) is ergodic with respect to \( \lambda \).

In the construction given in Section \( \text{III} \) \( k = 2 \), and we define a specific partition \( (\pi_0, \pi_1) \) of \([0, 1] \).

We will use a natural correspondence between points of \([0, 1] \) and infinite binary sequences representing dyadic representations of these points (see Section \( \text{II} \)). This correspondence is one-to-one besides a measure zero.

Let \( T \) be the left shift on \( \Omega \), and \( P \) be the computable stationary measure (i.e. the left shift \( T \) is invariant w.r.t \( P \)) defined by \( \text{II} \) using the partition \( (\pi_0, \pi_1) \).

Recent result of Hochman \( \text{I} \) implies an algorithmic version of the Shannon–McMillan–Breiman theorem for any computable stationary ergodic measure \( P \) with entropy \( H \),

\[
Km(\omega^n) \geq - \log P(\omega^n) - O(1)
\]

implies

\[
\lim_{n \to \infty} \frac{Km(\omega^n)}{n} = \lim_{n \to \infty} \frac{- \log P(\omega^n)}{n} = H \quad (3)
\]

The construction given in Section \( \text{III} \) shows also an instability property of the relation \( \text{III} \) (this was first shown in \( \text{[9]} \)).

**Theorem 3:** Let \( \sigma(n) \) as in Theorem \( \text{I} \) \( \epsilon \) be a sufficiently small positive real number. A computable stationary ergodic measure \( P \) (defined by \( \text{II} \) and by the partition \( (\pi_0, \pi_1) \)) with entropy \( 0 < H \leq \epsilon \) and an infinite binary sequence \( \omega \) exist such that for all \( n \)

\[
Km(\omega^n) \geq - \log P(\omega^n) - \sigma(n),
\]

\[
\limsup_{n \to \infty} \frac{Km(\omega^n)}{n} \geq \frac{1}{4}, \quad (4)
\]

\[
\liminf_{n \to \infty} \frac{Km(\omega^n)}{n} \leq \epsilon. \quad (5)
\]

A code \( \{\phi_n\} \) is called universal with respect to a class of stationary ergodic sources if for any computable stationary ergodic measure \( P \) (with entropy \( H \))

\[
\lim_{n \to \infty} \frac{I(\phi_n(\omega^n))}{n} = H \text{ a.s.}, \quad (6)
\]

where \( I(x) \) is length of a word \( x \).

Lempel–Ziv coding scheme is an example of such universal coding scheme.

Since universal coding schemes are asymptotically optimal, i.e., for almost all sequences, they have compressing ratio \( \text{IV} \) the same as Kolmogorov complexity compressing ratio \( \text{III} \), we have an instability property for any universal coding schemes.

**Theorem 4:** Let \( \sigma(n) \) and \( \epsilon \) as in Theorem \( \text{I} \) \( \text{A} \) computable stationary ergodic measure \( P \) with entropy \( 0 < H \leq \epsilon \) exists such that for each universal code \( \{\phi_n\} \) an infinite binary sequence \( \omega \) exists such that for all \( n \)

\[
Km(\omega^n) \geq - \log P(\omega^n) - \sigma(n),
\]

\[
\limsup_{n \to \infty} \frac{I(\phi_n(\omega^n))}{n} \geq \frac{1}{4},
\]

\[
\liminf_{n \to \infty} \frac{I(\phi_n(\omega^n))}{n} \leq \epsilon. \quad (6)
\]

**III. PROOF**

We will define an ergodic transformation \( T \) and prove that an \( \omega \in \Omega \) exists such that two conditions hold

\[
\limsup_{n \to \infty} \frac{1}{n} \sum_{i=0}^{n-1} \chi(T^i\omega) \geq 1/16, \quad (7)
\]

\[
\liminf_{n \to \infty} \frac{1}{n} \sum_{i=0}^{n-1} \chi(T^i\omega) \leq 2r, \quad (8)
\]

where \( r \) is an arbitrary small real number and an indicator function \( \chi \) is defined below.

We use cutting and stacking method to define an ergodic transformation \( T \) of the unit interval \([0, 1] \) (see Section \( \text{IV-B} \) for details of this method).

Let \( r > 0 \) be a sufficiently small rational number. Define a partition \( \pi_0 = [0, \frac{1}{2}] \cup \left[\frac{1}{2} + r, 1\right] \) and \( \pi_1 = [\frac{1}{2}, \frac{1}{2} + r] \). The ergodic transformation \( T \) will be defined by a sequence of gadgets \( \Delta_s, \Pi_s \), where \( s = 0, 1, \ldots \).
A computable with respect to $\sigma$ sequence of positive integer numbers exists such that $0 < h_{s-2} < h_{s-1} < h_0 < h_1 < \ldots$ and

$$\sigma(h_{i-1}) - \sigma(h_{i-2}) > i - \log r + 8$$

for all $i = 0, 1, \ldots$.

The gadgets will be defined by mathematical induction on steps. The gadget $\Delta_0$ is defined by cutting of the interval $[\frac{1}{2} - r, \frac{1}{2} + r]$ on $2h_0$ equal parts and by stacking them. Let $\Pi_0$ be a gadget defined by cutting of the intervals $[0, \frac{1}{2} - r]$ and $[\frac{1}{2} + r, 1]$ in $2h_0$ equal subintervals and stacking them. The purpose of this definition is to construct initial gadgets of height $2h_0$ with supports satisfying $\lambda(\Delta_0) = 2r$ and $\lambda(\Pi_0) = 1 - 2r$.

The sequence of gadgets $\{\Delta_s\}$, $s = 0, 1, \ldots$, will define an approximation of the uniform Bernoulli measure concentrated on the names of their trajectories. The sequence of gadgets $\{\Pi_s\}$, $s = 0, 1, \ldots$, will define a measure with sufficiently small entropy.

We use a natural correspondence between points of $[0, 1]$ and infinite binary sequences representing dyadic representations of these points. This correspondence is defined using correspondence between finite binary sequences and subintervals of $[0, 1]$ with dyadically rational endpoints. A finite sequence $\alpha$ of length $n$ corresponds to an interval $[a, a + 2^{-n}]$ with dyadically rational endpoints, where $a = \sum_{i=1}^{n} \alpha_i 2^{-i}$. We choose $r$ such that both elements $\pi_0$ and $\pi_1$ of the partition can be represented (up to a set of measure zero) as finite unions of pairwise disjoint intervals with dyadically rational endpoints.

The purpose of the construction is to suggest conditions under which an infinite sequence $\omega \in \Omega$ exists such that $Km(\omega^n) \geq n - \sigma(n)$ for all $n$ and the limit (1) does not exist for a name $\chi(\omega)\chi(T(\omega))\chi(T^2(\omega)) \ldots$ of its trajectory, where $\chi(\omega) = i$ if $\omega \in \pi_i$, $i = 0, 1, 2$. A construction. Let at step $s - 1$ ($s > 0$) gadgets $\Delta_{s-1}$ and $\Pi_{s-1}$ were defined. Cut of the gadget $\Delta_{s-1}$ into two copies $\Delta'$ and $\Delta''$ of equal width (i.e. we cut of each column into two subcolumns of equal width) and join $\Pi_{s-1} \cup \Delta''$ in one gadget. Find a number $R_s$ and do $R_s$-fold independent cutting and stacking of the gadget $\Pi_{s-1} \cup \Delta''$ and also of the gadget $\Delta'$ to obtain new gadgets $\Pi_s$ and $\Delta_s$ of height $\geq 2h_s$ such that the gadget $\Pi_{s-1} \cup \Delta''$ is $(1 - 1/s)$-well-distributed in the gadget $\Pi_s$. The needed number $R_s$ exists by Lemma 2 (Section (V-A)).

Properties of the construction. Define a transformation $T = T(\Pi_s)$. Since the sequence of the gadgets $\{\Pi_s\}$ is complete (i.e. $\lambda(\Pi_s) \to 1$ and $\omega(\Pi_s) \to 0$ as $s \to \infty$), $T$ is defined for $\omega$ almost all $\omega$. The transformation $T$ is ergodic by Lemma 3 (Section (V)), where $T_s = \Pi_s$, since the sequence of gadgets $\Pi_s$ is complete. Besides, the gadget $\Pi_{s-1} \cup \Delta''$, and the gadget $\Pi_{s-1}$ are $(1 - 1/s)$-well-distributed in $\Pi_s$ for any $s$. By construction $\lambda(\Delta_i) = 2^{-i+1}r$ and $\lambda(\Pi_i) = 1 - 2^{-i+1}r$ for all $i = 0, 1, \ldots$.

We will define by induction on steps $s$ a sequence $\omega$ satisfying the conclusion of Theorem 3 as the union of an increasing sequence of initial fragments

$$\omega(0) \subset \cdots \subset \omega(k) \subset \cdots$$

Using Proposition 2 (Section (V-A)), define $\omega(0)$ such that $d(\omega(0)^j) \leq 2$ for all $j \leq l(\omega(0))$. Define $s(-1) = s(0) = 0$.

Induction hypotheses. Suppose that $k > 0$ and a sequence $\omega(0) \subset \cdots \subset \omega(k-1)$ is already defined. Also, the interval with dyadically rational endpoints corresponding to $\omega(k-1)$ is a subset of an interval from the support of the gadget $\Pi_{s(k-1)}$.

We suppose that $l(\omega(k-1)) > h_{s(k-1)}$, and $d(\omega(k-1)) \leq \sigma(h_{s(k-2)}) - 3$ if $k$ is odd and $d(\omega(k-1)) \leq \sigma(h_{s(k-2)})$ if $k$ is even.

Consider an odd $k$. Denote $a = \omega(k-1)$.

Let us consider a sufficiently large $s$ and a set of all finite strings extending $a$ such that the corresponding intervals with dyadically rational endpoints are subsets of intervals of columns of the gadget $\Pi_{s-1}$ and generate $\Pi_{s-1}$-trajectories with frequency of visiting the element $\pi_0$ of the partition less than $2r$. By the ergodic theorem the total measure of these subintervals tends to $2^{-l(a)}$ as $s \to \infty$.

We consider a sufficiently fine dyadically rational discretization of unit interval such that for all sufficiently large $s$ the total measure of the set of all such intervals with dyadically rational endpoints locating in the lower half of the gadget $\Pi_s$ is $\geq (1/4)2^{-l(a)}$. Let $C_a$ be a set of strings corresponding to these subintervals.

Fix some such $s$ and define $s(k) = s$.

By Proposition 2 (Section (V-A)) an $b \in C_a$ exists such that $d(b^j) \leq d(a) + 3$ for each $l(a) \leq j \leq l(b)$. Define $\omega(k) = b$. By induction hypotheses inequalities $d(a) \leq \sigma(h_{s(k-2)}) - 3$ and $l(a) \geq h_{s(k-1)} > h_{s(k-2)}$ hold. Then $d(b^j) \leq \sigma(h_{s(k-2)}) \leq \sigma(l(a)) \leq \sigma(j)$ for all $l(a) \leq j \leq l(b)$.

Also, we can take $l(b) \geq h_{s(k)}$. Therefore, the induction hypotheses and condition (9) are valid.

Let $k$ be even. Put $b = \omega(k-1)$ and $s = s(k-1) + 1$. Define $s(k) = s$.

Let us consider an arbitrary column from the gadget $\Delta_{s-1}$. Divide all its intervals into two equal parts: upper half and lower half. Any interval of the lower part has a trajectory of the length $\geq L/2$, where $L \geq 2h_{s-1}$ is the height of the gadget $\Delta_{s-1}$. The uniform measure of all such intervals is equal to $1/2\lambda(\Delta_{s-1})$.

By Chernoff inequality the measure $\lambda$ of all points of support of the gadget $\Delta_{s-1}$ whose trajectories have length $\geq L/2$ and frequency of ones $\leq 1/4$ is less than $2^{-\frac{L}{2} - \frac{3}{8}\lambda(\Delta_{s-1})} \leq \frac{1}{4}\lambda(\Delta_{s-1})$ (let $L$ be sufficiently large). Then all such points from intervals of the lower part of the gadget $\Delta_{s-1}$ whose trajectories have length $\geq L/2$ and frequency of ones more than $1/4$ have total measure at least $\frac{1}{4}\lambda(\Delta_{s-1})$.

Consider a set of all binary strings correspondent to intervals from the lower part of the gadget generated by cutting and
stacking of $\Pi_{s-1}\cup \Delta''$ such that trajectories starting from these intervals pass through an upper subcolumn of the gadget $\Delta''$ and have frequencies of ones at least 1/4. For any such subinterval consider a set of all subintervals from a discretization set with dyadically rational endpoints containing in it (we consider a sufficiently fine dyadically rational discretization). Let $D_b$ be a set of all binary strings correspondent to these intervals. By definition the trajectory of any such interval has length at most $2L$ and some fragment of its name has at least $L/8$ ones.

Hence, frequency of ones in the name of any such trajectory is at least $\frac{1}{8}$. Since $\gamma \leq 1$, the total measure of all such intervals is at least $\frac{1}{2} \cdot 2^{-l(b)}$. By Proposition 2 an $c \in D_b$ exists such that $d(c^j) \leq d(b) + 1 - \log \frac{c}{b} \leq d(b) + (\sigma(h_{s-1}) - \sigma(h_{s-2}) - 9) + 6 \leq \sigma(h_{s-1}) - 3 = \sigma(h_{s(k-1)}) - 3$ for all $j$ such that $l(b) \leq j \leq l(c)$. Here we have $d(b) \leq \sigma(h_{s(k-2)}) \leq \sigma(h_{s-2})$ by induction hypothesis. We also have used the inequality (11).

Besides, by the induction hypothesis we have $l(b) \geq h_{s-1}$. Therefore, $d(c^j) \leq \sigma(h_{s-1}) - \sigma(l(b)) \leq \sigma(j)$ for all $j$ such that $l(b) \leq j \leq l(c)$. Define $\omega(k) = c$. It is easy to see that all induction hypotheses are true for $\omega(k)$.

An infinite sequence $\omega$ is defined by a sequence of its initial fragments (10). We have proved that $d(\omega^j) \leq \sigma(j)$ for all $j$.

By the construction there are infinitely many initial fragments of trajectory of the sequence $\omega$ with frequency of ones $\geq 1/16$ in their names. Hence, the condition (7) holds.

The proof of Theorem 3 is analogous: since the sequence $\omega$ has sufficiently long fragments of $\Delta_{s-1}$-names we have (4) for infinitely many $n$; we have also (5) for infinitely many $n$, since $\omega$ has infinitely many fragments of $\Pi_{s-1}$-trajectories with frequency of zeros less than $2r$.

IV. AUXILIARY NOTIONS AND ASSERTIONS

A. Bounded increase of deficiency of randomness

Let $\lambda$ be a uniform measure on $\Omega$, $\lambda(\Gamma_x) = 2^{-l(x)}$ and a set $A$ consists of words $y$ such that $x \subseteq y$. Define $\lambda(A) = \lambda(\cup_{y \in A} \Gamma_y)$. Recall that $d(x) = l(x) - Km(x)$.

Proposition 2: Assume a set $A$ consists of words $y$ such that $x \subseteq y$ and $0 < \mu < 1$. Then a subset $A' \subseteq A$ exists such that $\lambda(A') > \mu \lambda(A)$ and

$$d(y^n) \leq d(x) - \log(1 - \mu) - \log \lambda(A) + l(x)$$

for all $y \in A'$ and $l(x) \leq n \leq l(y)$.

Proof: We will use in the proof the notion of supermartingale.

A function $M$ is called $\lambda$-supermartingale if: (1) $M(A) \leq 1$, (2) $M(x) \leq \frac{1}{2}(M(x0) + M(x1))$ for all $x$. We will consider only nonnegative supermartingales.

Lemma 1: There exists a supermartingale $M$ such that $d(x) \geq \log M(x)$ for all $x$.

Proof: Let some optimal function $\psi$ defines the monotone complexity $Km(x)$. Define

$$Q(\alpha) = \lambda(\cup\{\Gamma_p : \alpha \subseteq \psi(p)\}).$$

(12)

Since $\psi(p)$ is monotone (with respect to relation $\subseteq$ on strings), $Q(\Lambda) \leq 1$ and $Q(\alpha) \geq Q(\alpha0) + Q(\alpha1)$ for all $\alpha$. Then the function $M(\alpha) = 2^h(\alpha)Q(\alpha)$ is a supermartingale.

Since for any $\alpha$ the shortest string $p$ exists such that $\alpha \subseteq \psi(p)$ and $Km(\alpha) = l(p)$ is an element of the set from (12), we have $Q(\alpha) \geq 2^{-Km(\alpha)}$. Hence, $\lambda(\alpha) \leq \log M(x)$.

Let $d(x) \leq \log M(x)$, where $M$ is a supermartingale from Lemma 1. Denote $c = 2^{-l(x)}/(1 - \mu)\lambda(A)$ and define

$$A_1 = \{y \in A : (\exists j \geq l(x))(M(y^j) > cM(x))\}.$$

A set of words $B$ is called prefix free if for any two distinct words $x, y \in B$ conditions $x \not\subseteq y$ and $y \not\subseteq x$ hold.

It is easy to verify that for any prefix free set $B$ such that $x \subseteq y$ for all $y \in B$ the inequality

$$2^{-l(x)}M(x) \geq \sum_{y \in B} M(y)2^{-l(y)}$$

(13)

holds. For any $y \in A_1$ let $y^p$ be the initial fragment of $y$ of maximal length such that $M(y^p) > c$. The set $\{y^p : y \in A_1\}$ is prefix free. Then by (13) we have

$$1 \geq \sum_{y \in A_1} M(y^p)2^{-l(y^p) + l(x)} > c2^{l(x)} \sum_{y \in A_1} 2^{-l(y^p)} = c2^{l(x)}\lambda(A_1).$$

From this the inequality $\lambda(A_1) < c^{-1}2^{-l(x)}(1 - \mu)\lambda(A)$ follows. Define

$$A' = A - \{y \in A : z \subseteq y \text{ for some } z \in A_1\}.$$

Then $\lambda(A') > \mu \lambda(A)$. For any $y \in A'$ we have $M(y^p) \leq cM(x)$ for all $j$ such that $l(x) \leq j \leq l(y)$. The proposition follows from the inequality $d(x) \leq \log M(x)$.

B. METHOD OF CUTTING AND STACKING

An arbitrary measurable mapping of the a probability space into itself is called a transformation or a process. A transformation $T$ preserves a measure $P$ if $P(T^{-1}(A)) = P(A)$ for all measurable subsets $A$ of the space. A subset $A$ is called invariant with respect to $T$ if $T^{-1}A = A$. A transformation $T$ is called ergodic if each invariant with respect to $T$ subset $A$ has measure 0 or 1.

Recall the main notions and properties of cutting and stacking method (see Shields 5, 6). A column is a sequence $E = (L_1, \ldots, L_h)$ of pairwise disjoint subintervals of $[0, 1]$ of equal width (with rational endpoints); $L_1$ is the base, $L_h$ is the top of the column, $\hat{E} = \cup_{i=1}^h L_i$ is the support of the column, $w(E) = \lambda(L_1)$ is the width of the column, $h$ is the height of the column, $\lambda(\hat{E}) = \lambda(\cup_{i=1}^h L_i)$ is the measure of the column.

Any column defines an algorithmically effective transformation $T$ which linearly transforms $L_j$ to $L_{j+1}$ for all $j = 1, \ldots, h - 1$. This transformation $T$ is not defined outside all intervals of the column and at all points of the top $L_h$ interval of this column. Denote $T^0w = w$, $T^{i+1}w = T(T^iw)$. For any $1 \leq j < h$ an arbitrary point $\omega \in L_j$ generates a finite trajectory $\omega, T\omega, \ldots, T^{h-j-1}\omega$.

Since all points of the interval $L_j$ generate the identical trajectories, we refer to this trajectory as to trajectory of generated by the interval $L_j$. 


A partition $\pi = (\pi_1, \ldots, \pi_k)$ is compatible with a column $E$ if for each $j$ there exists an $i$ such that $L_j \subseteq \pi_i$. This number $i$ is called the name of the interval $L_j$, and the corresponding sequence of names of all intervals of the column is called the name of the column $E$. For any point $\omega \in L_j$, where $1 \leq j < h$, by $E$-name of the trajectory $\omega, T\omega, \ldots, T^{h-j}\omega$ we mean a sequence of names of intervals $L_j, \ldots, L_h$ from the column $E$. The length of this sequence is $h - j + 1$.

A gadget is a finite collection of disjoint columns. The width of the gadget $w(\Upsilon)$ is the sum of the widths of its columns. A union of gadgets $\Upsilon_i$ with disjoint supports is the gadget $\Upsilon = \bigcup \Upsilon_i$ whose columns are the columns of all the $\Upsilon_i$. The support of the gadget $\Upsilon$ is the union $\Upsilon$ of the supports of all its columns. A transformation $T(\Upsilon)$ is associated with a gadget $\Upsilon$ if it is the union of transformations defined on all columns of $\Upsilon$. With any gadget $\Upsilon$ the corresponding set of finite trajectories generated by points of its columns is associated. By $\Upsilon$-name of a trajectory we mean its $E$-name, where $E$ is that column of $\Upsilon$ to which this trajectory corresponds. A gadget $\Upsilon$ extends a column $\Lambda$ if the support of $\Upsilon$ extends the support of $\Lambda$, the transformation $T(\Upsilon)$ extends the transformation $T(\Lambda)$ and the partition corresponding to $\Upsilon$ extends the partition corresponding to $\Lambda$.

The cutting and stacking operations that are common used will now be defined. The distribution of a gadget $\Upsilon$ with columns $E_1, \ldots, E_n$ is a vector of probabilities

$$
\left( \frac{w(E_1)}{w(\Upsilon)}, \ldots, \frac{w(E_n)}{w(\Upsilon)} \right). \quad (14)
$$

A gadget $\Upsilon$ is a copy of a gadget $\Lambda$ if they have the same distribution and the corresponding columns have the same partition names. A gadget $\Upsilon$ can be cut into $M$ copies of itself $\Upsilon_{m, m = 1, \ldots, M}$, according to a given probability vector $(\gamma_1, \ldots, \gamma_M)$ of type $[2]$ by cutting each column $E_i = (L_{i, j} : 1 \leq j \leq h(E_i))$ (and its intervals) into disjoint subcolumns $E_{i,m} = (L_{i,j,m} : 1 \leq j \leq h(E_i))$ such that $w(E_{i,m}) = w(L_{i,j,m}) = \gamma_m w(L_{i,j})$. The gadget $\Upsilon_m = \{ E_{i,m} : 1 \leq i \leq L \}$ is called the copy of the gadget $\Upsilon$ of width $\gamma_m$. The action of the gadget transformation $T$ is not affected by the copying operation.

Another operation is the stacking gadgets onto gadgets. At first we consider the stacking of columns onto columns and the stacking of gadgets onto columns.

Let $E_1 = (L_{1, j} : 1 \leq j \leq h(E_1))$ and $E_2 = (L_{2, j} : 1 \leq j \leq h(E_2))$ be two columns of equal width whose supports are disjoint. The new column $E_1 \ast E_2 = (L_j : 1 \leq j \leq h(E_1)+h(E_2))$ is defined as $L_j = L_{1, j}$ for all $1 \leq j \leq h(E_1)$ and $L_j = L_{2,j-h(E_1)+1}$ for all $h(E_1) \leq j \leq h(E_1)+h(E_2)$. Let a gadget $\Upsilon$ and a column $E$ have the same width, and their supports are disjoint. A new gadget $E \ast \Upsilon$ is defined as follows. Cut $E$ into subcolumns $E_i$ according to the distribution of the gadget $\Upsilon$ such that $w(E_i) = w(U_i)$, where $U_i$ is the $i$-th column of the gadget $\Upsilon$. Stack $U_i$ on the top of $E_i$ to get the new column $E_i \ast U_i$. A new gadget consists of the columns $(E_i \ast U_i)$.

Let $\Upsilon$ and $\Lambda$ be two gadgets of the same width and with disjoint supports. A gadget $\Upsilon \ast \Lambda$ is defined as follows. Let the columns of $\Upsilon$ are $(E_i)$. Cut $\Lambda$ into copies $\Lambda_i$ such that $w(\Lambda_i) = w(E_i)$ for all $i$. After that, for each $i$ stack the gadget $\Lambda_i$ onto column $E_i$, i.e. consider a gadget $E_i \ast \Lambda_i$. The new gadget is the union of gadgets $E_i \ast \Lambda_i$ for all $i$. The number of columns of the gadget $\Upsilon \ast \Lambda$ is the product of the number of columns of $\Upsilon$ on the number of columns of $\Lambda$.

The $M$-fold independent cutting and stacking of a single gadget $\Upsilon$ is defined by cutting $\Upsilon$ into $M$ copies $\Upsilon_i$, $i = 1, \ldots, M$, of equal width and successively independently cutting and stacking them to obtain $\Upsilon^{(M)} = \Upsilon_1 \ast * \Upsilon_M$.

A sequence of gadgets $\{ \Upsilon_m \}$ is complete if

- $\lim_{m \to \infty} w(\Upsilon_m) = 0$;
- $\lim_{m \to \infty} \lambda(\Upsilon_m) = 1$;
- $\Upsilon_{m+1}$ extends $\Upsilon_m$ for all $m$.

Any complete sequence of gadgets $\{ \Upsilon_n \}$ determines a transformation $T = T(\Upsilon_n)$ which is defined on interval $[0, 1)$ almost surely.

By definition $T$ preserves the measure $\lambda$. In $[5]$ the conditions sufficient a process $T$ to be ergodic were suggested. Let a gadget $\Upsilon$ is constructed by cutting and stacking from a gadget $\Lambda$. Let $E$ be a column from $\Upsilon$ and $D$ be a column from $\Lambda$. Then $E \cap D$ is defined as the union of subcolumns from $D$ of width $w(E)$ which were used for construction of $E$.

Let $0 < \epsilon < 1$. A gadget $\Lambda$ is $(1 - \epsilon)$-well-distributed in $\Upsilon$ if

$$
\sum_{D \in \Lambda} \sum_{E \in \Upsilon} |\lambda(\hat{E} \cap \hat{D}) - \lambda(\hat{E})\lambda(\hat{D})| < \epsilon. \quad (15)
$$

We will use the following two lemmas.

**Lemma 2:** ($[5]$, Corollary 1), ($[6]$, Theorem A.1). Let $\{ \Upsilon_n \}$ be a complete sequence of gadgets and for each $n$ the gadget $\{ \Upsilon_n \}$ is $(1 - \epsilon_n)$-well-distributed in $\{ \Upsilon_{n+1} \}$, where $\epsilon_n \to 0$. Then $\{ \Upsilon_n \}$ defines the ergodic process.

**Lemma 3:** ($[6]$, Lemma 2.2). For any $\epsilon > 0$ and any gadget $\Upsilon$ there is an $\hat{M}$ such that for each $m \geq \hat{M}$ the gadget $\Upsilon$ is $(1 - \epsilon)$-well-distributed in the gadget $\Upsilon^{(m)}$ constructed from $\Upsilon$ by $m$-fold independent cutting and stacking.

**ACKNOWLEDGEMENT**

This research was partially supported by Russian foundation for fundamental research: 09-07-00180-a.

**REFERENCES**

[1] M. Hochman, “Upcrossing inequalities for stationary sequences and applications to entropy and complexity,” Ann. Probab. vol. 37, no. 6, pp. 2135–2149, 2009.

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

[3] B. Ryabko, “Twice universal coding,” Probl. Inform. Transm., vol. 20, pp. 173–178, 1984.

[4] C.P. Schnorr, “A unified approach to the definition of random sequences,” Mathematical Systems Theory, vol. 5, pp. 246–258, 1971.

[5] P.C. Shields, “Cutting and stacking: a method for constructing stationary processes,” IEEE Trans. Inform. Theory, vol. 37, no. 6, pp. 1605–1617, 1991.

[6] P.C. Shields “Two divergence-rate counterexamples,” J. Theoret. Probability, vol. 6, pp. 521–545, 1993.
[7] V.G. Vovk, “The law of the iterated logarithm for random Kolmogorov, or chaotic sequences,” *SIAM Theory Probab. Applic.*, vol. 32, pp. 413–425, 1987.

[8] V.V. V'yugin, “Ergodic theorems for individual random sequences,” *Theoretical Computer Science*, vol. 207, no. 4, pp. 343–361, 1998.

[9] V.V. V'yugin, “Problems of robustness for universal coding schemes,” *Probl. Inform. Transm.*, vol. 39, no. 1, pp. 32–46, 2003.