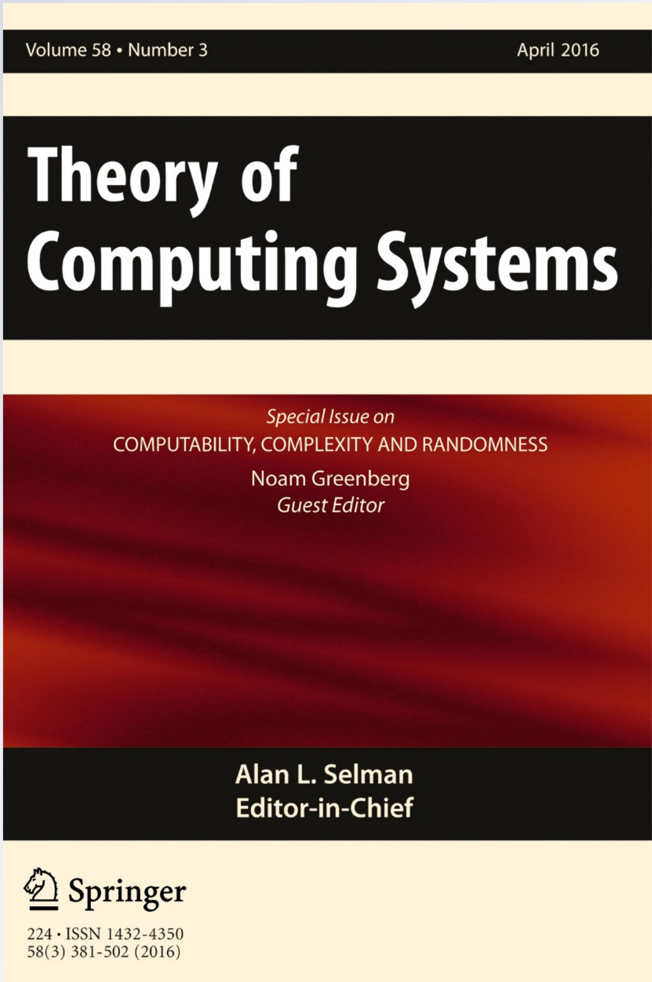
On Stability of Probability Laws with Respect to Small Violations of Algorithmic Randomness

Vladimir V. V'yugin

Theory of Computing Systems

ISSN 1432-4350
Volume 58
Number 3

Theory Comput Syst (2016) 58:403-423
DOI 10.1007/s00224-015-9632-6



Springer

Theory Comput Syst (2016) 58:403–423
DOI 10.1007/s00224-015-9632-6

On Stability of Probability Laws with Respect to Small Violations of Algorithmic Randomness

Vladimir V. V'yugin[1,2]

Published online: 24 April 2015

Abstract We study a stability property of probability laws with respect to small violations of algorithmic randomness. Some sufficient condition of stability is presented in terms of Schnorr tests of algorithmic randomness. Most probability laws, like the strong law of large numbers, the law of iterated logarithm, and even Birkhoff's pointwise ergodic theorem for ergodic transformations, are stable in this sense. Nevertheless, the phenomenon of instability occurs in ergodic theory. Firstly, the stability property of Birkhoff's ergodic theorem is non-uniform. Moreover, a computable non-ergodic measure-preserving transformation can be constructed such that the ergodic theorem is non-stable.

1 Introduction

In this paper we study a stability property of probability laws with respect to small violations of randomness. By a probability law we mean any property $\Phi(\omega)$ of infinite binary sequences ω which holds almost surely. We define the notion of stability of a probability law in terms of algorithmic theory of randomness. Within the frame-

This paper is an extended version of the talk at the Eighth International Conference on Computability, Complexity and Randomness (CCR 2013), September 23-27, 2013, Moscow, Russia; see also the conference paper V'yugin [26]. This work was partially supported by RFBR grants 13-01-12458 and 13-01-12447.

✉ Vladimir V. V'yugin
vyugin@iitp.ru

1 Institute for Information Transmission Problems, Moscow, Russia

2 National Research University Higher School of Economics, Moscow, Russia

work of this theory, the probability laws are formulated in a "pointwise" form. It is well known that the main laws of probability theory are valid not only almost surely but for each individual Martin-Löf random sequence.

Some standard notions of algorithmic randomness are given in Section 2. We use the definition of a random sequence in complexity terms. An infinite binary sequence $\omega_1\omega_2\dots$ is Martin-Löf random with respect to the uniform (or 1/2-Bernoulli) measure if and only if $Km(\omega^n) \ge n - O(1)$ as $n \to \infty$, where $Km(\omega^n)$ is the monotonic Kolmogorov complexity of a binary string $\omega^n = \omega_1 \dots \omega_n$ and the constant $O(1)$ depends on ω but not on n.

The same property also holds if we replace the monotonic complexity $Km(\omega^n)$ with the prefix complexity $KP(\omega^n)$. The difference is that the inequality $Km(\omega^n) \le n+O(1)$ holds for monotonic complexity but this is not true for the prefix complexity. The main results of this paper, Theorems 2 and 3, also hold for the prefix complexity.

A probability law $\Phi(\omega)$ is said to be stable if an unbounded computable function $\sigma(n)$ exists such that $\Phi(\omega)$ is true for each infinite sequence ω such that $Km(\omega^n) \ge n - \sigma(n) - O(1)$ as $n \to \infty$. We assume that this function is non-decreasing and refer to the function $\sigma(n)$ as to a degree of stability.

A stability property under small violations of algorithmic randomness of the main limit probability laws was discovered by Schnorr [16] and Vovk [20]. They have shown that the law of large numbers for the uniform Bernoulli measure holds for a binary sequence $\omega_1\omega_2\dots$ if $Km(\omega^n) \ge n - \sigma(n) - O(1)$, 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 $Km(\omega^n) \ge n - \sigma(n) - O(1)$, where $\sigma(n)$ is an arbitrary computable function such that[1] $\sigma(n) = o(\log\log n)$. V'yugin [22] has shown that the law of the length of the longest head-run in an individual random sequence is stable with degree of stability $\sigma(n) = o(\log\log n)$. It was shown in these papers that the corresponding degrees of stability are tight.

We present in Proposition 4 a sufficient condition of stability in terms of Schnorr tests of randomness. We mention that if a computable rate of convergence almost surely exists, then the corresponding probability law holds for any Schnorr random sequence. In turn, the latter property implies a stability property of this law. Using this sufficient condition, we prove that most probability laws, like the strong law of large numbers and the law of iterated logarithm, are stable under small violations of algorithmic randomness. Theorem 1 shows that Birkhof's ergodic theorem is also stable if the measure preserving transformation is ergodic.

In Section 4 we show that the phenomenon of instability occurs in ergodic theory. First, there are no universal stability bounds in ergodic theorems for ergodic transformations. The Birkhoff ergodic theorem is non-stable for some non-ergodic stationary measure-preserving transformation.

We note that there is some analogy with the lack of universal convergence rate or redundancy estimates in ergodic theory. A lack of universal convergence bounds

[1] In what follows all logarithms are on the base 2.

is typical for asymptotic results of ergodic theory like the Birkhoff ergodic theorem – Krengel [12], or the Shannon–McMillan–Breiman theorem and universal compressing schemes –Ryabko [15].

2 Preliminaries

Let $\{0,1\}^*$ be the set of all finite binary sequences (strings) and $\Omega = \{0,1\}^\infty$ the set of all infinite binary sequences. Denote by Λ the empty sequence. Let $l(\alpha)$ denote the length of a sequence α ($l(\alpha) = \infty$ for $\alpha \in \Omega$).

For any finite or infinite sequence $\omega = \omega_1\omega_2\dots$, we write $\omega^n = \omega_1\omega_2\dots\omega_n$, where $n \le l(\omega)$. Also, we write $\alpha \subseteq \beta$ if $\alpha = \beta^n$ for some n. Two finite sequences α and β are incomparable if $\alpha \not\subseteq \beta$ and $\beta \not\subseteq \alpha$. A set $A \subseteq \{0,1\}^*$ is prefix-free if any two distinct sequences from A are incomparable.

The complexity of a string $x \in \{0,1\}^*$ is equal to the length of the shortest binary codeword $p \in \{0,1\}^*$ from which the string x can be reconstructed: $K_\psi(x) = \min\{l(p) : \psi(p) = x\}$. We suppose that $\min\emptyset = +\infty$.

By this definition the complexity of x depends on a computable (partial recursive) function ψ, a method of decoding. Kolmogorov proved that the optimal decoding algorithm ψ exists such that $K_\psi(x) \le K_{\psi'}(x) + O(1)$ holds for each computable decoding function ψ' and for all strings x. We fix some optimal decoding function ψ. The value $K(x) = K_\psi(x)$ is called the Kolmogorov complexity of x.

If domains of decoding algorithms are prefix-free sets, the same construction gives us the definition of prefix complexity $KP(x)$.

Let $\mathcal{R}$ be the set of all real numbers and $\mathcal{Q}$ the set of all rational numbers.

A function $f : \{0,1\}^* \to \mathcal{R}$ is said to be computable if there exists an algorithm which, given a finite string x and a rational number $\epsilon > 0$, computes a rational approximation of a number $f(x)$ with accuracy ϵ.

For a general reference on algorithmic randomness, see Li and Vitányi [13]. We confine our attention to the Cantor space Ω with the uniform Bernoulli measure $B_{1/2}$. Hoyrup and Rojas [9] proved 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.

The topology on Ω is generated by binary intervals $\Gamma_x = \{\omega \in \Omega : x \subset \omega\}$, where x is a finite binary sequence.

A probability measure P on Ω can be defined by the values $P(x) = P(\Gamma_x)$, $x \in \{0,1\}^*$. Also, $P(\Lambda) = 1$ and $P(x) = P(x0) + P(x1)$ for all x. A measure P is computable if the function $x \to P(x)$ is computable.

An important example of a computable probability measure is the uniform Bernoulli measure $B_{1/2}$, where $B_{1/2}(\Gamma_x) = 2^{-l(x)}$ for any finite binary sequence x.

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

A Martin-Löf test of randomness with respect to a computable measure P is an effectively enumerable sequence U_n, $n = 1, 2, \ldots$, of effectively open sets such that $P(U_n) \leq 2^{-n}$ for all n. If the real numbers $P(U_n)$ are uniformly computable, then the test U_n is called a Schnorr test of randomness.[2]

An infinite binary sequence ω passes the test U_n, $n = 1, 2, \ldots$, if $\omega \notin \bigcap U_n$. A sequence ω is Martin-Löf random with respect to a computable measure P if it passes each Martin-Löf test of randomness. The notion of a Schnorr random sequence is defined analogously.

In what follows we mainly consider the notion of randomness with respect to the uniform Bernoulli measure $B_{1/2}$.

An equivalent definition of randomness can be obtained using Solovay tests of randomness. A computable sequence $\{x_n : n = 1, 2, \ldots\}$ of binary strings is called a Solovay test of randomness with respect to the uniform measure if the series $\sum\limits_{n=1}^{\infty} 2^{-l(x_n)}$ converges.

An infinite sequence ω passes a Solovay test of randomness $\{x_n : n = 1, 2, \ldots\}$ if $x_n \not\subseteq \omega$ for almost all n.

The Martin-Löf and Solovay tests define the same class of random sequences.

Proposition 1 *An infinite sequence $\omega = \omega_1\omega_2\ldots$ is Martin-Löf random if and only if it passes each Solovay test of randomness.*

Proof Assume that ω is not Martin-Löf random. Then a Martin-Löf test U_n, $n = 1, 2, \ldots$, exists such that $\omega \in \bigcap U_n$. Define the Solovay test of randomness as follows. Since U_n is effectively open and $B_{1/2}(U_n) \leq 2^{-n}$ for all n, we can effectively compute a prefix-free sequence of strings x_n, $n = 1, 2, \ldots$, such that $\bigcup_n \Gamma_{x_n} = \cup_n U_n$ and the series $\sum\limits_{n=1}^{\infty} 2^{-l(x_n)}$ converges. Obviously, $x_n \subset \omega$ for infinitely many n.

On the other hand, assume that for some Solovay test x_n, $n = 1, 2, \ldots$, $x_n \subset \omega$ for infinitely many n. Let $\sum\limits_{n=1}^{\infty} 2^{-l(x_n)} < 2^K$, where m is a positive integer number. Let U_n be the set of all infinite ω such that $|\{m : x_m \subset \omega\}| \geq 2^{n+K}$. It is easy to verify that U_n is a Martin-Löf test of randomness and that $\omega \in \bigcap U_n$. □

We also consider total Solovay tests of randomness, which leads to the same definition of randomness as with Schnorr tests of randomness (see Downey and Griffiths [4]). A series $\sum\limits_{i=1}^{\infty} r_i$ converges with a computable rate of convergence if a

[2]Uniform computability of $P(U_n)$ means that there is an algorithm which, given n and $\epsilon > 0$, outputs a rational approximation of $P(U_n)$ up to ϵ.

computable function $m(\delta)$ exists such that $|\sum_{i=m(\delta)}^{\infty} r_i| \le \delta$ for each positive rational number δ. A Solovay test of randomness $\mathcal{T} = \{x_n : n = 1, 2, \dots\}$ is said to be total if the series $\sum_{n=1}^{\infty} 2^{-l(x_n)}$ converges with a computable rate of convergence.

Proposition 2 *An infinite sequence $\omega = \omega_1\omega_2\dots$ is Schnorr random if and only if it passes each total Solovay test of randomness.*

The proof is similar to the proof of Proposition 1.

An equivalent definition of a Martin-Löf random sequence is obtained in terms of algorithmic complexity (see Li and Vitanyi [13]).

In terms of prefix complexity the following definition is known. An infinite sequence ω is Martin-Löf random with respect to a computable measure P if and only if $KP(\omega^n) \ge -\log P(\omega^n) + O(1)$.

An analogous definition can be obtained in terms of monotonic complexity. Let us define the notion of a monotonic computable transformation of binary sequences. A computable representation of such an operation is a set $\hat{\psi} \subseteq \{0,1\}^* \times \{0,1\}^*$ such that (i) the set $\hat{\psi}$ is recursively enumerable; (ii) for any $(x, y), (x', y') \in \hat{\psi}$, if $x \subseteq x'$, then $y \subseteq y'$ or $y' \subseteq y$; (iii) if $(x, y) \in \hat{\psi}$, then $(x, y') \in \hat{\psi}$ for all $y' \subseteq y$.

The set $\hat{\psi}$ defines a monotonic (with respect to $\subseteq$) decoding function[3] $\psi(p) = \sup\{x : \exists p'(p' \subseteq p \& (p', x) \in \hat{\psi})\}$.

Any computable monotonic function ψ determines the corresponding measure of complexity $Km_\psi(x) = \min\{l(p) : x \subseteq \psi(p)\} = \min\{l(p) : (x, p) \in \hat{\psi}\}$. An invariance property also holds for monotonic measures of complexity: an optimal computable operation ψ exists such that $Km_\psi(x) \le Km_{\psi'}(x) + O(1)$ for all computable operations ψ' and for all finite binary sequences x.

An infinite sequence ω is Martin-Löf random with respect to a computable measure P if and only if $Km(\omega^n) = -\log P(\omega^n) + O(1)$. In particular, an infinite binary sequence ω is Martin-Löf random (with respect to the uniform measure) if and only if $Km(\omega^n) = n + O(1)$.

A randomness criterium can also be formulated in terms of prefix complexity: an infinite sequence ω is Martin-Löf random with respect to a computable measure P if and only if $KP(\omega^n) \ge -\log P(\omega^n) + O(1)$. Also, $Km(x) \le KP(x) + O(1)$ (see for details Li and Vitanyi [13]). The main results of this paper, Theorems 2 and 3, also hold if we replace $Km(x)$ with $KP(x)$.

The function $dm_P(\omega^n) = -\log P(\omega^n) - Km(\omega^n)$ is called the universal deficiency of randomness (with respect to a computable measure P). For the uniform measure, $dm(\omega^n) = n - Km(\omega^n)$.

[3] Here by the supremum we mean a finite or an infinite sequence extending all comparable finite x.

3 Algorithmically Stable Laws

Let $\Phi(\omega)$ be an asymptotic probability law, i.e., a property of infinite binary sequences which holds almost surely.

Kolmogorov's algorithmic approach to probability theory offers a new paradigm for logic of probability. We can formulate any probabilistic law in a pointwise form: $Km(\omega^n) \geq n - O(1) \Longrightarrow \Phi(\omega)$.[4]

In this paper we present a more deep analysis. We call a law $\Phi(\omega)$ stable if there exists an unbounded nondecreasing computable function $\alpha(n)$ such that $Km(\omega^n) \geq n - \alpha(n) - O(1) \Longrightarrow \Phi(\omega)$. The function $\alpha(n)$ is called the degree of stability of the law $\Phi(\omega)$.

3.1 Sufficient Condition of Stability

We present in this section some sufficient condition of stability of a probability law and consider examples of such laws with different degrees of stability. We formulate this sufficient condition in terms of Schnorr's [16] definition of an algorithmic random sequence. The choice of Schnorr's definition is justified by an observation that the vast majority of such laws hold for Schnorr random sequences.

An algorithmic effective version of almost sure convergence of functions f_n of type $\Omega \rightarrow \mathcal{R}^+$ was considered by V'yugin [21]. A sequence of functions f_n effectively converges to a function f almost surely if a computable function $m(\delta, \epsilon)$ exists such that

$$B_{1/2}\{\omega : \sup_{n \geq m(\delta,\epsilon)} |f_n(\omega) - f(\omega)| > \delta\} < \epsilon \quad (1)$$

for all positive rational numbers δ and ϵ.

A function $f : \Omega \rightarrow \mathcal{R}$ is said to be computable if the sets $\{(r, \omega) : r \in \mathcal{Q}, \omega \in \Omega, r < f(\omega)\}$ and $\{(r, \omega) : r \in \mathcal{Q}, \omega \in \Omega, r > f(\omega)\}$ are effectively open in the product topology[5] on $\mathcal{Q} \times \Omega$. A notion of a computable sequence of functions $f_n : \Omega \rightarrow \mathcal{R}, n = 1, 2, \dots,$ is defined analogously.

The following simple proposition was formulated in [21] for the Martin-Löf notion of randomness. It holds also for Schnorr random sequences (see Galatolo et al. [10]).

Proposition 3 *Let a computable sequence of functions f_n effectively converge almost surely to some function f. Then a Schnorr test of randomness $\mathcal{T}$ can be constructed such that $\lim_{n\rightarrow\infty} f_n(\omega) = f(\omega)$ for each infinite sequence ω passing the test $\mathcal{T}$.*

Proof By (1) we have $B_{1/2}\{\omega : \sup_{n,n' \geq m(\delta/2,\epsilon)} |f_n(\omega) - f_{n'}(\omega)| > \delta\} < \epsilon$ for all positive rational numbers δ and ϵ. Denote $W_{n,n',\delta} = \{\omega : |f_n(\omega) - f_{n'}(\omega)| > \delta\}$. This set can be represented as the union $\bigcup_i \Gamma_{x_i}$, where $x_i, i = 1, 2, \dots,$ is a computable

[4] An equivalent form is $Km(\omega^n) = n + O(1) \Longrightarrow \Phi(\omega)$.

[5] We consider a discrete topology on $\mathcal{Q}$ and a topology on Ω generated by the intervals $\Gamma_x, x \in \{0, 1\}^*$. The notion of an effectively open set is defined as Section 2.

sequence of finite sequences. Define $V_i = \bigcup_{n,n' \ge m(1/i,2^{-i})} W_{n,n',1/i}$ for all i and $U_i = \bigcup_{j>i} V_j$. Then $B_{1/2}(U_i) \le 2^{-i}$ for all i.

Note that the measure $B_{1/2}(U_i)$ can be computed with an arbitrary degree of precision. Indeed, by (1), to compute $P(U_i)$ with a given degree of precision $\epsilon > 0$ it is sufficient to compute $B_{1/2}(\bigcup_{i' \ge j \ge i} \bigcup_{m' \ge n,n' \ge m(1/i,2^{-j})} W_{n,n',1/j})$ for some sufficiently large i' and m'. Therefore, $\mathcal{T} = \{U_i\}$ is a Schnorr 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 $n, n' \ge m(1/j, 2^{-j})$ exist such that $\omega \in W_{n,n',1/j} \subseteq V_j$. Hence, the sequence ω is rejected by the Schnorr test $\mathcal{T}$. □

In the following proposition some sufficient condition of stability of a probability law is given in terms of total Solovay tests randomness. This proposition also follows from the proof of Proposition 13 of Bienvenu and Merkle [2].

Proposition 4 *For any total Solovay test of randomness $\mathcal{T}$, a computable unbounded function $\sigma(n)$ exists such that for any infinite sequence ω, if $Km(\omega^n) \ge n - \sigma(n) - O(1)$, then the sequence ω passes the test $\mathcal{T}$.*

Proof Let $\mathcal{T} = \{x_n : n = 1, 2, \dots\}$. Denote $l_s = l(x_s)$. Since $\sum\limits_{s=1}^{\infty} 2^{-l_s} < \infty$ with a uniform computable rate of convergence $m(\epsilon)$, an unbounded nondecreasing computable function $\nu(n)$ exists such that $\sum\limits_{s=1}^{\infty} 2^{-l_s+\nu(l_s)} < \infty$. We can define $\nu(n) = i$, where i is such that $m(2^{-2i}) \le n < m(2^{-2(i+1)})$. Then

$$\sum_{s=1}^{\infty} 2^{-l_s+\nu(l_s)} = \sum_{i=1}^{\infty} 2^i \sum_{m(2^{-2i}) \le l_s < m(2^{-2(i+1)})} 2^{-l_s} \le \sum_{i=1}^{\infty} 2^{-i} \le 1.$$

By the generalized Kraft inequality (see Li and Vitanyi [13]), we can define the corresponding prefix-free code such that $Km(x_m) \le l(x_m) - \nu(l(x_m)) + O(1)$. Assume $x_m \subseteq \omega$ for infinitely many m. For any such m, $\omega^n = x_m$, where $n = l(x_m)$.

Let $\sigma(n)$ be a unbounded nondecreasing computable function such that $\sigma(n) = o(\nu(n))$ as $n \to \infty$. Let also ω be an infinite binary sequence such that $Km(\omega^n) \ge n - \sigma(n) - O(1)$ for all n. For $n = l(x_m)$,

$$\sigma(n) \ge n - Km(\omega^n) \ge n - l(x_m) + \nu(l(x_m)) - O(1) \ge \nu(n) - O(1)$$

for infinitely many n. On the other hand, $\sigma(n) = o(\nu(n))$ as $n \to \infty$. This contradiction proves the theorem. □

By Proposition 4 the stability property holds for main probability laws like the strong law of large numbers and the law of iterated logarithm.

By a computable sequence of total Solovay tests of randomness we mean a computable double-indexed sequence of finite binary strings $\mathcal{T}_k = \{x_{k,n} : n =$

$1, 2, \ldots\}$, $k = 1, 2, \ldots$, such that the series $\sum\limits_{n=1}^{\infty} 2^{-l(x_{k,n})}$ converges with a uniformly (with respect to k) computable rate of convergence. This means that there exists a computable function $m(\delta, k)$ such that $\sum\limits_{i=m(\delta,k)}^{\infty} 2^{-l(x_{k,i})} \le \delta$ for each k and each rational[6] δ.

In applications, it is often convenient to use computable sequences of tests. One can easily modify Proposition 4 for computable sequences of tests.

Proposition 5 *For any computable sequence of Solovay total tests of randomness $\mathcal{T}_k$, $k = 1, 2, \ldots$, a computable unbounded function $\sigma(n)$ exists such that for any infinite sequence ω, if $Km(\omega^n) \ge n - \sigma(n) - O(1)$, then the sequence ω passes all tests $\mathcal{T}_k$.*

The proof is analogous to the proof of Proposition 4.

It is well known that the Schnorr randomness satisfies the strong law of large numbers and the law of iterated logarithm. Now we show details of how Proposition 5 can be applied to these laws.

Hoeffding's [8] inequality for the uniform probability distribution

$$B_{1/2}\left\{\omega \in \Omega : \left|\frac{1}{n}\sum_{i=1}^{n}\omega_i - \frac{1}{2}\right| \ge \epsilon\right\} \le 2e^{-2n\epsilon^2} \tag{2}$$

serves as a tool for constructing total Solovay tests of randomness.

Let ϵ_k be a computable sequence of positive rational numbers such that $\epsilon_k \to 0$ as $k \to \infty$. For any k, let $\bigcup_n \{x : l(x) = n \& |\frac{1}{n}\sum\limits_{i=1}^{n} x_i - \frac{1}{2}| \ge \epsilon_k\} = \{x_{k,m} : m = 1, 2, \ldots\}\}$. This is a total Solovay test of randomness, since by (2) we have $\sum\limits_{m=1}^{\infty} 2^{-l(x_{k,m})} \le \sum\limits_{n=1}^{\infty} 2e^{-2n\epsilon_k^2} < \infty$ with a computable rate of convergence.

The strong law of large numbers $\lim\limits_{n\to\infty} \frac{1}{n}\sum\limits_{i=1}^{n}\omega_i = \frac{1}{2}$ holds for an infinite sequence $\omega = \omega_1\omega_2\ldots$ if and only if it passes the test $\{x_{k,m} : m = 1, 2, \ldots\}$ for each k. By Proposition 4 an unbounded nondecreasing computable function $\sigma(n)$ exists such that if $Km(\omega^n) \ge n - \sigma(n) - O(1)$ as $n \to \infty$, then the strong law of large numbers holds for this ω.

We can find a specific form of this function $\sigma(n)$ using the proof of Proposition 4. By inequality (2) we have the bound $\sum\limits_{n=1}^{\infty} 2^{-l(x_{k,n})} < \sum\limits_{n=1}^{\infty} 2e^{-2n\epsilon_k^2} < \infty$ for the corresponding total Solovay test of randomness $\mathcal{T}_k = \{x_{k,n}\}$. Also,

[6] We can combine all tests of a computable sequence $\mathcal{T}_k$, $k = 1, 2, \ldots$, into a single total test $\mathcal{T} = \{x_{k,n} : k = 1, 2, \ldots, n = m(2^{-k}, k), m(2^{-k}, k) + 1, \ldots\}$ such that if any ω passes the test $\mathcal{T}$, then it passes the test $\mathcal{T}_k$ for each k. $\mathcal{T}$ is a test, since $\sum\limits_{k=1}^{\infty}\sum\limits_{n=m(2^{-k},k)}^{\infty} 2^{-l(x_{k,n})} \le \sum\limits_{k=1}^{\infty} 2^{-k} \le 1$.

$\sum_{n=1}^{\infty} 2e^{-2n\epsilon_k^2+\nu(n)} < \infty$ for any function $\nu(n)$ such that $\nu(n) = o(n)$ as $n \to \infty$. The remaining part of the proof coincides with the proof of Proposition 4. Hence, any function $\sigma(n) = o(n)$ can serve as a degree of stability for the strong law of large numbers.

An analogous construction can be developed for the law of iterated logarithm:

$$\limsup_{n\to\infty} \frac{\left|\sum_{i=1}^{n} \omega_i - \frac{n}{2}\right|}{\sqrt{\frac{1}{2}n \ln\ln n}} = 1. \tag{3}$$

Here we consider here only the inequality $\le$ in (3).[7] This inequality is violated if and only if a rational number $\delta > 1$ exists such that $S_n - \frac{n}{2} > \delta\sqrt{\frac{1}{2}n \ln\ln n}$ for infinitely many n, where $S_n = \sum_{i=1}^{n} \omega_i$.

For any rational number δ such that $\delta > 1$ and for $m_n = \lceil \delta^n \rceil$, let[8]

$$U_{\delta,n} = \left\{\omega \in \Omega : \exists k(m_n \le k \le m_{n+1} \& S_k - k/2 > \delta\sqrt{(1/2)m_n \ln\ln m_n}\right\}.$$

Using the inequality $B_{1/2}\{\max_{1\le k\le m} S_k > a\} \le 2B_{1/2}\{S_m > a\}$, we obtain

$$B_{1/2}(U_{\delta,n}) \le 2B_{1/2}\left(\left\{\omega \in \Omega : S_{m_{n+1}} - m_{n+1}/2 > \delta\sqrt{(1/2)m_n \ln\ln m_n}\right\}\right) \le \\ \le ce^{-\delta \ln\ln m_n} \approx \frac{1}{n^{\delta}}, \tag{4}$$

where $c > 0$. We have used in (4) the Hoeffding inequality.

We can effectively construct a prefix-free set $\tilde{U}_{\delta,n}$ of finite sequences such that for each $\omega \in U_{\delta,n}$ a number m exists such that $\omega^m \in \tilde{U}_{\delta,n}$.

The sequence $\bigcup_n \tilde{U}_{\delta,n} = \{x_{\delta,k} : k = 1, 2, \dots\}$ is a total Solovay test of randomness, since the series $\sum_n 2^{-l(x_{\delta,n})} = \sum_n B_{1/2}(U_{\delta,n}) \le \sum_n \frac{1}{n^{\delta}}$ converges (with a computable rate of convergence) for any $\delta > 1$.

By definition, the law of iterated logarithm (3) holds for $\omega = \omega_1\omega_2\dots$ if and only if it passes the test $\{x_{\delta,k} : k = 1, 2, \dots\}$ for each $\delta > 1$.

By Proposition 4 an unbounded nondecreasing computable function $\sigma(m)$ exists such that the inequality $\le$ in (3) holds for any ω satisfying $Km(\omega^m) \ge m - \sigma(m) - O(1)$ as $m \to \infty$.

We can also find a specific form of the degree of stability for the law of iterated logarithm. Let $\alpha(m)$ be an unbounded nondecreasing computable function such that $\alpha(m) = o(\ln\ln m)$ as $m \to \infty$. Then the series $\sum_n e^{-\delta \ln\ln m_n + \alpha(m_n)} \approx \sum_n \frac{o(\ln n)}{n^{\delta}}$ converges for any $\delta > 1$. The proof of Proposition 4 shows that any computable

[7]The converse inequality is studied in Vovk [20].

[8]For any real number r, $\lceil r \rceil$ denotes the least positive integer number m such that $m \ge r$.

unbounded function $\sigma(n) = o(\log\log n)$ can serve as a measure of stability of the law of iterated logarithm.

3.2 Stability of Birkhoff's Theorem in the Ergodic Case

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

A transformation T of the set Ω is computable if a computable representation $\hat{\psi}$ exists such that (i)-(iii) hold and $T(\omega) = \sup\left\{y : x \subseteq \omega \& (x, y) \in \hat{\psi})\right\}$ for all infinite $\omega \in \Omega$.

Denote $T^0\omega = \omega, T^{i+1}\omega = T(T^i\omega)$. Any point $\omega \in \Omega$ generates an infinite trajectory $\omega, T\omega, T^2\omega, \ldots$.

Using Bishop's [3] analysis, V'yugin [21], [23] presented an algorithmic version of Birkhoff's pointwise ergodic theorem:

Let T be a computable measure-preserving transformation and f a computable real-valued bounded function defined on the set of binary sequences. Then for any infinite binary sequence ω the following implication is valid:

$$Km(\omega^n) \geq n - O(1) \Longrightarrow \lim_{n\to\infty} \frac{1}{n}\sum_{i=0}^{n-1} f(T^i\omega) = \hat{f}(\omega) \qquad (5)$$

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

Later this result was extended for non-computable f and generalized for more general metric spaces. For a further development see Nandakumar [14], Galatolo et al. [11], and Gacs et al. [5].

Let $f \in L^1$ be computable and assume that $\sup_\omega |f(\omega)| < \infty$; By $\|f\|$ denote the norm in L^1 (or in L^2). Let P be a computable measure and T a computable ergodic transformation preserving the measure P.

Define the sequence of ergodic averages $A_n^f, n = 1, 2, \ldots,$ where $A_n^f(\omega) = \frac{1}{n}\sum\limits_{k=0}^{n-1} f(T^k\omega)$.

Galatolo at al. [10] and Avigad et al. [1] showed that the ergodic averages $\{A_n^f\}$ effectively converge to a computable real number $c = \int f(\omega)dP$ almost surely as $n \to \infty$. Then the stability property of the ergodic theorem in the case where the transformation T is ergodic is a consequence of this result and Propositions 3 and 4. We present this result for completeness of exposition.

Proposition 6 *Let T be a computable measure preserving ergodic transformation. Then the sequence of ergodic averages $\left\{A_n^f\right\}$ effectively converges almost surely as $n \to \infty$.*

Proof We suppose without loss of generality that $\int f dP = 0$.[9] The sequence $\|A_n^f\|$ is computable and converges to 0 by the ergodic theorems.

The maximal ergodic theorem says that $P\{\omega : \sup\limits_n |A_n^f(\omega)| > \delta\} \le \frac{1}{\delta}\|f\|$ for any ergodic transformation T preserving the measure P.

Given $\epsilon, \delta > 0$, compute a $p = p(\delta\epsilon)$ such that $\|A_p^f\| \le \delta\epsilon/2$. By the maximal ergodic theorem for $g = A_p^f$ we have $P\{\omega : \sup\limits_n |A_n^g(\omega)| > \delta/2\} \le \frac{2}{\delta}\|A_p^f\| \le \epsilon$.

Now we check that A_n^g is not too far from A_n^f. Expanding A_n^g, one can check that

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

This implies that $\sup\limits_\omega |A_n^g(\omega) - A_n^f(\omega)| \le \frac{2}{np}\sum\limits_{k=1}^{p-1}(p-k)\sup\limits_\omega |f(\omega)| = \frac{p-1}{n}\sup\limits_\omega |f(\omega)| \le \delta/2$ for all $n \ge m(\delta,\epsilon) = 2(p(\delta\epsilon)-1)\sup\limits_\omega |f(\omega)|/\delta$. If $|A_n^f(\omega)| > \delta$ for some $n \ge m(\delta,\epsilon)$ then $|A_n^g(\omega)| > \delta/2$. Hence, $P\{\omega : \sup\limits_{n \ge m(\delta,\epsilon)} |A_n^f(\omega)| > \delta\} \le \epsilon$. The proposition is proved. □

Propositions 3, 4, and 6 imply a stable version of the ergodic theorem for the case where the transformation T is ergodic and $P = B_{1/2}$.

Theorem 1 *Let f be a computable observable and T a computable ergodic transformation preserving the uniform measure $B_{1/2}$. Then a computable unbounded nondecreasing function $\sigma(n)$ exists such that for any infinite sequence ω the condition $Km(\omega^n) \ge n - \sigma(n) - O(1)$ implies that the limit $\lim\limits_{n\to\infty}\frac{1}{n}\sum\limits_{k=0}^{n-1} f(T^k\omega)$ exists.*

Propositions 3, 4, and 6 imply also that in the case where the transformation T is ergodic, the Birkhof ergodic theorem holds for any Schnorr random sequence. Moreover, a total Solovay test $\mathcal{T}$ exists such that if an infinite sequence ω passes $\mathcal{T}$, then $A_n^f(\omega)$ converges as $n \to \infty$. This result

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

should probably be attributed to Galatolo et al. [11] (see also Franklin and Towsner [6]).

4 Instability in Ergodic Theory

The phenomenon of instability occurs in ergodic theory. In this section we present a property of uniform instability of the ergodic theorem and absolute instability for a non-ergodic measure-preserving transformation.

4.1 Instability Results

The degree of stability $\sigma(n)$ from Theorem 1 may depend on the observable f and transformation T. Theorem 2 below shows that there is no uniform degree of stability $\sigma(n)$ for the ergodic theorem.

The phenomenon of instability of the ergodic theorem was first discovered by V'yugin [24]. Compared with a "symbolic dynamics type" result from [24], this result is "measure free"; it is formulated in terms of transformations and the Kolmogorov complexity.

Theorem 2 *Let $\sigma(n)$ be a nondecreasing unbounded computable function. Then there exist a computable ergodic measure-preserving transformation T and an infinite sequence $\omega \in \Omega$ such that the inequality $Km(\omega^n) \geq n - \sigma(n)$ holds for all n and the limit*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} f(T^i \omega) \tag{6}$$

does not exist for some computable indicator function f.

Gacs et al. [5] showed that for every infinite sequence ω which is not Schnorr random, a measure preserving ergodic transformation T exists such that the limit (6) does not exist (i.e., the ergodic theorem does not hold). This result and Theorem 1 show that Theorem 2 is equivalent to the following statement: for every nondecreasing unbounded computable function $\sigma(n)$ there is some infinite sequence ω such that $Km(\omega^n) \geq n - \sigma(n)$ but ω is not Schnorr random.

In the next theorem a uniform (with respect to $\sigma(n)$) result is presented. In this case, we will lose the ergodic property of a transformation T.

Theorem 3 *A computable measure-preserving transformation T can be constructed such that for any nondecreasing unbounded computable function $\sigma(n)$ an infinite sequence ω exists such that $Km(\omega^n) \geq n - \sigma(n)$ holds for all n and the limit (6) does not exist for some computable indicator function f.*

A constructive version of the ergodic theorem by V'yugin [21] shows that this sequence ω is not Martin-Löf random. A closely related result was obtained by Franklin and Towsner [6]. Using the cutting and stacking method, they showed that for every infinite sequence ω which is not Martin-Löf random, a measure preserving transformation T exists such that the limit (6) does not exist.

A construction of the transformation T is given in Section 4.3; the proof of Theorem 2 is given in Section 4.4. In Section 4.2 we consider the main technical concept, the method of cutting and stacking.

4.2 Method of Cutting and Stacking

In this section we consider the main notions and properties of cutting and stacking method (see Shields [17, 18]).

A column is a sequence $E = (L_1, \ldots, L_h)$ of pairwise disjoint intervals of the unit interval $[0, 1]$ of equal width. We refer to L_1 as to the base and to L_h as to the top of the column; $\hat{E} = \bigcup_{i=1}^{h} L_i$ is the support of the column; $w(E) = \lambda(L_1)$ is the width of the column; h is the height of the column; $\lambda(\hat{E}) = \lambda(\bigcup_{i=1}^{h} L_i)$ is the measure of the column, where λ is the uniform measure on $[0, 1]$.

Any column defines a transformation T which linearly transforms L_j to L_{j+1}, namely, $T(x) = x + c$ for all $x \in L_j$, where c is the corresponding constant and $1 \leq j \leq h$. This transformation T is not defined outside all intervals of the column and at all points of the top interval L_h of this column.

Denote $T^0\omega = \omega$, $T^{i+1}\omega = T(T^i\omega)$. For any $1 \leq j < h$, an arbitrary point $\omega \in L_j$ generates a finite trajectory $\omega, T\omega, \ldots, T^{h-j}\omega$.

By a partition of the unit interval $[0, 1]$ we mean any pair $\pi = (\pi_0, \pi_1)$ of disjoint subsets of this interval such that $\pi_0 \cup \pi_1 = [0, 1]$. In what follows we suppose that some partition $\pi = (\pi_0, \pi_1)$ is given.

A partition $\pi = (\pi_0, \pi_1)$ is compatible with a column E if for each j there exists a number 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 the 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 Υ is a finite collection of columns with disjoint supports. The width of the gadget $w(\Upsilon)$ is the sum of the widths of its columns. The support of the gadget Υ is the union $\hat{\Upsilon}$ of the supports of all its columns. We suppose that the partition $\pi = (\pi_0, \pi_1)$ is compatible with each column of the gadget Υ.

The union of gadgets Υ_i with disjoint supports is the gadget $\Upsilon = \bigcup \Upsilon_i$ whose columns are the columns of all the Υ_i. A transformation $T = T(\Upsilon)$ is associated with a gadget Υ if it is the union of transformations defined on all columns of Υ.

Any point of the support $\hat{\Upsilon}$ of a gadget Υ generates a finite trajectory. By the Υ-name of this trajectory we mean its E-name, where E is the column of Υ to which this trajectory corresponds. A gadget Υ extends a column Λ if the support of Υ extends the support of Λ and the transformation $T(\Upsilon)$ extends the transformation $T(\Lambda)$.

Since all points of the interval L_j of the column generate trajectories with the same names, we refer to the name of any such trajectory as to the name generated by the interval L_j.

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

$$\left(\frac{w(E_1)}{w(\Upsilon)}, \ldots, \frac{w(E_n)}{w(\Upsilon)}\right). \tag{7}$$

A gadget Υ is a copy of a gadget Λ if they have the same distributions and the corresponding columns have the same partition names.

A gadget Υ 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 (7) 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 Υ of width γ_m. The action of the gadget transformation T is not affected by the copying operation.

Another operation is stacking of gadgets onto gadgets. First we consider stacking of columns onto columns and 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. A 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 $L_j = L_{2,j-h(E_1)}$ for all $h(E_1) < j \leq h(E_1) + h(E_2)$.

Let a gadget Υ and a column E have the same width, and assume that 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 Υ such that $w(E_i) = w(U_i)$, where U_i is the ith column of the gadget Υ. Stack U_i on the top of E_i to get a new column $E_i * U_i$. A new gadget consists of the columns $(E_i * U_i)$.

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

The M-fold independent cutting and stacking of a single gadget Υ is defined by cutting Υ into M copies $\Upsilon_i, i = 1, \ldots, M$, of equal width and successive independent cutting and stacking to obtain $\Upsilon^{*(M)} = \Upsilon_1 * \cdots * \Upsilon_M$. A sequence of gadgets $\{\Upsilon_m\}$ is complete if

- $\lim_{m\to\infty} w(\Upsilon_m) = 0$;
- $\lim_{m\to\infty} \lambda(\hat{\Upsilon}_m) = 1$;
- Υ_{m+1} extends Υ_m for all m.

Any complete sequence of gadgets $\{\Upsilon_s\}$ determines a transformation $T = T\{\Upsilon_s\}$ which is defined almost surely.

By the definition, T preserves the measure λ. Shields [17] gives the following sufficient conditions for a process T to be ergodic. Let a gadget Υ be constructed by cutting and stacking from a gadget Λ. Let E be a column from Υ and D a column from Λ. Then the set $\hat{E} \bigcap \hat{D}$ is the union of subcolumns from D of width $w(E)$ which were used to construct E.

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

$$\sum_{D\in\Lambda}\sum_{E\in\Upsilon} |\lambda(\hat{E} \bigcap \hat{D}) - \lambda(\hat{E})\lambda(\hat{D})| < \epsilon. \tag{8}$$

We will use the following two lemmas.

Lemma 1 *([17], Corollary 1), ([18], Theorem A.1). Let $\{\Upsilon_n\}$ be a complete sequence of gadgets and assume that 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 an ergodic process.*

Lemma 2 *([18], Lemma 2.2). For any $\epsilon > 0$ and any gadget Υ there is a number M such that for each $m \ge M$ the gadget Υ is $(1-\epsilon)$-well-distributed in the gadget $\Upsilon^{*(m)}$ constructed from Υ by m-fold independent cutting and stacking.*

We refer the reader to Shields [17] for the proof.

Several examples of stationary and ergodic transformation constructed using the cutting and stacking method are given in Shields [17, 18].

4.3 Construction

Let $r > 0$ be a sufficiently small rational number. Define a partition $\pi = (\pi_0, \pi_1)$ of the unit interval $[0, 1]$, where $\pi_0 = [0, 0.5) \bigcup (0.5 + r, 1)$ and $\pi_1 = [0.5, 0.5 + r]$.

Let $\sigma(n)$ be a computable unbounded nondecreasing function. A computable sequence of positive integer numbers exists such that $0 < h_{-2} < h_{-1} < h_0 < h_1 < \dots$ and $\sigma(h_{i-1}) - \sigma(h_{i-2}) > i - \log r + 8$ for all $i = 0, 1, \dots$.

The gadgets Δ_s, Π_s, where $s = 0, 1, \dots$, will be defined by mathematical induction on the number of steps. The gadget Δ_0 is defined by cutting the interval $[0.5 - r, 0.5 + r]$ into equal parts and stacking them. Let Π_0 be a gadget defined by cutting the intervals $[0, 0.5 - r)$ and $(0.5 + r, 1]$ into equal subintervals and stacking them. The purpose of this definition is to construct initial gadgets of width $\le 2^{-h_0}$ with supports satisfying $\lambda(\hat{\Delta}_0) = 2r$ and $\lambda(\hat{\Pi}_0) = 1 - 2r$.

The sequence of gadgets $\{\Delta_s\}$, $s = 0, 1, \dots$, will define an approximation of the uniform Bernoulli measure concentrated on the names of their trajectories. The sequence of gadgets $\{\Pi_s\}$, $s = 0, 1, \dots$, will define a measure with sufficiently small entropy. The gadget Π_{s-1} will be extended at each step of the construction by a half of the gadget Δ_{s-1}. After that, the independent cutting and stacking process will be applied to this extended gadget to obtain the gadget Π_s. This process eventually defines infinite trajectories starting from points of [0, 1]. The sequence of gadgets $\{\Pi_s\}$, $s = 0, 1, \dots$, will be complete and will define a transformation T. Lemmas 1 and 2 from Section 4.2 ensure the transformation T to be ergodic.

Construction Let at step $s-1$ ($s > 0$) gadgets Δ_{s-1} and Π_{s-1} be defined. Cut the gadget Δ_{s-1} into two copies Δ' and Δ'' of equal width (i.e., cut each column into two subcolumns of equal width) and join $\Pi_{s-1} \bigcup \Delta''$ in to one gadget. Find a sufficiently large number R_s and do R_s-fold independent cutting and stacking of the gadget $\Pi_{s-1} \bigcup \Delta''$ and also of the gadget Δ' to obtain new gadgets Π_s and Δ_s of width $\le 2^{-h_s}$ such that the gadget $\Pi_{s-1} \bigcup \Delta''$ is $(1 - 1/s)$–well–distributed in the gadget Π_s. The needed number R_s exists by Lemma 2 (Section 4.2).

By the construction, the endpoints of all subintervals of [0, 1] used in this construction are rational numbers, and so the construction is algorithmically effective.

Properties of the construction Define a transformation $T = T\{\Pi_s\}$. Since the sequence of the gadgets $\{\Pi_s\}$ is complete (i.e. $\lambda(\hat{\Pi}_s) \to 1$ and $w(\Pi_s) \to 0$ as $s \to \infty$), T is defined almost surely.

The transformation T is ergodic by Lemma 1, since the sequence of gadgets Π_s is complete. Furthermore, the gadget $\Pi_{s-1} \bigcup \Delta''$ and the gadget Π_{s-1} are $(1 - 1/s)$-well distributed in Π_s for any s. By the construction, $\lambda(\hat{\Delta}_i) = 2^{-i+1}r$ and $\lambda(\hat{\Pi}_i) = 1 - 2^{-i+1}r$ for all $i = 0, 1, \dots$.

We need to interpret the transformation T as a transformation of infinite binary sequences. To do this, we identify real numbers from [0, 1] with their infinite binary representations. This correspondence is one-to-one except for the countable set of infinite sequences corresponding to dyadic rational numbers: for example, 0.0111... = 0.10000.... Such sequences can be ignored, since their set is countable.

From the point of view of this interpretation, the Bernoulli measure $B_{1/2}$ and the uniform measure λ are identical, and the transformation T constructed above preserves the uniform Bernoulli measure and is defined almost surely.

4.4 Proof of Theorem 2

For technical convenience, in the proof of Theorem 2 we replace the deficiency of randomness $dm(x)$ with a nonnegative supermartingale (see Schiryaev [19]). A function $M : \{0, 1\}^* \to \mathcal{R}$ is called a supermartingale if $M(\Lambda) \le 1$ and $M(x) \ge \frac{1}{2}(M(x0) + M(x1))$ for all x. Also, we require $M(x) \ge 0$ for all x. A more general property holds: $M(x) \ge \sum\limits_{y \in B} M(xy)2^{-l(y)}$ for any prefix-free set B.

Recall the proof of that the deficiency of randomness is bounded by the logarithm of some supermartingale: $dm(x) \le \log M(x)$ for all x.

Let ψ be the optimal function which defines the monotone complexity $Km(x)$. Define $Q(x) = B_{1/2}(\bigcup\{\Gamma_p : x \subseteq \psi(p)\})$. It is easy to verify that $Q(\Lambda) \le 1$ and $Q(x) \ge Q(x0) + Q(x1)$ for all x. Then the function $M(x) = 2^{l(x)}Q(x)$ is a supermartingale and $M(x) \ge 2^{l(x)-Km(x)}$ for all x.

Denote $d(x) = \log M(x)$. Using the following lemma, we will construct an infinite binary sequence such that the randomness deficiency of its initial segments grows arbitrarily slowly.

Lemma 3 *For any set of binary strings A and for any string x, a string $y \in A$ exists such that $d(xy^n) \le d(x) - \log B_{1/2}(\tilde{A}) + 1$ for all n such that $1 \le n \le l(y)$, where $\tilde{A} = \bigcup\{\Gamma_y : y \in A\}$.*

Proof Define $A_1 = \left\{y \in A : \exists j(1 \le j \le l(y) \& M(xy^j) > 2M(x)/B_{1/2}(\tilde{A}))\right\}$. For any $y \in A_1$, let y^p be the initial segment of y of the minimal length such that $M(xy^p) > 2M(x)/B_{1/2}(\tilde{A})$. The set $\{y^p : y \in A_1\}$ is prefix-free. Then we have

$$1 \ge \sum_{y\in A_1} \frac{M(xy^p)}{M(x)} 2^{-l(y^p)} \ge \frac{2}{B_{1/2}(\tilde{A})} \sum_{y\in A_1} 2^{-l(y^p)} \ge \frac{2B_{1/2}(\tilde{A}_1)}{B_{1/2}(\tilde{A})}.$$

From this we obtain $B_{1/2}(\tilde{A}_1) \le \frac{1}{2}B_{1/2}(\tilde{A})$ and $B_{1/2}(\tilde{A} \setminus \tilde{A}_1) > \frac{1}{2}B_{1/2}(\tilde{A})$.

For any $y \in A \setminus A_1$, we have $M(xy^j) \le 2M(x)/B_{1/2}(\tilde{A})$ for all x such that $l(x) \le j \le (y)$. □

We will use the construction of Section 4.3 to show that an infinite binary sequence ω exists such that $d(\omega^n) \le \sigma(n)$ for all n and the limit (6) does not exist for the name $\chi(\omega)\chi(T\omega)\chi(T^2\omega)\dots$ of its trajectory, where $\chi(\omega) = i$ if $\omega \in \pi_i$, $i = 0, 1$. More precisely, we prove that

$$\limsup_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} \chi(T^i\omega) \ge 1/16, \tag{9}$$

$$\liminf_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} \chi(T^i\omega) \le 2r, \tag{10}$$

where r is sufficiently small.

By induction on the number of steps we will define the sequence ω as a union of an increasing sequence of initial segments

$$\omega(0) \subset \dots \subset \omega(k) \subset \dots. \tag{11}$$

We will also define an auxiliary sequence of integer numbers $s(-1) = s(0) = 0 < s(1) < \dots$.

Using Lemma 3, define $\omega(0)$ such that $d(\omega(0)^j) \le 2$ for all $j \le l(\omega(0))$.

Let us consider dyadic intervals of the form $[m2^{-n}, (m+1)2^{-n})$, where $0 \le m < 2^n$. Any such interval corresponds to some finite sequence $x = x_1 \dots x_n$ and the corresponding binary interval $\Gamma_x = \{\omega \in \Omega : x \subset \omega\}$ in Ω.

Induction hypothesis Suppose that a binary sequence $\omega(0) \subset \cdots \subset \omega(k-1)$ of strings and a sequence of integer numbers $s(-1) = s(0) = 0 < s(1) < \cdots < s(k-1)$ are already defined.

Suppose also that the dyadic interval corresponding to the string $\omega(k-1)$ is a subset of one of the intervals of the gadget $\Pi_{s(k-1)}$. By the construction, $w(\Pi_{s(k-1)}) \le 2^{-h_{s(k-1)}}$. Then $l(\omega(k-1)) \ge h_{s(k-1)}$.

We also suppose that $d(\omega(k-1)) \le \sigma(h_{s(k-2)}) - 4$ if k is odd and $d(\omega(k-1)) \le \sigma(h_{s(k-2)})$ if k is even.

Consider an odd k. Denote $a = \omega(k-1)$ and let I_a be the dyadic interval corresponding to a. Any point of I_a generates the Π_s-trajectory. By the ergodic theorem, for a.e. points of I_a, the frequency of visiting the element π_1 of the partition converges to r as $s \to \infty$.

Let s be sufficiently large such that $s > s(k-1)$ and the total measure of all points of I_a generating Π_s-trajectories with frequency $\le 2r$ of visiting the element π_1 is at least $(1/2)2^{-l(a)}$.

The intersection of intervals from Π_s with I_a can be represented as a union of pairwise disjoint intervals $[r_1, r_2]$. It is easy to see that any such interval $[r_1, r_2]$ contains a dyadic subinterval of length at least $\frac{1}{4}(r_2 - r_1)$ corresponding to a binary string b. Let C_a be a set of such strings b. The Bernoulli measure of C_a is at least $(1/8)2^{-l(a)}$.

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

By Lemma 3 a sequence $b \in C_a$ exists such that $d(b^j) \le d(a) + 4$ for each $l(a) \le j \le l(b)$. Define $\omega(k) = b$. By the induction hypothesis, $d(a) \le \sigma(h_{s(k-2)}) - 4$ and $l(a) \ge h_{s(k-1)}$. Then $d(b^j) \le \sigma(h_{s(k-2)}) < \sigma(h_{s(k-1)}) \le \sigma(l(a)) \le \sigma(j)$ for all $l(a) \le j \le l(b)$. Also, since $w(\Pi_s) \le 2^{-h_s}$, we have $l(b) \ge h_{s(k)}$. Therefore, the induction hypothesis and condition (10) are valid for the next step of induction.

Let k be even. Put $b = \omega(k-1)$ and $s(k) = s(k-1) + 1$. Let $s = s(k)$ and Π_s be the gadget generated by the R_s-fold independent cutting and stacking of the gadget $\Pi_{s-1} \bigcup \Delta''$.

Let $E = (L_1, \ldots, L_k)$ be a column of the gadget Δ'' defined at step s and let $E^* = (L_1, \ldots, L_{\lceil k/2 \rceil})$ be its lower half.

Divide all intervals of the gadget Δ'' into two equal parts: an upper half and lower half. Any interval of the lower half of Δ'' generates a trajectory of length $\ge M/2$, where M is the height of the gadget Δ''. The uniform measure of the support of the lower half is $\ge \frac{1}{2}\lambda(\hat{\Delta}'')$. By the construction, for all sufficiently large s, the measure of all points from this set, whose Δ''-trajectories have length $\ge M/2$ and frequency of ones $\ge 1/4$, is at least $\frac{1}{4}\lambda(\hat{\Delta}'')$.

By the construction,

$$\gamma = \frac{\lambda\left(\hat{\Delta}''\right)}{\lambda\left(\hat{\Pi}_{s-1}\right)} = \frac{\lambda\left(\hat{\Delta}_{s-1}\right)}{2\lambda\left(\hat{\Pi}_{s-1}\right)} = \frac{2^{-s+1}r}{1 - 2^{-s+2}r} >$$

$$> 2^{-s+1}r \ge 2^{-(\sigma(h_{s-1}) - \sigma(h_{s-2}) - 9)}. \tag{12}$$

Let I_b be the dyadic interval corresponding to the string b. By the construction, I_b is a subset of some interval of Π_{s-1}. Consider a subset of I_b such that Π_s-trajectories starting from points of these intervals pass through the corresponding upper subcolumns of the gadget Δ'' and have frequencies of ones at least $1/4$ in substrings defined by Δ''. Since the gadgets Π_{s-1} and Δ'' have the same heights M, some initial segment of the trajectory starting from such a point has length at most $2M$ and its name has at least $M/4$ ones. Hence, the frequency of ones in the name of any such initial segment is at least $\frac{1}{8}$. The total measure of this subset of I_b is at least $\frac{\gamma}{4}2^{-l(b)}$. The intersection of this subset with intervals of the gadget Π_s can be represented as a union of pairwise disjoint intervals $[r_1, r_2]$. Any such interval contains a dyadic subinterval of length at least $\frac{1}{4}(r_2 - r_1)$ corresponding to a binary string extending b. The measure of these dyadic subintervals is at least $\frac{\gamma}{16}2^{-l(b)}$. The set D_b of the corresponding binary strings has the same Bernoulli measure.

By Lemma 3 some $c \in D_b$ exists such that $d(c^j) \le d(b) + 1 - \log\frac{\gamma}{16} \le d(b) + (\sigma(h_{s-1}) - \sigma(h_{s-2}) - 9) + 5 \le \sigma(h_{s-1}) - 4$ for all j such that $l(b) \le j \le l(c)$. Here we have used the induction hypothesis, the inequality $d(b) \le \sigma(h_{s(k-2)}) \le \sigma(h_{s-2})$, and inequality (12). Besides, $l(b) \ge h_{s-1}$. Therefore, $d(c^j) < \sigma(h_{s-1}) \le \sigma(l(b)) \le \sigma(j)$ for all j such that $l(b) \le j \le l(c)$. Define $\omega(k) = c$.

It is easy to see that the induction hypothesis is valid for this k.

The infinite sequence ω is defined by a sequence of its initial segments (11). We have proved that $d(\omega^j) \le \sigma(j)$ for all j.

By the construction, there are infinitely many initial segments of the trajectory of the sequence ω with frequency of ones $\ge 1/8$ in their names. Also, there are infinitely many initial segments of this trajectory with frequency of ones $\le 2r$. Hence, condition (9) holds.

The proof of Theorem 3 is more complicated. Consider a sequence of pairwise disjoint subintervals J_i of the unit interval $[0, 1]$ of lengths 2^{-i}, $i = 1, 2, \dots$, and a uniform computable sequence $\sigma_i(n)$ of all partial recursive functions (candidates for a degree of instability). For any i, we apply the construction of Section 4.3 to the subinterval J_i and to a function $\sigma_i(n)$ in order to define a computable ergodic measure-preserving transformation T_i on J_i for each i. The needed transformation is defined as a union of all these transformations T_i. We omit details of this construction.

4.5 Instability of Universal Compression Schemes

Note that an infinite sequence ω is Martin-Löf random with respect to a computable measure P if and only if $Km(\omega^n) = -\log P(\omega^n) + O(1)$ as $n \to \infty$.

A recent result of Hochman [7] implies an algorithmic version of the Shannon–McMillan–Breiman theorem for Martin-Löf random sequences: for any computable stationary ergodic measure P with entropy H, $Km(\omega^n) \ge -\log P(\omega^n) - O(1)$ as $n \to \infty$ implies

$$\lim_{n\to\infty} \frac{Km(\omega^n)}{n} = \lim_{n\to\infty} \frac{-\log P(\omega^n)}{n} = H. \quad (13)$$

Clearly, the same property holds for plain and prefix Kolmogorov complexities and for a sequence ω Martin-Löf random with respect to the uniform measure.

The construction given in Section 4.3 can be applied to show the instability of relation (13): for any computable function $\sigma(n)$ as in Theorem 2 and for any sufficiently small $\epsilon > 0$ a computable stationary ergodic measure P with entropy $0 < H \le \epsilon$ and an infinite binary sequence ω exist such that $Km(\omega^n) \ge -\log P(\omega^n) - \sigma(n)$ for all n and

$$\limsup_{n\to\infty} \frac{Km(\omega^n)}{n} \ge \frac{1}{4},$$
$$\liminf_{n\to\infty} \frac{Km(\omega^n)}{n} \le \epsilon.$$

By a prefix-free code we mean a computable sequence of one-to-one functions $\{\phi_n\}$ from $\{0,1\}^n$ to a prefix-free set of finite sequences. In this case a decoding method $\hat{\phi}_n$ also exists such that $\hat{\phi}_n(\phi_n(\alpha)) = \alpha$ for each α of length n.

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

$$\lim_{n\to\infty} \frac{l(\phi_n(\omega^n))}{n} = H \text{ almost surely.}$$

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

We have also an instability property for any universal coding scheme: for any computable function $\sigma(n)$ as in Theorem 2 and for any sufficiently small $\epsilon > 0$ a computable stationary ergodic measure P with entropy $0 < H \le \epsilon$ exists such that for each universal code $\{\phi_n\}$ an infinite binary sequence ω exists such that $Km(\omega^n) \ge -\log P(\omega^n) - \sigma(n)$ for all n and

$$\limsup_{n\to\infty} \frac{l(\phi_n(\omega^n))}{n} \ge \frac{1}{4},$$
$$\liminf_{n\to\infty} \frac{l(\phi_n(\omega^n))}{n} \le \epsilon.$$

The proof of these statements is based on the construction of Section 4.3. For further details we refer the reader to V'yugin [25].

Acknowledgments The author is grateful to an anonymous referee for numerous corrections and suggestions on improving the content of this paper.

References

1. Avigad, J., Gerhardy, P., Towsner, H.: Local stability of ergodic averages. Trans. Am. Math. Soc. **362**(1), 261–288 (2010)
2. Bienvenu, L., Merkle, W.: Reconciling data compression and Kolmogorov complexity. Lecture Notes in Computer Science **4596**, 643–654 (2007)
3. Bishop, E.: Foundation of Constructive Analysis. McGraw-Hill, New York (1967)
4. Downey, R.G., Griffiths, E.G.: On Schnorr randomness. J. Symb. Log. **69**(2), 533–554 (2004)

5. Gacs, P., Hoyrup, M., Rojas, C.: Randomness on computable probability spaces-a dynamical point of view. Theory of Computing Systems **48**(3), 465–485 (2011)
6. Franklin, J.N.Y., Towsner, H.: Randomness and non-ergodic systems (2012). arXiv:1206.2682v1 [math.LO]
7. Hochman, M.: Upcrossing inequalities for stationary sequences and applications to entropy and complexity. Ann. Probab. **37**(6), 2135–2149 (2009)
8. Hoeffding, W.: Probability inequalities for sums of bounded random variables. J. Am. Stat. Assoc. **58**(301), 13–30 (1963)
9. Hoyrup, M., Rojas, C.: Computability of probability measures and Martin-Löf randomness over metric spaces. Inf. Comput. **207**(7), 830–847 (2009)
10. Galatolo, S., Hoyrup, M., Rojas, C.: Computing the speed of convergence of ergodic averages and pseudorandom points in computable dynamical systems. In: Proceedings Seventh International Conference on Computability and Complexity in Analysis, CCA 2010, Zhenjiang, China, 21-25th June 2010, pp. 7–18 (2010). doi:10.4204/EPTCS.24.6
11. Galatolo, S., Hoyrup, M., Rojas, C.: Effective symbolic dynamics, random points, statistical behavior, complexity and entropy. Inf. Comput. **208**(1), 23–41 (2010)
12. Krengel, U.: Ergodic Theorems, Berlin, New York: de Cruyter (1984)
13. Li, M., Vitányi, P.: An Introduction to Kolmogorov Complexity and Its Applications. Springer-Verlag, New York (1997)
14. Nandakumar, S.: An effective ergodic theorem and some applications. In: Proceeding STOC'08, pp. 39–44 (2008)
15. Ryabko, B.: Twice universal coding. Probl. Inform. Transm. **20**, 173–178
16. Schnorr, C.P.: A unified approach to the definition of random sequences. Mathematical Systems Theory **5**, 246–258 (1971)
17. Shields, P.C.: Cutting and stacking: a method for constructing stationary processes. IEEE Trans. Inform. Theory **37**(6), 1605–1617 (1991)
18. Shields, P.C.: Two divergence-rate counterexamples. J. Theoret. Probability **6**, 521–545 (1993)
19. Shiryaev, A.N.: Probability, Berlin, Springer (1980)
20. Vovk, V.G.: The law of the iterated logarithm for random Kolmogorov, or chaotic sequences. SIAM Theory Probab. Applic. **32**, 413–425 (1987)
21. V'yugin, V.V.: Effective Convergence in Probability and an Ergodic Theorem for Individual Random Sequences. Theory Probab. Appl. **42**(1), 39–50 (1998)
22. V'yugin, V.V.: On the longest head-run in an individual random sequence. Theory Probab. Appl. **42**(3), 541–546 (1998)
23. V'yugin, V.V.: Ergodic theorems for individual random sequences. Theor. Comput. Sci. **207**(4), 343–361 (1998)
24. V'yugin, V.V.: Non-robustness property of the individual ergodic theorem. Probl. Inf. Transm. **37**(2), 27–39 (2001)
25. V'yugin, V.V.: Problems of robustness for universal coding schemes. Probl. Inform. Transm. **39**(1), 32–46 (2003)
26. V'yugin, V.V.: On Instability of the Ergodic Limit Theorems with Respect to Small Violations of Algorithmic Randomness. In: Proceedings of the IEEE International Symposium on Information Theory (ISIT 2011), St. Petersburg, Russia, August (2011) ISBN 978–1–4577–0594–6 1614–1618 (2011)