Problems of robustness for universal coding schemes

V.V.V’yugin

Abstract

The Lempel–Ziv universal coding scheme is asymptotically optimal for the class of all stationary ergodic sources. A problem of robustness of this property under small violations of ergodicity is studied. A notion of deficiency of algorithmic randomness is used as a measure of disagreement between data sequence and probability measure. We prove that universal compressing schemes from a large class are non-robust in the following sense: if the randomness deficiency grows arbitrarily slowly on initial fragments of an infinite sequence then the property of asymptotic optimality of any universal compressing algorithm can be violated. Lempel–Ziv compressing algorithms are robust on infinite sequences generated by ergodic Markov chains when the randomness deficiency of its initial fragments of length \( n \) grows as \( o(n) \).

1 Introduction

Well known data compression schemes universal for classes of stationary ergodic sources, like Lempel–Ziv algorithms, are asymptotically optimal [1, 2]. In particular, for almost every infinite binary sequence \( \omega_1 \omega_2 \ldots \) generated by an ergodic source with unknown statistics the average length of codeword related to one bit of input sequence tends to entropy of the source when the block length tends to infinity. It looks significant a property of coding algorithms to be robust under small variations of its parameters. We consider in this paper a problem of robustness of the asymptotic optimality property under small violations of ergodicity of a source. A notion of deficiency of algorithmic randomness \( d_P(\omega_1 \ldots \omega_n) \) is used as a measure of
disagreement between data sequence $\omega_1\omega_2\ldots\omega_n\ldots$ and probability distribution $P$. This notion is considered in Kolmogorov theory of algorithmic complexity and randomness [3, 4, 5]. In the framework of this theory we can formulate laws of probability theory, i.e. statements which hold almost surely, in a “pointwise” form as statements which hold for individual objects. The set of Martin-Löf [6] random sequences is used at the present time as a standard set of such individual objects. The measure of this set is equal 1 and laws of probability theory, like the law of large numbers, the law of iterated logarithm and others, hold for each sequence from this set. A sequence $\omega_1\omega_2\ldots$ is algorithmic random with respect to a computable measure $P$ if and only if the randomness deficiency $d_P(\omega_1\ldots\omega_n)$ of its initial fragments of length $n$ is bounded then $n$ increases (exact definition of the randomness deficiency will be given in Section 2).

“Robustness” under small violations of algorithmic randomness of some probability laws was studied in [7, 8]. These statements hold not only for random sequences but they hold also for sequences from more broader sets: the law of large numbers for symmetric Bernoulli scheme holds for any sequence $\omega_1\omega_2\ldots$ such that $d_P(\omega_1\ldots\omega_n) = o(n)$; the law of iterated logarithm holds if $d_P(\omega_1\ldots\omega_n) = o(\log \log n)$. Small variations of these conditions imply violations of these laws. Robustness property can be failed for laws of more general type. It is proved in [9] that Birkhoff’s ergodic theorem is non-robust in this sense – any small growing of the deficiency of randomness on initial fragments of an infinite sequence $\omega_1\omega_2\ldots$ can imply the violation of the statement of this theorem.

We prove that for any unbounded, nonnegative, and nondecreasing function $\sigma(n)$ a stationary ergodic (and computable with respect to $\sigma$) measure $P$ exists such that for any universal code for some infinite binary sequence $\omega_1\omega_2\ldots$ inequality $d_P(\omega_1\ldots\omega_n) \leq \sigma(n)$ holds for all sufficiently large $n$ and the property of asymptotic optimality of this code is violated for this sequence.

2 Algorithmic complexity and randomness

Main notions and results on computability can be found in [10]. In this paper we consider algorithms working with constructive objects (that is integer and rational numbers, or words in finite alphabet). Let $B$ be some finite alphabet and $B^*$ be the set of all words (finite sequences of letters) in it. Empty word
Λ is also an element of $B^*$. Let $l(x)$ be the length (number of letters) of a word $x \in B^*$. We write $x \subseteq y$ if a word $x$ is a prefix of a word $y$. Two words $x$ and $x'$ are comparable if $x \subseteq x'$ or $x' \subseteq x$. Let $bx$ be a concatenation of $b$ and $x$ (i.e. all letters of $x$ follow after all letters of $b$ in $bx$).

Kolmogorov (algorithmic) complexity of a word $x \in B^*$ (with respect to a word $y \in B^*$) is equal to the length of the shortest binary codeword $p$ (i.e. $p \in \{0, 1\}^*$) by which given $y$ the word $x$ can be reconstructed

$$K_\psi(x|y) = \min \{ l(p) : \psi(p, y) = x \}.$$ 

By this definition the complexity depends on partial computable function $\psi$ – method of decoding. A.N.Kolmogorov proved that an optimal decoding algorithm $\psi$ exists such that for any positive constant $c$ (do not depending from $x$, $y$ and $\psi'$)

$$K_\psi(x|y) \leq K_{\psi'}(x|y) + 2K(\psi') + c$$

holds for any computable decoding function $\psi'$ and for all words $x$ and $y$. Here $K(\psi')$ is the length of the shortest program computing values of $\psi'$. We fix some optimal decoding function $\psi$. The value $K(x|y) = K_\psi(x|y)$ is called (conditional) Kolmogorov complexity of $x$ given $y$. Unconditional complexity of $x$ is defined $K(x) = K(x|\Lambda)$.

It follows from [11] that a corresponding to $\psi$ coding algorithm (in sense of Section 4) computing by $x$ a codeword $p$ of minimal length such that $\psi(p) = x$ does not exist.

We will use some properties of Kolmogorov complexity [5, 11]. Incompressibility property asserts that for any positive integer numbers $n$ and $m$ a portion of all sequences $x$ of length $n$ such that

$$K(x) < n - m,$$

is less than $2^{-m}$. Indeed, the number of all $x$ satisfying this inequality does not exceed the number of all binary programs generating them. Since the length of any such program is less than $n - m$ the number of these programs is less than $2^{n-m}$.

\footnote{We suppose that $\min \emptyset = +\infty$.}
\footnote{We suppose that some universal programming language is fixed, and all decoding programs are written in this language (the constant $c$ depends on this language).}
Let \( x \) and \( b \) be finite words. It is easy to construct a function which given any program computing \( bx \) and the length of \( b \) computes the word \( x \). Therefore,

\[
K(x) \leq K(bx) + 2 \log l(b) + c
\]

for any \( x \), where \( c \) is a positive constant not depending from \( b \) and \( x \).

We consider a probability space \((\Omega, F, P)\), where \( \Omega = \{0,1\}^\infty \), Borel field \( F \) is generated by balls \( \Gamma_x = \{\omega \in \Omega : x \subseteq \omega\} \), where \( x \in \{0,1\}^* \). To define a probability measure \( P \) on the space \( \Omega \) it is sufficient to define the concordant values \( P(\Gamma_x) = P(x) \) such that \( P(\Lambda) = 1 \) and \( P(x) = P(x0) + P(x1) \) for all \( x \), where \( x\nu \) denotes a word obtained from \( x \) by adding \( \nu \) on right. After that, the function \( P \) can be extended by Kolmogorov extension theorem \cite{12}. A uniform Bernoulli probability distribution on binary sequences is defined \( B_{1/2}(x) = 2^{-l(x)} \). A measure \( P \) is called computable if there exists an algorithm which given a finite sequence \( x \) and a degree of accuracy, a rational \( \epsilon > 0 \), outputs a rational approximation to \( P(x) \) with the accuracy \( \epsilon \).

A notion of algorithmic random sequence is defined using an algorithmic analogue of a set of measure 0. Let \( P \) be a computable probability measure on a set of all infinite binary sequences \( \Omega \).

A set \( M \subseteq \Omega \) has \( P \)-measure 0 if for each rational \( \epsilon > 0 \) there is a sequence \( x(1), x(2), \ldots \) of elements of \( \Xi \) such that the set \( U_\epsilon = \cup_i \Gamma_{x(i)} \) satisfies \( M \subseteq U_\epsilon \) and \( P(U_\epsilon) < \epsilon \). A \( P \)-null set is called effectively \( P \)-null if there exists a computable function \( x(\epsilon, i) \) such that \( M \subseteq U_\epsilon = \cup_i \Gamma_{x(\epsilon, i)} \) and \( P(U_\epsilon) < \epsilon \) for each rational \( \epsilon > 0 \). It can be proved that for any computable measure \( P \) there exists the largest with respect to the measure-theoretic inclusion effectively \( P \)-null set \cite{4, 5, 6}. The complement of this largest effectively \( P \)-null set is called the constructive support of the measure \( P \). An infinite sequence \( \omega \in \Omega \) is called algorithmic random with respect to the measure \( P \) (random in the sense of Martin-Löf) if it belongs to the constructive support of the measure \( P \).

Using some modification of decoding algorithms we can define a notion of algorithmic random sequence in terms of complexity \cite{4, 5, 13}. Let us consider monotonic computable transformations of sequences. Let \( A \) and \( B \) be finite alphabets, and let a set \( \hat{\psi} \subseteq A^* \times B^* \) is (recursively) enumerable (by means of some algorithm) and such that for any \( (x, y), (x', y') \in \hat{\psi} \) if \( x \) and \( x' \) are comparable then \( y \) and \( y' \) are also comparable. Let also \( A = \{0,1\} \).

\footnote{We will consider in the following logarithms on the base 2.}
The set \( \hat{\psi} \) defines some monotonic with respect to \( \subseteq \) decoding function

\[
\psi(p) = \sup\{x : (p, x) \in \hat{\psi}\}.
\] (4)

The class of such monotonic functions \( \psi \) determines the corresponding algorithmic complexity

\[
Km_\psi(x) = \min\{l(p) : x \subseteq \psi(p)\}.
\]

The corresponding optimal complexity \( Km(x) \) is differ from complexity \( K(x) \) by a term of order of logarithm from \( l(x) \). We have

\[
K(x) - 2\log l(x) - c \leq Km(x) \leq K(x) + 2\log K(x) + c \quad (5)
\]

for all \( x \), where \( c \) is a positive constant [4, 5].

For any sequence \( \omega \) denote by \( \omega^n = \omega_1 \ldots \omega_n \) its initial fragment of length \( n \). The following fundamental assertion (which at first was proved in [13]) holds.

**Proposition 1** Let \( P \) be some computable measure. Then

1) for any infinite sequence \( \omega \) a constant \( c \) exists such that for all \( n \) inequality \( Km(\omega^n) \leq -\log P(\omega^n) + c \) holds, besides, for any \( m \)

\[
P(\bigcup \{\Gamma_x : -\log P(x) - Km(x) \geq m\}) \leq 2^{-m};
\]

2) a sequence \( \omega \) is random with respect to a measure \( P \) in sense of Martin-Löf if and only if for some constant \( c \) it holds \( Km(\omega^n) \geq -\log P(\omega^n) - c \) for all \( n \).

These proposition shows that asymptotic behaviour of the function

\[
d_P(\omega^n) = -\log P(\omega^n) - Km(\omega^n)
\]

can be used as a quantitative measure of nonrandomness of the sequence \( \omega \). By Proposition 1 a sequence \( \omega \) is algorithmic random with respect to a computable measure \( P \) if and only if \( \sup d_P(\omega^n) < \infty \). The value \( d_P(\omega^n) \) is called the deficiency of algorithmic randomness of a word (finite sequence) \( \omega^n \) with respect to a measure \( P \) [4, 5, 14].

Basic notions of ergodic theory can be found in [15] (see also Appendix 2 to this paper). A property of “asymptotic optimality of compression” by means of the shortest codeword defining the Kolmogorov complexity holds.

\textsuperscript{4} Here the by supremum we mean an union of all comparable \( x \) in one sequence.
Corollary 1 Let $P$ be an arbitrary computable stationary ergodic measure, and let $H$ be its entropy. Then for $P$-almost all infinite sequences $\omega$ the following limits exist and the corresponding equalities hold
\[
\lim_{n \to \infty} \frac{K(\omega^n)}{n} = \lim_{n \to \infty} \frac{Km(\omega^n)}{n} = \lim_{n \to \infty} -\frac{\log P(\omega^n)}{n} = H.
\]

This corollary follows from Proposition 1, relation (5) and Shannon–McMillan–Breiman theorem [15]. At first this corollary was proved for $K(x)$ in [11]. In [16] a variant of (6) for algorithmic random sequence was obtained: for any infinite sequence $\omega$ random with respect to a computable ergodic measure $P$ with entropy $H$ relations (6) hold where the limit is replaced on upper limit.

3 Non-robustness property of the universal data compression scheme

It looks important a property of compressing algorithms to be robust under small variations of its parameters. The following Theorem 1 can be interpreted as an assertion of that “optimal compression scheme” corresponding to Kolmogorov complexity is non-robust in the class of all stationary ergodic sources. As consequences of this theorem we obtain in Section 4 results on non-robustness of computable universal coding schemes (see Propositions 2 and 3).

Theorem 1 For any nonnegative, nondecreasing, and unbounded function $\sigma(n)$ and for any real number $0 < \varepsilon < 1/4$ a computable with respect to $\sigma$ stationary ergodic measure $P$ with entropy $0 < H \leq \varepsilon$ and an infinite binary sequence $\alpha$ exist such that
\[
d_P(\alpha^n) \leq \sigma(n)
\]
for almost all $n$. It holds also
\[
\limsup_{n \to \infty} \frac{K(\alpha^n)}{n} \geq \frac{1}{4},
\]
\[
\liminf_{n \to \infty} \frac{K(\alpha^n)}{n} \leq \varepsilon.
\]
Proof. Let \( r > 0 \) be a sufficiently small rational number. Let us consider a partition

\[
\pi_0 = \left[ 0, \frac{1}{2} \right) \cup \left( \frac{1}{2} + r, 1 \right], \pi_1 = \left[ \frac{1}{2}, \frac{1}{2} + r \right]
\]

of semiopen interval \([0, 1)\) (the number \( r \) will be specified later). Using cutting and stacking method (basic definitions for this method will be given in Appendix 2) we will define an ergodic transformation \( T \) of interval \([0, 1)\) which will generate a stationary ergodic measure \( P \) on the set \( \Omega \). To define the measure \( P \) consider

\[
P(a_1a_2 \ldots a_n) = \lambda\{\omega : \omega \in [0, 1), T^i(\omega) \in \pi_{a_i}, i = 1, 2, \ldots, n\},
\]

where \( a_1a_2 \ldots a_n \) is an arbitrary binary sequence, \( \lambda \) is the uniform measure on the interval \([0, 1)\). The measure \( P \) is extended on arbitrary Borel subsets of \( \Omega \) by a natural fashion \([12]\).

The ergodic transformation \( T \) will be defined by a sequence of gadgets \( \Delta_s, \Pi_s \), where \( s = 0, 1, \ldots \). Let a gadget \( \Phi_s \) be the union of these two gadgets. We define at step \( s \) an approximation \( T_s = T(\Phi_s) \) of the transformation \( T \) and corresponding approximation \( P^s \) of the measure \( P \) analogously to (10). The transformation \( T_s \) determines finite trajectories starting in the points of internal intervals of these gadgets and finishing in the top intervals. Any such trajectory has a name which is a word in the alphabet \( \{0, 1\} \). By definition for any word \( a \) (for any set of words \( D \)) the number \( P^s(a) \) (\( P^s(D) \) accordingly) is equal to the sum of lengths of all intervals of the gadget \( \Pi_s \) from which trajectories with names extending \( a \) (extending words from \( D \)) start.

Since the function \( \sigma \) is nondecreasing and unbounded a computable with respect to it sequence of positive integer numbers exists such that \( 0 < h_{-2} < h_{-1} < h_0 < h_1 < \ldots \) and

\[
\sigma(h_{i-1}) - \sigma(h_{i-2}) > -\log r + i + 13
\]

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 \( \left[ \frac{1}{2} - r, \frac{1}{2} + r \right) \) on \( 2h_0 \) equal parts and by stacking them. Let \( \Pi_0 \) be a gadget defined by cutting of intervals \( [0, \frac{1}{2} - r) \) and \( (\frac{1}{2} + r, 1] \) in \( 2h_0 \) equal parts 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. The gadget \( \Pi_{s-1} \) will be extended at each step of the construction by a half part of the gadget \( \Delta_{s-1} \). After that, the independent cutting and stacking process will be applied to this extended gadget. This process eventually defines infinite trajectories of points from interval \([0, 1)\). The sequence of gadgets \( \{\Pi_s\} \), \( s = 0, 1, \ldots \), will be complete and will define the needed measure \( \mathcal{P} \). Lemmas 2 and 3 will ensure the transformation \( T \) and measure \( \mathcal{P} \) to be ergodic.

The purpose of the construction is to suggest conditions under which there exists a point in interval \([0, 1)\) having an infinite trajectory with a name \( \alpha \) satisfying (7), (8) and (9). To implement (8) we periodically extend initial fragments of \( \alpha \) by names of trajectories of gadgets \( \Delta_{s-1} \) (for suitable \( s \)) which have the maximal complexity. To bound the deficiency of randomness of initial fragment of length \( n \) by the value \( \sigma(n) \) we suggest with the help of condition (11) some relation between the height of the gadget \( \Delta_s \) and the measure of the support of this gadget. We will use Proposition 5 to define an extension with sufficiently small deficiency of randomness. To implement condition (9) it is sufficient to extend names in long runs of the construction only in account of trajectories of gadgets \( \{\Pi_s\} \), \( s = 0, 1, \ldots \). For any \( s \) only a portion \( \leq r \) of the support of such gadget belongs to element \( \pi_1 \) of the partition. Then by ergodic theorem the most part of (sufficiently long) trajectories of this gadget will visit \( \pi_1 \) according to this frequency, and the names of these trajectories will have the frequency of ones bounded by a small number \( 2r \), that ensures the bound (9).

**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 \( \Delta' \Delta'' \) 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 \( 2h_s \) 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 3 (Appendix 2).

**Properties of the construction.** Define \( T = T\{\Pi_s\} \). Since the sequence of the gadgets \( \{\Pi_s\} \) is complete (i.e. \( \lambda(\Pi_s) \rightarrow 1 \) and \( w(\Pi_s) \rightarrow 0 \) as \( s \rightarrow \infty \)) the transformation \( T \) is defined for \( \lambda \)-almost all \( \omega \). The measure \( \mathcal{P} \) is defined by (10). The measure \( \mathcal{P} \) is stationary, since the transformation \( T \) preserves the
uniform measure $\lambda$. Measure $P$ is ergodic by Lemma 2 (Appendix 2), where $\Upsilon_s = \Pi_s$, 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

$$\lambda(\hat{\Delta}_i) = 2^{-i+1}r \quad \text{and} \quad \lambda(\hat{\Pi}_i) = 1 - 2^{-i+1}r$$  \quad (12)$$

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

This construction is algorithmic effective, so the measure $P$ is computable with respect to $\sigma$.

Let us prove that entropy $H$ of the measure $P$ do not exceed $\epsilon$. Since $\lambda(\pi_1) = r$ and the transformation $T$ preserves the measure $\lambda$, by ergodic theorem in almost all points of interval $[0,1)$ a trajectory starts such that the limit of the frequency of visiting the element $\pi_1$ by this trajectory is equal $r$, when the length of initial fragment of such trajectory tends to infinity.\footnote{For any $\omega \in [0,1)$ the frequency of visiting of $\pi_1$ by trajectory starting in $\omega$ is equal to $(1/l) \sum_{i=1}^{l} \chi_1(T^i \omega)$, where $l$ is the length of this trajectory and $\chi_1(r) = 1$ if $r \in \pi_1$, and $\chi_1(r) = 0$, otherwise.}

Thus for any $\delta > 0$ for all sufficiently large $n$ the measure $P$ of all sequences $x$ of length $n$ with portion of ones $\leq 2r$ is $\geq 1 - \delta$. Let us consider any such sequence $x$ as an element a finite set consisting of all sequences of length $n$ and containing no more than $2rn \leq \frac{n}{2}$ ones. Then we obtain a standard upper bound

$$\frac{K(x)}{n} \leq \frac{1}{n} \log \left(2rn \left(\frac{n}{2rn}\right)\right) + \frac{2 \log n}{n} \leq -3r \log r$$  \quad (13)$$

for all sufficiently large $n$. By this inequality and by (13) we obtain upper bound $H \leq -3r \log r \leq \epsilon$ for entropy $H$ of the measure $P$, where $r$ is sufficiently small.

Let us prove that an infinite sequence $\alpha$ exists such that the conclusion of Theorem 1 holds. We will define $\alpha$ by induction on steps $s$ as the union of an increasing sequence of initial fragments

$$\alpha(0) \subset \ldots \subset \alpha(k) \subset \ldots$$  \quad (14)$$

For all sufficiently large $k$ the Kolmogorov complexity of initial fragment $\alpha(k)$ will be small if $k$ is odd, and complexity of $\alpha(k)$ will be large, otherwise.
Define \(\alpha(0)\) be equal to \(\Pi_0\)-name of some trajectory of length \(\geq h_0\) such that \(d_P(\alpha(0)) \leq 2\). This is possible to do by Proposition 5 (Appendix 1). Define \(s(-1) = s(0) = 0\).

**Induction hypotheses.** Suppose that \(k > 0\) and a sequence \(\alpha(0) \subset \ldots \subset \alpha(k-1)\) is already defined, and for some step \(s(k-1)\) of the construction the word \(\alpha(k-1)\) is \(\Pi_{s(k-1)}\)-name of a trajectory of some point from the support of the gadget \(\Pi_{s(k-1)}\). We suppose that \(l(\alpha(k-1)) > h_{s(k-1)}\), and if \(k\) is odd then \(d_P(\alpha(k-1)) \leq \sigma(h_{s(k-2)}) - 4\). If \(k\) is even then \(d_P(\alpha(k-1)) \leq \sigma(h_{s(k-2)})\) and \(P^{s(k-1)}(\alpha(k-1)) > (1/8)P(\alpha(k-1))\).

Let us consider any odd \(k\). Define \(a = \alpha(k-1)\).

Let us consider a set of all intervals (from columns) of the gadget \(\Pi_{s-1}\) with the following property: for any trajectory starting from this interval with \(\Pi_{s-1}\)-names extending \(a\), the frequency of visiting the element \(\pi_1\) of the partition is \(\leq 2r\). For the name \(\gamma\) of any such trajectory an inequality

\[
K(\gamma)/l(\gamma) \leq -3r \log r \leq \epsilon
\]

(15)

(analogous to (13)) holds, where \(r\) is sufficiently small. As in the proof of the inequality \(H \leq \epsilon\) we obtain by ergodic theorem that for all sufficiently large \(s\) total length of all interval from this set is \(\geq (1/2)P(a)\).

Let us consider an arbitrary column from the gadget \(\Pi_s\). Divide all its intervals on two equal parts: upper part and lower part. We will consider only intervals from the lower part. Any trajectory starting from a point of an interval from this part has length \(\geq h_s\). Fix some \(s\) as above and define \(s(k) = s\). Let \(U_s(a)\) be all intervals from the lower part of the gadget \(\Pi_s\) such that trajectories starting from them and having \(\Pi_s\)-names extending \(a\) satisfy the inequality (13). Let \(D_a\) be a set of all \(\Pi_s\)-names of all these trajectories. Inequality \(P^s(D_a) = P^a > (1/4)P(a)\) holds for the total length \(P^s(D_a)\) of all intervals from \(U_s(a)\).

Define \(\bar{D} = \cup_{x \in D} \Gamma_x\). It is easy to prove that a set \(C_a \subseteq D_a\) exists such that \(P(\bar{C}_a) > (1/8)P(\bar{D}_a)\) and \(P^s(b) > (1/8)P(b)\) for all \(b \in C_a\). By Proposition 5 (Appendix 1) an \(b \in C_a\) exists such that \(d_P(b') \leq d_P(a) + 4\) when \(l(a) \leq j \leq l(b)\). Define \(\alpha(k) = b\). By induction hypotheses inequalities \(d_P(a) \leq \sigma(h_{s(k-2)}) - 4\) and \(l(a) \geq h_{s(k-1)} > h_{s(k-2)}\) hold. Then \(d_P(b') \leq \sigma(h_{s(k-2)}) \leq \sigma(l(a)) \leq \sigma(j)\) for all \(l(a) \leq j \leq l(b)\).

Notice, that \(l(b) \geq h_{s(k)}\), since any trajectory defining \(b\) starts from an interval of the lower part of the gadget \(\Pi_s\), and the height of this gadget is \(\geq 2h_s\). The rest induction hypotheses are proved above.
The condition (9) is true, since condition (15) holds for infinite number of initial fragments $\alpha(k)$ of the sequence $\alpha$.

Let $k$ be even. Put $b = \alpha(k - 1)$. Let $s = s(k - 1) + 1$. 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 part and lower part. Any trajectory starting from an interval of the lower part have 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 $\frac{1}{2}\lambda(\hat{\Delta}_{s-1})$. Let us consider the names $x^{L/2}$ of initial fragments of length $L/2$ of all these trajectories. By incompressibility property of Kolmogorov complexity (2) and by choice of $L$ the uniform Bernoulli measure of all sequences of length $L/2$ satisfying

$$K(x^{L/2}) \leq 1 - \frac{2}{h_{s-2}},$$

is less than $2^{-L/h_{s-2}} \leq 1/4$. Names of initial fragments (of length $L/2$) of the rest part of trajectories starting from intervals of lower part of the gadget $\Delta_{s-1}$ satisfy

$$K(x^{L/2}) \geq 1 - \frac{2}{h_{s-2}}. \quad (16)$$

It is noted in Appendix 2 (Remark 1), for any step $s$ of the construction the equality $P_{s-1}(x) = 2^{-l(x)}\lambda(\Delta_{s-1})$ holds for the name $x$ of any trajectory of the gadget $\Delta_{s-1}$. We conclude from this equality that the uniform measure of all intervals from the lower part of the gadget $\Delta_{s-1}$, such that trajectories with names (more correctly, with initial fragments $x^{L/2}$ of such names) satisfying (16) start from these intervals, is at least $\frac{1}{4}\lambda(\hat{\Delta}_{s-1})$.

By (11) and (12)

$$\gamma = \frac{\lambda(\hat{\Delta}'')}{\lambda(\hat{\Pi}_{s-1})} = \frac{\lambda(\hat{\Delta}_{s-1})}{2\lambda(\hat{\Pi}_{s-1})} = \frac{2^{-s+1}r}{1 - 2^{-s+2}} > 2^{-(s+2)} + 1/2 \geq 2^{-(\sigma(h_{s-1}) - \sigma(h_{s-2})+12} \quad (17)$$

Let us consider $R_s$-fold independent cutting and stacking of the gadget $\Pi_{s-1} \cup \Delta''$ in more details. At first, we cut out this gadget on $R_s$ copies. When we stack the next copy on already defined part of the gadget the portion of all trajectories of any column from the previously constructed part, which go to a subcolumn from the gadget $\Delta''$, is equal to

$$\frac{\lambda(\hat{\Delta}'')}{\lambda(\hat{\Pi}_{s-1}) + \lambda(\hat{\Delta}'')} = \frac{\gamma}{1 + \gamma}. \quad (18)$$
This is true, since by definition any column is covered by a set of subcolumns with the same distribution as the gadget $\Pi_{s-1} \cup \Delta''$ has. Total length of all intervals of the gadget $\Pi_{s-1}$ such that trajectories with names extending $b$ start from these intervals is equal to $P^{s-1}(b)$.

Consider the lower half of all subintervals generated by cutting and stacking of the gadget $\Pi_{s-1}$ in which trajectories with $\Pi_{s-1}$-names extending $b$ start. The length of any such trajectory (in $\Pi_s$) is at least $h_s$. By this reason some inductive hypothesis will be true. The measure of all remaining subintervals decreases twice. After that, we consider a subset of these subintervals, such that trajectories starting from subintervals of this subset go into subcolumns of the gadget $\Delta''$. The measure of remaining subintervals is multiplied by a factor $\gamma/(1 + \gamma)$. Further, consider subintervals from the remaining part generating trajectories whose names have in $\Delta''$ fragments satisfying (16). The measure of the remaining part can be at least $1/4$ from the previously considered part. We obtain this bound from previous estimate of the portion of subintervals generating trajectories in the gadget $\Delta''$ of length $\geq L/2$ satisfying (16). Let $D_b$ be a set of all $\Pi_{s-1}$-names of all trajectories starting from subintervals remaining after these selection operations. Then

$$P^s(D_b) \geq \frac{\gamma}{8(1 + \gamma)} P^{s-1}(b).$$

(19)

The name of any such trajectory has initial fragment of type $bx'x^{L/2}$, where $x'x^{L/2}$ is the name of a fragment of this trajectory corresponding to its path in the gadget $\Delta_{s-1}$. The word $x^{L/2}$ has length $L/2$ and satisfies (16). The word $x'$ is the name of a fragment of the trajectory which goes from lower interval to an interval generating trajectory with name $x^{L/2}$. We have $l(bx'x^{L/2}) \leq 2L = 4l(x^{L/2})$. By (3) and (16) we obtain for these initial fragments of sufficiently large length

$$\frac{K(bx'x^{L/2})}{l(bx'x^{L/2})} \geq \frac{K(x^{L/2}) - 2 \log l(bx')}{{4l(x^{L/2})}} \geq \frac{1}{4} - \frac{1}{h_{s-2}}.\quad (20)$$

We have $P^{s-1}(b) > (1/8)P(b)$ by induction hypothesis. After that, taking into account that $\gamma \leq 1$, we deduce from (19)

$$P(D_b) \geq P^{s-1}(D_b) \geq \frac{\gamma}{128} P(b).$$

6 Remember, that $L (\geq 2h_{s-1})$ is the height of gadgets $\Pi_{s-1}$, $\Delta_{s-1}$. 

12
By Proposition 5 an $c \in D_b$ exists such that

$$d_P(c^j) \leq d_P(b) + 1 - \log \frac{\gamma}{128} \leq d_P(b) + (\sigma(h_{s-1}) - \sigma(h_{s-2}) - 12) + 8 \leq \sigma(h_{s-1}) - 4 = \sigma(h_{s(k-1)}) - 4$$

for all $l(b) \leq j \leq l(c)$. Here we have $d_P(b) \leq \sigma(h_{s(k-2)}) \leq \sigma(h_{s-2})$ by induction hypothesis. We also used inequality (17). Besides, by induction hypothesis we have $l(b) \geq h_{s-1}$. Therefore,

$$d_P(c^j) < \sigma(h_{s-1}) \leq \sigma(l(b)) \leq \sigma(j)$$

for $l(b) \leq j \leq l(c)$. Define $\alpha(k) = c$. It is easy to see that all induction hypotheses are true for $\alpha(k)$.

An infinite sequence $\alpha$ is defined by a sequence of initial fragments (14). We proved that $d_P(\alpha^j) \leq \sigma(j)$ for all $j \geq l(\alpha(1))$.

By the construction there are infinitely many initial fragments of the sequence $\alpha$ satisfying (20). The sequence $h_s$, where $s = 0, 1, \ldots$, is monotone increased. So, the condition (8) hold. △

4 Non-robustness property of universal codes

Let $A$ and $B$ be finite alphabets. By a code we mean a computable family of functions $\phi_n : A^n \rightarrow B^*$, where $n = 1, 2, \ldots$. Suppose that $B = \{0, 1\}$. We will consider decodable codes. A computable family of decoding functions $\psi_n : \phi_n(A^n) \rightarrow A^n$ such that $\alpha = \psi_n(\phi_n(\alpha))$ for all $n$ and for all $\alpha \in A^n$ is associated with this code. A separating property of the code is required. An algorithm must exist decoding any sequence of concatenated codewords. Prefix codes satisfy to this requirement. Any two codewords $\phi_n(\alpha)$ and $\phi_n(\alpha')$ are incomparable under prefix method of coding. For any code $\{\phi_n\}$ a compressing ratio $\rho_{\phi_n}(\alpha^n) = l(\phi_n(\alpha^n))/(n \log |A|)$ of input word $\alpha^n \in A^n$ is defined. We suppose for simplicity that $A = \{0, 1\}$.

In [17, 18] codes universal in the mean for some classes of sources were considered, in [1, 2] a code universal almost everywhere for the class of all stationary ergodic sources was defined. We consider codes universal almost everywhere.

7 A function $\phi_n(\alpha)$ is computable by both arguments $n$ and $\alpha$. 

13
A code \( \{ \phi_n \} \) is called \textit{universal} with respect to a class of stationary ergodic sources if for any computable stationary ergodic measure \( P \) from this class

\[
\lim_{n \to \infty} \rho_{\phi_n}(\omega^n) = H
\]  

(21)

holds \( P \)-almost every infinite sequence \( \omega = \omega_1 \omega_2 \ldots \), where \( H \) is the entropy of the measure \( P \). There exist several types of Lempel - Ziv universal coding scheme \([1, 2]\). Let us recall two of them.

A coding algorithm is fed with a word \( \omega_1 \ldots \omega_N \) of length \( N \). By the first variant of the algorithm a sequence of letters \( \omega_1, \omega_2 \ldots \omega_n \) is read beginning at the left and is divided on subblocks as follows: a pointer on \( k \)-th subblock is inserted after \( \omega_i(k) \) if subblock \( \omega_i(k-1)+1 \omega_i(k-1)+2 \ldots \omega_i(k) \) was already seen between previous pointers and subblock \( \omega_i(k-1)+1 \omega_i(k-1)+2 \ldots \omega_i(k) \) was not seen. To encode new subblock it is sufficient to memorize coordinate of the beginning of the sequence \( \omega_i(k-1)+1 \omega_i(k-1)+2 \ldots \omega_i(k) \), its length, and new letter \( \omega_i(k) \).

The same idea is used in the second variant of the algorithm but a sub-block \( \omega_i(k-1)+1 \omega_i(k-1)+2 \ldots \omega_i(k) \) is deemed to have appeared if it occurs at all - not necessary between pointers.

The following proposition on non-robustness of universal codes is an analog of Theorem 1

**Proposition 2** For any nonnegative, nondecreasing, and unbounded function \( \sigma(n) \) and for any real number \( 0 < \epsilon < 1/4 \) a computable with respect to \( \sigma \) stationary ergodic measure \( P \) with entropy \( 0 < H \leq \epsilon \) exists such that for each universal (for class of all stationary ergodic sources) code \( \{ \phi_n \} \) an infinite binary sequence \( \alpha \) exists such that \( d_P(\alpha^n) \leq \sigma(n) \) for almost all \( n \) and

\[
\limsup_{n \to \infty} \rho_{\phi_n}(\alpha^n) \geq \frac{1}{4};
\]  

(22)

\[
\liminf_{n \to \infty} \rho_{\phi_n}(\alpha^n) \leq \epsilon.
\]  

(23)

**Proof.** For any \( n \) a decoding algorithm \( \psi_n \) of the code \( \{ \phi_n \} \) is defined by \( \log n + O(1) \) bits. Then we have

\[
K(\alpha^n) \leq l(\phi_n(\alpha)) + O(\log n).
\]  

(24)

Inequality (22) follows from the inequality (8) of Theorem 1. The proof of the inequality (23) is analogous to the proof of the inequality (9) of Theorem 1. We must only replace condition (15) from the proof of Theorem 1 on
Let \( \{ \phi_n \} \) be a code. Under block realization of the code any sequence of letters \( \omega^n = \omega_1 \ldots \omega_n \) is divided in consecutive blocks \( \omega = \tilde{\omega}_1 \ldots \tilde{\omega}_k \), where \( n = (k - 1)N + q \), \( 0 \leq q < N \) and \( \tilde{\omega}_i = \omega_{(i-1)N} \ldots \omega_{iN}, i = 1, 2, \ldots k - 1 \), is a block of length \( N \), and \( \tilde{\omega}_k = \omega_{(k-1)N} \ldots \omega_{(k-1)N+q} \) is the last incomplete block. Any block \( \tilde{\omega}_i \) is encoded by a binary word \( \phi_N(\tilde{\omega}_i) \). In asymptotic estimates (when \( n \to \infty \)) method of coding of this last block \( \tilde{\omega}_k \) is unessential (we fix some of these methods). We write \( \phi_N(\omega^n) = \phi_N(\tilde{\omega}_1) \ldots \phi_N(\tilde{\omega}_k) \) and \( \rho_{\phi_N}(\omega^n) = l(\phi_N(\omega^n))/n \).

It is proved in [2] (Theorem 4) that for any stationary ergodic measure \( P \) with entropy \( H \) a property of asymptotic optimality holds for block realization of Lempel–Ziv code \( \{ \phi_N \} \) with blocks of length \( N \). Relation

\[
\lim_{N \to \infty} \limsup_{n \to \infty} \rho_{\phi_N}(\omega^n) = H
\]

holds for \( P \)-almost all \( \omega \). We can prove that equality \( 25 \) holds also for any sequence \( \omega \) random in sense of Martin-Löf with respect to a measure \( P \) (i.e. when \( d_P(\omega^n) = O(1) \) as \( n \to \infty \)).

The following analogue of Theorem 1 holds for block realization of codes with block length \( N \) and for codes using sliding window of length \( N \) (when a new letter of codeword depends only from \( N \) preceding letters of input word).

**Proposition 3** For any nonnegative, nondecreasing, and unbounded function \( \sigma(n) \) and for any real number \( 0 < \epsilon < 1/4 \) a computable with respect to \( \sigma \) stationary ergodic measure \( P \) with entropy \( 0 < H \leq \epsilon \) exists such that for each universal (for class of all stationary ergodic sources) code \( \{ \phi_N \} \) or for each universal code with sliding window of length \( N \) an infinite binary sequence \( \alpha \) exists such that \( d_P(\alpha^n) \leq \sigma(n) \) for almost all \( n \) and for any \( N \)

\[
\limsup_{n \to \infty} \rho_{\phi_N}(\alpha^n) \geq \frac{1}{4},
\]

and for all sufficiently large \( N \)

\[
\liminf_{n \to \infty} \rho(\phi_N(\alpha^n)) \leq \epsilon.
\]

The proof of this proposition is a small complication of the proof of Proposition 2.
Notice, that the property (26) is also hold for adaptive coding scheme, i.e. when coding algorithm depends on preceding blocks.

Using Theorem 1 it can be proved that non-robustness property holds for other well-known universal codes. For example, in [19] a universal forecasting measure \( \rho(\omega_1\ldots\omega_n) \) and a code \( \psi_n \) such that \( l(\psi_n(\omega_1\ldots\omega_n)) \leq -\log \rho(\omega_1\ldots\omega_n) + 1 \) were defined. This measure is defined as a mixture \( \rho(y) = \sum_{k=0}^{\infty} \lambda_k \rho_k(y) \) of measures \( \rho_k \) universal for Markov sources of order \( k \) constructed in the theory of universal coding [20]. Here \( \lambda_k \) is some optimal probability distribution on positive integer numbers (it can be defined \( \lambda_k = c_k k^{-1} \log^{-2} k \), where \( c \) is a constant) and \( \phi(k) \) is the corresponding codeword for a positive integer number \( k \): \( l(\phi(k)) = \log k + O(\log \log k) \). In [21] an universal code was constructed \( \psi(u) = \phi(l(u)) \psi_{l(u)}(u) \), where \( u \in B^* \). The universality conditions for the measure \( \rho \) and for the code \( \psi \) is the following: \( \text{for any stationary measure } \mu \text{ with entropy } H(\mu) \text{ for } \mu \text{-almost all } \omega \in \Omega \text{ the mean error of the forecast by measure } \mu \text{ tends to zero} \)

\[
\lim_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T} \log \frac{\mu(\omega_{t+1} | \omega_1 \ldots \omega_t)}{\rho(\omega_{t+1} | \omega_1 \ldots \omega_t)} = \lim_{t \to \infty} \log \frac{\mu(\omega_1 \ldots \omega_t)}{\rho(\omega_1 \ldots \omega_t)} = 0, \tag{28}
\]

and \( \lim_{n \to \infty} l(\psi(\omega^n))/n = \lim_{n \to \infty} -\log \rho(\omega^n)/n = H(\mu) \). It is easy to derive from the definition of the deficiency of randomness that the condition (28) is “robust under violation of randomness”, more correctly, it holds for any computable stationary measure \( \mu \) and for any infinite sequence \( \omega \) such that \( d_\mu(\omega^n) = o(n) \) as \( n \to \infty \). But the corresponding universal code \( \psi \) is non-robust for the class of all stationary ergodic sources. Since a decoding algorithm exists for the code \( \psi \) it holds \( K(\omega_1 \ldots \omega_n) \leq l(\psi(\omega_1 \ldots \omega_n)) + O(1) \leq -\log \rho(\omega_1 \ldots \omega_n) + O(\log n) \). Then by Proposition 2 there exists an \( \alpha \in \Omega \), such that the conclusion of this proposition holds, in particular, the condition (22) holds. The property (23) can be obtained as in the proof of Proposition 2 by universality of the code.

The property of asymptotic optimality can be robust for more narrow classes of stationary ergodic sources such that as i.i.d sequences of random variables or stationary Markov chains.

**Proposition 4** Let \( P \) be an arbitrary computable probability measure representing a stationary ergodic Markov chain of fixed order (in particular, i.i.d
sequence of random variables), $H$ is its entropy, $\{\phi_n\}$ is a variant of Lempel–Ziv compressing algorithm. Then for any infinite sequence $\omega$ if $d_P(\omega^n) = o(n)$ then equality (21) holds, and for block realization of this compressing scheme equality (22) holds.

The proof is based on constructive feature of the proof of results from [2]. The Birghoff’s ergodic theorem is also used in this proof that is in the case of Markov sources is a variant of the law of large numbers. This law holds for individual sequence $\omega$ when $d_P(\omega^n) = o(n)$ as $n \to \infty$.

5 Appendix 1

Bounded increase of the deficiency of randomness. In the proof of Theorem 1 a proposition on a bounded increase of the deficiency of randomness was used. Let $P$ be a measure, $P(x) \neq 0$ and a set $A$ consists of words $y$ such that $x \subseteq y$. Recall, that $P(\tilde{A}) = P(\cup\{\Gamma_y : y \in A\})$ for any $A \subseteq \{0,1\}^*$. Define $P(\tilde{A}|x) = P(\tilde{A})/P(x)$.

Proposition 5 Let $P$ be a measure, $x$ be a word, $P(x) \neq 0$ and a set $A$ consists of words $y$ such that $x \subseteq y$ and $P(\tilde{A}) > 0$. Then for any $0 < \mu < 1$ a subset $A' \subseteq A$ exists such that $P(\tilde{A'}) > \mu P(\tilde{A})$ and

$$d_P(\nu^n) \leq d_P(x) - \log(1 - \mu) - \log P(\tilde{A}|x)$$

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

Proof. We will use in the proof a notion of supermartingale [12]. A function $M$ is called $P$–supermartingale if it is defined on $\{0,1\}^*$ and satisfies conditions:

$$M(\Lambda) \leq 1;$$

$$M(x) \geq M(x0)P(0|x) + M(x1)P(1|x)$$

for all $x$,

where $P(\nu|x) = P(x\nu)/P(x)$ for $\nu = 0, 1$ (we put here $0/0 = 0 \ast \infty = 0$).

A supermartingale $M$ is lower semicomputable if the set $\{(r,x) : r < M(x)\}$, where $r$ is a rational number, is a range of some computable function. We will consider only nonnegative supermartingales.

Let us prove that the deficiency of randomness is bounded by a logarithm of some lower semicomputable supermartingale.

Lemma 1 Let $P$ be a computable probability measure. Then there exists a lower semicomputable $P$–supermartingale $M$ such that $d_P(x) \leq \log M(x)$ for all $x$. 

17
Proof. Let some optimal function \( \psi \) satisfying (4) defines the monotone complexity \( K_m(x) \). Define
\[
Q(\alpha) = \frac{B_1/2(\cup \{ \Gamma_p : \alpha \subseteq \psi(p) \})}{2 - \ell(\alpha)},
\]
where \( B_1/2(\Gamma_\alpha) = 2^{-\ell(\alpha)} \) is the uniform Bernoulli measure on the set of all binary sequences. It is easy to verify that \( Q(\Lambda) \leq 1 \) and \( Q(\alpha) \geq Q(\alpha 0) + Q(\alpha 1) \) for all words \( \alpha \). Then the function \( M(\alpha) = Q(\alpha)/P(\alpha) \) is a \( P \)-supermartingale.

Since for any \( \alpha \) the shortest \( p \) such that \( \alpha \subseteq \psi(p) \) is an element of the set from (29), we have inequality \( Q(\alpha) \geq 2^{-K_m(\alpha)} \), and so, \( d_P(\alpha) \leq \log M(\alpha) \).

Let \( d_P(x) \leq \log M(x) \), where \( M \) is lower semicomputable \( P \)-supermartingal.

Let us define a set \( A_1 = \{ y \in A : \exists j (l(x) \leq j \leq l(y) \text{ and } M(y^j) > \frac{1}{(1 - \mu)P(A|x)} M(x) ) \} \).

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.

By definition of supermartingale for any prefix free set \( B \) such that \( x \subseteq y \) for all \( y \in B \) inequality
\[
M(x) \geq \sum_{y \in B} M(y) P(y|x)
\]
holds. For any \( y \in A_1 \) let \( y^p \) be the initial fragment of \( y \) of maximal length such that \( \frac{M(y^p)}{M(x)} > \frac{1}{(1 - \mu)P(A|x)} \). The set \( \{ y^p : y \in A_1 \} \) is prefix free. Then by (30) we have
\[
1 \geq \sum_{y \in A_1} \frac{M(y^p)}{M(x)} P(y^p|x) > \frac{1}{(1 - \mu)P(A|x)} \sum_{y \in A_1} P(y^p|x) \geq \frac{1}{(1 - \mu)P(A|x)} P(\tilde{A}_1|x).
\]

From this we obtain \( P(\tilde{A}_1|x) < (1 - \mu)P(\tilde{A}|x) \). Define \( A' = A - \{ y \in A : z \subseteq y \text{ for some } z \in A_1 \} \).

Then \( P(\tilde{A}|x) > \mu P(\tilde{A}|x) \). For any \( y \in A' \) we have
\[
M(y^j) \leq M(x) \frac{1}{(1 - \mu)P(A|x)}
\]
for all \( l(x) \leq j \leq (y) \). The result of the proposition follows from inequality \( d_P(x) \leq \log M(x) \). △
Method of cutting and stacking. An arbitrary measurable mapping of the probability space into itself is called a transformation or a process. A transformation \( T \) preserves a measure \( P \) if \( P(T^{-1}(A)) = T(A) \) 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.

The simplest example of such transformation of the space \( A^\infty \) of all infinite sequences, where \( A = \{0, 1, \ldots, k - 1\} \) is some finite alphabet, is the (left) shift \( T \) defined by \( (T\omega)_i = \omega_{i+1} \) for all \( i = 1, 2, \ldots \). If the shift \( T \) preserves the measure \( P \) then this measure is called stationary, i.e.

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

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

Recall some notions of symbolic dynamics. We use a cutting and stacking method of constructing of ergodic processes [22] [23]. Recall the main notions and properties of this method. A column is a sequence of letters from \( A \). A subset \( A \) of the column is called the name of the column, \( L \) column and at all points of the top \( T \) trajectory \( \omega, T\omega, \ldots, T^{h-1}\omega \). 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 \( \hat{\Upsilon} \) 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 commonly used will now be defined. The distribution of a gadget \( \Upsilon \) with columns \( E_1, \ldots, E_n \) is a vector of probabilities

\[
\left( \frac{w(E_1)}{w(\Upsilon)}, \ldots, \frac{w(E_n)}{w(\Upsilon)} \right).
\]

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 \( \Upsilon_i, i = 1, \ldots, M \), according to a given probability vector \((\gamma_1, \ldots, \gamma_n)\) by cutting each column \( E_i = (L_{i,j} : 1 \leq j \leq h(E_i)) \) (and its intervals) into disjoint subcolumns \( E_{i,m} = (L_{i,j,m} : 1 \leq j \leq h(E_{i,m})) \) 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 \( E_1 * E_2 = (L_j : 1 \leq j \leq h(E_1) + h(E_2)) \) 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 \( E * \Upsilon \) 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 \( E_i * U_i \). A new gadget consists of the columns \( (E_i * U_i) \).

Let \( \Upsilon \) and \( \Lambda \) be two gadgets of the same width and with disjoint supports. A gadget \( \Upsilon * \Lambda \) is defined as follows. Let the columns of \( \Upsilon \) are \( (E_i) \). Cut \( \Lambda \) into

20
copies Λᵢ such that \( w(Λᵢ) = w(Eᵢ) \) for all \( i \). After that, for each \( i \) stack the gadget Λᵢ onto column \( Eᵢ \), i.e. we consider a gadget \( Eᵢ * Λᵢ \). The new gadget is the union of gadgets \( Eᵢ * Λᵢ \) for all \( i \). The number of columns of the gadget \( Y * Λ \) is the product of the number of columns of \( Y \) on the number of columns of \( Λ \).

The \( M \)-fold independent cutting and stacking of a single gadget \( Y \) is defined by cutting \( Y \) into \( M \) copies \( Yᵢ, i = 1, \ldots, M \), of equal width and successively independently cutting and stacking them to obtain \( Y^*(M) = Y₁ * \ldots * Y_M \).

**Remark 1.** Several examples of stationary measures constructed using cutting and stacking method are given in [22, 23]. We use in Section 3 a construction of a sequence of gadgets defining the uniform Bernoulli distribution on trajectories generated by them. This sequence is constructed using the following scheme. Let a partition \( π = (π₀, π₁) \) be given. Let also \( Δ \) be a gadget such that its columns have the same width and are compatible with the partition \( π \). Let \( λ(Δ ∩ π₀) = λ(Δ ∩ π₁) \). Suppose that for some \( M \) a gadget \( Δ' \) is constructed from the gadget \( Δ \) by means of \( M \)-fold independent cutting and stacking and \( P \) be a measure on trajectories of the gadget \( Δ' \) defined by (31). Then by the method of cutting and stacking \( P(x) = 2^{-l(x)}λ(Δ) \) for the trajectory \( x \) of any point from the support of \( Δ' \).

A sequence of gadgets \( \{Yᵢ\} \) is complete if

- \( \lim_{m \to \infty} w(Yᵢ) = 0; \)
- \( \lim_{m \to \infty} λ(Ŷᵢ) = 1; \)
- \( Yᵢ+₁ \) extends \( Yᵢ \) for all \( m \).

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

By definition \( T \) preserves the measure \( λ \). In [22] and [23] the conditions sufficient a process \( T \) to be ergodic were suggested. Let a gadget \( Y \) is constructed by cutting and stacking from a gadget \( Λ \). Let \( E \) be a column from \( Y \) and \( D \) be a column from \( Λ \). Then \( E ∩ D \) 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 \( Λ \) is \((1 - \epsilon)\)-well-distributed in \( Y \) if

\[
\sum_{D \in Λ} \sum_{E \in Y} |λ(Ê ∩ Ê) - λ(Ê)λ(Ê)| < \epsilon.
\] (32)

We will use the following two lemmas.

**Lemma 2** ([22], Corollary 1), ([23], Theorem A.1). Let \( \{Yᵢ\} \) be a complete sequence of gadgets and for each \( n \) the gadget \( \{Yᵢ\} \) is \((1 - \epsilon_n)\)-well-distributed in \( \{Yᵢ+₁\} \), where \( \epsilon_n \to 0 \). Then \( \{Yᵢ\} \) defines the ergodic process.
Lemma 3 ([23], Lemma 2.2). For any $\epsilon > 0$ and any gadget $\Upsilon$ there is an $M$ such that for each $m \geq M$ the gadget $\Upsilon$ is $(1 - \epsilon)$-well-distributed in the gadget $\Upsilon^*(m)$ constructed from $\Upsilon$ by $m$-fold independent cutting and stacking.

References

[1] Lempel A., Ziv J. A Universal Algorithm for Sequential Data Compression // IEEE Trans. Inform. Theory. 1977. V.23. N3. P.337–343.

[2] Lempel A., Ziv J. Compression of Individual Sequences via Variable Rate Coding // IEEE Trans. Inform. Theory. 1978. V.24. N5. P.530–536.

[3] Kolmogorov A.N. The Logical Basis for Information Theory and Probability Theory // IEEE Trans. Inf. Theory. 1968. V. 14, P. 662–664.

[4] Uspensky V.A., Semenov A.L., Shen A.Kh. Can an Individual Sequence of Zeros and Ones be Random? // Russian Math. Surveys. 1990. V. 45. P. 121–189.

[5] Li M., Vitányi P. An introduction to Kolmogorov complexity and its applications. New York: Springer–Verlag. 1997.

[6] Martin-Löf P. The Definition of Random Sequences // Inform. and Control. 1966. V.9. N6. P.602–619.

[7] Vovk V.G. The Law of the Iterated Logarithm for Random Kolmogorov, or Chaotic Sequences // SIAM Theory Probab. Applic. 1987, V. 32. P. 413–425.

[8] Schnorr C.P. A Unified Approach to the Definition of Random Sequences // Mathematical Systems Theory. 1971. V.5 P.246–258.

[9] V’yugin V.V. Non-robustness Property of the Individual Ergodic Theorem // Probl. Inform. Transm. 2001. V.37. P.27–39.

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

[11] Zvonkin A.K. and Levin L.A. The Complexity of Finite Objects and the Algorithmic Concepts of Information and Randomness // Russ. Math. Surv. V.25. P.83–124.
[12] Shiryaev A.N. Probability. Berlin: Springer. 1984.

[13] Levin L.A. On the Notion of Random Sequence // Soviet Math. Dokl. V.14. P.1413–1416.

[14] Kolmogorov A.N., Uspensky, V.A. Algorithms and Randomness // Theory Probab. Applic. 1987. V. 32. P. 389–412.

[15] Billingsly P. Ergodic theory and Information. New York: Wiley. 1965.

[16] V’yugin V.V. Ergodic Theorems for Individual Random Sequences // Theoretical Computer Science. 1998. V.207. N4. P.343-361.

[17] Fittinghof B.M. Optimal Coding in the Case of Unknown and Changing Message Statistics// Probl. Inform. Transm. 1966. V.2. N2. P.3–11.

[18] Davisson L.D. Universal Noiseless Coding // IEEE Trans. Inform. Theory. 1973. V.19. P.783–795.

[19] Ryabko B. Prediction of Random Sequences and Universal Coding // Probl. Inform. Transm. 1988. V.24. P.3–14.

[20] Krichevsky R.E., Trofimov V.K. The Performance of Universal Coding // IEEE Trans. Inform. Theory. 1981. V.27. N2. P.199–207.

[21] Ryabko B. Twice Universal Coding // Probl. Inform. Transm. 1984. V.20. P.173–178.

[22] Shields P.C. Cutting and Stacking: a Method for Constructing Stationary Processes // IEEE Trans. Inform. Theory. 1991. V.37. N6. P.1605–1617.

[23] Shields P.C. Two Divergence-Rate Counterexamples // J. Theoret. Probability. 1993. V.6. P.521–545.