Ornstein Isomorphism and Algorithmic Randomness

Mrinalkanti Ghosh *, Satyadev Nandakumar †, and Atanu Pal ‡

1Toyota Technological Institute at Chicago, Chicago IL 60637, USA
2Department of Computer Science and Engineering, Indian Institute of Technology Kanpur, Kanpur, UP, 208016, India
3Strand Genomics, Bangalore, India

April 4, 2014

Abstract

In 1970, Donald Ornstein proved a landmark result in dynamical systems, viz., two Bernoulli systems with the same entropy are isomorphic except for a measure 0 set [19]. Keane and Smorodinsky [14] gave a finitary proof of this result. They also indicated how one can generalize the result to mixing Markov Shifts in [12]. We adapt the construction given in [14] to show that if two computable mixing Markov systems have the same entropy, then there is a layerwise lower semicomputable isomorphism defined on all Martin-Löf random points in the system. Since the set of Martin-Löf random points forms a measure 1 set, it implies the classical result for such systems.

This result uses several recent developments in computable analysis and algorithmic randomness. Following the work by Braverman [3], Nandakumar [18], and Hoyrup and Rojas [9] introduced discontinuous functions into the study of algorithmic randomness. We utilize Hoyrup and Rojas' elegant notion of layerwise computability to produce the test of randomness in our result. Further, we use the recent result of the effective Shannon-McMillan-Breiman theorem, independently established by Hochman [7] and Hoyrup [11] to prove the properties of our construction.

We show that the result cannot be improved to include all points in the systems - only trivial computable isomorphisms exist between systems with the same entropy.

*mkghosh@ttic.edu
†satyadev@cse.iitk.ac.in
‡palatanu@cse.iitk.ac.in
1 Introduction

In the Kolmogorov program for algorithmic randomness, Martin-Löf established that there is a smallest constructive measure 1 set, whose objects are the set of individual random objects. Every effectively computable probabilistic law, i.e. law which holds with probability 1, specifies a “majority rule”. Thus it is reasonable to ask if every such law is satisfied by every individual random object. This will *a fortiori* imply the classical theorem, since the set of random objects has probability 1. The effective versions have more intuitive content, since they show that if any object fails the particular law, then there is an algorithm which can “bet” and win unbounded amounts of money on it.

Indeed, very general theorems like the Strong Law of Large Numbers [29], the Law of Iterated Logarithm [30], and Birkhoff’s Ergodic Theorem [31], [18], [6], [1] have been effectivized. Prior to the work of Braverman [3], only continuous functions were considered. Following the work of Braverman, Nandakumar [18] and Hoyrup and Rojas [10] have considerably broadened the class of functions to deal with discontinuities. Recently, Hochman [7] and Hoyrup [11] independently resolved the long-standing open problem of the effectivization of the Shannon-McMillan-Breiman theorem.

In a recent line of work, Gács [4], and Gács, Hoyrup and Rojas [5], [10], [9] have extended the field of study of randomness to fairly general spaces other than the finite alphabet spaces which have traditionally formed the subject of algorithmic randomness. This also enables us to study the *relationships between* the random objects of different probability spaces. In this paper, we utilize this theory to study the isomorphism between effective dynamical systems. We prove an effective version of the celebrated Ornstein Isomorphism Theorem [19], by adapting the finitary proof of Keane and Smorodinsky [14].

Consider two dynamical systems \((X, \mathcal{B}, \mu, T)\) and \((Y, \mathcal{C}, \nu, S)\) where \(X, Y\) are the sample space, \(\mathcal{B}, \mathcal{C}\) the \(\sigma\)-algebras, \(\mu, \nu\) the probabilities, and \(T\) and \(S\) the measure-preserving transformation on \(X\) and \(Y\) respectively. A map \(\phi: X \to Y\) is a factor map if \(\phi T(x) = S \phi(x)\) for almost every \(x \in X\). If \(\phi\) is invertible then we say that \(X\) and \(Y\) are isomorphic. Isomorphism helps us to categorize dynamical systems into classes of systems which are essentially “encodings” of another system.

Kolmogorov and Sinai [16], [28] introduced the notion of the *entropy* of a dynamical system as an invariant of an isomorphism. They showed that if two systems are isomorphic to each other, then they have the same *Kolmogorov-Sinai entropy*. Ornstein and Weiss [21] show that this was a crucial insight - in a very broad sense, the Kolmogorov-Sinai entropy is the only invariant of the isomorphism. The Kolmogorov-Sinai theorem brought a fresh perspective to the study of dynamical systems. Formally, it justifies viewing purely deterministic dynamical systems as having positive entropy [23] - thus some deterministic systems can be viewed as “random”.

The converse of the result, *viz.* that systems with the same Kolmogorov-Sinai entropy are

---

1 definitions in Section 4.1
isomorphic to each other, does not hold in general (see Billingsley [2]). However, Ornstein showed in a celebrated result, that if we restrict the systems to the broad class of “Bernoulli systems”, then equal entropy systems are isomorphic to each other. Ornstein generalized this result to hold on the class of “finitely determined systems”. Numerous examples of deterministic dynamical systems are isomorphic to the Bernoulli system, which is intuitively the most random system possible. (For a recent survey, see Ornstein [20].)

However, the isomorphism Ornstein constructs is not continuous (it cannot be continuous in general [24]) and is not directly amenable to the theory of algorithmic randomness. In 1979, Keane and Smorodinsky gave a finitary version of Ornstein isomorphism theorem. A map is called finitary if it is continuous except on a measure 0 set. The concept involves viewing the underlying systems as both probability and topological spaces. We adapt this proof to establish our result.

Our main contributions are the following.

First, we show that there is a layerwise isomorphism defined on all Martin-Löf random points. 

**Theorem 1.** If two effective mixing Markov systems have the same Kolmogorov-Sinai entropy, then there is a “layerwise computable” isomorphism which is defined on all Martin-Löf random objects of both the systems.

Hoyrup and Rojas [9] have shown that layerwise computable functions can be used to characterize Martin-Löf randomness.

Further, in Section 6 we show that this cannot be improved substantially – if we insist on a computable transformation which is defined on all points, then we have no non-trivial isomorphism.

This work crucially employs the concept of layerwise computability, which affords us the luxury of ignoring uncomputability of a function on a large set of discontinuities. Our construction will diverge on many non-random points. (For example, if a computable point has only finitely many zeroes in its “encoding”, then our map is undefined at that point.) This is an important difference from the result of Keane and Smorodinsky (see Theorem 17 of [13]), where the points of divergence of the construction are immaterial. We show that for every Martin-Löf random object, the adapted Keane-Smorodinsky construction converges – in particular, in a layerwise computable manner. Consequently there is a pointwise isomorphism between the set of random objects in the two systems.

## 2 Assumptions and Notations

In this section we describe notations only relevant for the proof developed in section 4. In order to facilitate easy detection of parallel constructs and difference between our proof and that of [14], we closely follow notations of [24].

We are given two finite alphabet stationary mixing Markov systems \( \mathcal{A} \) and \( \mathcal{B} \) with the guarantee that they have same entropy. Let their alphabet sets be \( \Sigma_A \) and \( \Sigma_B \) respectively. Let \( m \) be the
maximum memory of the two Markov processes. Note that all the conditional probabilities are bounded away from 0 or 1.

Let us consider the sequence \( \langle \varepsilon_n \rangle_{n=1}^{\infty} \) defined by \( \varepsilon_n = \frac{1}{2^n} \) for any natural number \( r \).

We assume that the probabilities of the given systems are computable. To be precise, we assume that we have a Turing machine \( M_A \) for the system \( A \) (and \( M_B \) for \( B \)) so that given a string \( x \in \Sigma_A^* \) (correspondingly, \( x \in \Sigma_B^* \)) and a natural number \( n \), \( M_A(x,n) \ (M_B(x,n) \) for \( B \)) returns a rational number approximating the probability of a cylinder \( x \) within additive error \( \varepsilon_n \cdot P_A(x) \) (\( \varepsilon_n \cdot P_B(x) \) for \( B \)). We denote this approximation by \( P_A(x,n) \) and \( P_B(x,n) \) respectively. Similarly, we denote approximation of conditional probabilities as \( P_A(a|x,n) \) for \( a \in \Sigma_A, x \in \Sigma_A^* \) and similarly and \( P_B(b|y,n) \) for \( b \in \Sigma_B, y \in \Sigma_B^* \). Since the dynamical systems are assumed to be stationary, the position of the cylinder does not matter.

Given a finite-dimensional probability vector \( P \), we denote its entropy as \( H(P) \). From the above assumption, we can infer that the entropy of the systems is computable, i.e., we have a Turing machine \( M \), which on input \( n \), gives a \( \varepsilon_n \) approximation of the entropy \( H \).

## 3 Overview of the construction

First, we reduce the problem of construction of isomorphism between two mixing Markov systems of equal entropy to one where two systems have a common probability weight. We call this the Marker Lemma, analogous to Keane and Smorodinsky. Our construction differs in that all our systems are mixing Markov systems, unlike the Bernoulli systems in [14]. This lemma allows us to assume, without loss of generality, that the symbol 0 has identical probability in the two systems.

A remark is due here about a false lead – it may appear that if such an intermediate construction succeeds, we can iterate the construction and construct an isomorphism between the alphabets which a fortiori yields a pointwise measure-preserving isomorphism. This is not possible in general because the non-trivial cases of Ornstein isomorphism are precisely when \( |\Sigma_A| \neq |\Sigma_B| \), and we reach an impasse when we have an odd number of symbols in one alphabet, and an even number of symbols in the other.

Then, we construct an isomorphism between the random objects in two mixing Markov systems \( A \) and \( C \) with equal entropy and with identical probability for 0, in stages. First, for a random object \( x \), we call the pattern of 0s with all other symbols replaced by \( \cdot \) as the skeleton of \( x \). For \( x \in \mathcal{A} \), we identify potential images as those sequences \( y \in \mathcal{C} \), such that their pattern of 0s are identical. This is enabled by the effective Skeleton Lemma. This is the first step to identify potential images of \( x \) under the isomorphism. We now restrict the choices available progressively, until we remain with a unique image for \( x \), through the following stages.

Once we have identified sequences in \( A \) and \( C \) with identical skeletons, we have to “fill in” the non-zero positions by producing a measure-preserving bijection between equal length strings.

\footnote{There is little difference between the requirements of having additive error of \( \varepsilon_n \) and additive error of \( \varepsilon_n \cdot P_A(x) \), except that the later is more convenient for our purpose.}
from the two systems. The definition and technical results about these strings form the “effective 
filler lemma”. In this stage, we identify “filled-in” strings from $A$ and $C$ which could potentially 
be isomorphically mapped to each other. The existence of strings in the two systems with simul-
taneously the same length and approximately the same entropy is a consequence of the effective 
Shannon-McMillan-Breiman theorem. This portion of our proof varies in an essential manner 
from that of Keane and Smorodinsky.

This potential mapping between the strings of $A$ and $C$ can be naturally modeled as a bipartite 
graph. Finally, we prove a version of the Marriage Lemma to form the bijection between the 
strings in the two sequences, which forms a basis for the construction of the layerwise computable 
bijection between the two systems. In the limit, we will map every random infinite sequence $x$ in 
the first system to a unique random infinite sequence $y$ from the second and vice versa. We will 
justify that the overall construction is a layerwise computable function.

3.1 Relevance of the Assumptions

The two most crucial computability results that our result relies on are – first, the definition of 
layerwise computability and second, the effective Shannon-McMillan-Breiman theorem. We now 
broadly justify the appropriateness of these assumptions.

Our algorithm relies on the fact that for any point in the support of the isomorphism, we 
can find skeletons, that is blocks of 0s, of any given length. This is true for all Martin-Löf 
random points, which is crucial in ensuring that our construction is layerwise computable. On 
the other hand, for several computable points – for instance, for periodic sequences, skeletons of 
only finitely many lengths occur. Thus the set of points where our algorithm diverges is dense. 
Hence it seems difficult to adapt topologically inspired notions of discontinuous functions like 
that of Braverman [3] or Nandakumar [18] for our purpose, and measure-theoretic notions of 
computable discontinuous functions like layerwise computability are considerably more natural to 
deal with.

Second, the filler lemma for finding fillers for the skeleton relies on the fact that for every 
Martin-Löf random point, we can find filler strings satisfying a certain entropy bound. The 
classical Shannon-McMillan-Breiman theorem gives us only an almost everywhere behavior which 
leaves the possibility that the construction may fail for a nonempty measure 0 subset of Martin-Löf 
random points. The effective Shannon-McMillan-Breiman theorem of Hochman [7] and Hoyrup 
[11] provides the assurance that we can find such fillers for every Martin-Löf random point.

4 Preliminaries

In this section, we briefly explain the definition of concepts and notation which we use in our 
result.
4.1 Kolmogorov-Sinai Entropy

Kolmogorov [15] and Sinai [28] introduced the notion of the entropy of a transformation, analogous to Shannon entropy, which proved a fruitful tool in the classification of dynamical systems. This notion is, in an essential sense, the only invariant of a dynamical system – all other natural invariants are continuous functions of the entropy [21]. We now describe the notion of Kolmogorov-Sinai entropy.

A probability space is a triple \((X, \mathcal{B}, \mu)\), where \(X\) is a sample space, \(\mathcal{B}\) a \(\sigma\)-algebra on \(X\), and \(\mu\), a probability distribution on \(\mathcal{B}\). Let \(T : X \to X\) be a measurable map. The transformation \(T\) is called measure-preserving if for any measurable set \(B \in \mathcal{B}\), \(\mu(T^{-1}B) = \mu(B)\). A measure-preserving map \(T\) is called an ergodic map if every set \(B \in \mathcal{B}\) where \(T^{-1}B = B\) has measure either 0 or 1.

**Definition 4.1.** A quadruple \((X, \mathcal{B}, \mu, T)\) where \((X, \mathcal{B}, \mu)\) is a probability space and \(T : X \to X\) is an ergodic map, is called a dynamical system.

We now proceed to the definition of entropy of a dynamical system. The chief idea is to introduce a notion analogous to a finite alphabet. Given any dynamical system \((X, \mathcal{B}, \mu, T)\), we can associate it with a process involving finitely many states. Let \(\alpha = (A_1, A_2, \ldots, A_n)\) be a finite collection of measurable subsets of \(X\) which are pairwise disjoint except for measure 0 sets, and cover \(X\) except possibly for a measure 0 set. We can think of the partition containing \(x \in X\) as its 0th “character” – that is, if \(x \in A_i\), then we write \(x[0] = i\).

The entropy of a partition \(\alpha\) is defined to be \(H(\alpha) = -\sum_{i=1}^{n} \mu(A_i) \log_2 \mu(A_i)\). Then for any integer \(i\), \(T^{-i}\alpha\) is the set \((T^{-i}(A_1), \ldots, T^{-i}(A_n))\). This set also partitions \(X\), since \(T\) is a measure-preserving transformation. This partition specifies the “\(i\)th letter” of any point in \(X\).

Now, we need to define concepts analogous to “subsequences”. For this, we introduce the notion of refinement of partitions.

If \(\alpha = (A_1, \ldots, A_n)\) and \(\beta = (B_1, \ldots, B_m)\) are two partitions of \(X\), then the join of the partitions, \(\alpha \lor \beta\) is defined to be the partition

\[
(A_i \cap B_j \mid i = 1, \ldots, n; j = 1, \ldots, m).
\]

For any sequence of integers \(i_1, \ldots, i_k\), we then consider the “least common refinement” \(\alpha[-k+1\ldots0]\), denoted \(\alpha \lor T^{-1}\alpha \lor \cdots \lor T^{-k+1}\alpha\). \[3\] For any point \(x \in X\), the cell containing \(x\) in this refinement represents the characters in the positions \(-k+1, \ldots, -1, 0\).

Using this, for any \(k \in \mathbb{N}\), we define the \(k\)-entropy of the system as \(H_k(\alpha) = \frac{1}{k}H(\alpha \lor T^{-1}\alpha \lor \cdots \lor T^{-k+1}\alpha)\), which represents the average entropy rate of the letters \(x[-k+1\ldots0]\) of any point \(x \in X\). Finally, the asymptotic rate of entropy induced by the partition \(\alpha\) is defined \(\lim_{k \to \infty} H_k(\alpha)\). This limit exists for every stationary, in particular, ergodic systems.

\[3\] The convention of starting from negative indices is standard in the literature on dynamical systems.
Definition 4.2. The entropy of the ergodic \((X, \mathcal{B}, \mu, T)\) with respect to the partition \(\alpha\) is 
\[h(\alpha, T) = \lim_{k \to \infty} \frac{1}{k} H_k(\alpha).\]

Let \(\Pi(X)\) denote the set of all finite partitions of \(X\). The entropy of the transformation \(T\) is defined to be 
\[h(T) = \sup_{\alpha \in \Pi(X)} h(\alpha, T).\]  \(\text{(1)}\)

The supremum in (1) is not easy to compute in general. However, there is a case where the supremum is attained by a fairly simple partition \(\alpha\). We say that \(\alpha\) is a generator of \((X, \mathcal{B}, \mu, T)\) if \(\alpha \lor T^{-1} \alpha \lor \cdots = \mathcal{B}\) – that is, if \(\alpha\) generates the full \(\sigma\)-algebra \(\mathcal{B}\). In this case, we have the famous Kolmogorov-Sinai theorem.

The Kolmogorov-Sinai Theorem. \([15], [28]\) If \(\alpha\) is a generator with respect to \(T\), then 
\[h(\alpha, T) = h(T).\]

For a given dynamical system, from now on, we will assume that a generating partition is given and thus we can view the dynamical system as an alphabet process with left shift being the ergodic transform from the space to itself.

The notion of entropy was then used to settle an open question. This involves the relationship between two dynamical systems \((X, \mathcal{B}, \mu, T)\) and \((Y, \mathcal{C}, \nu, S)\).

Definition 4.3. Two dynamical systems \((X, \mathcal{B}, \mu, T)\) and \((Y, \mathcal{C}, \nu, S)\) are said to be isomorphic to each other if there is a measure preserving invertible map \(\phi: X \to Y\) such that \(\phi T(x) = S \phi(x)\) for \(\mu\)-almost every \(x \in X\).

Now let us observe the following: \(\phi(x)[i] = (S^i(\phi(x)))[0] = (\phi(T^ix))[0]\). Hence as long as we can compute the central coordinates of the images for \(T^ix\) (for all \(i \in \mathbb{Z}\)), we can compute the isomorphism \(\phi(x)\). So, from now on we only wish to determine the central alphabet of the image under the isomorphism.

Kolmogorov proved the following theorem.

Theorem 4.4. (Kolmogorov [15]) If two dynamical systems are isomorphic, then they have the same Kolmogorov-Sinai entropy.

He used this to negate the existence of a specific isomorphism by showing that the systems involved had different entropies. \([2]\)

The converse of the question does not hold in general. To see some examples, see Section 5 of Billingsley [2]. However, Ornstein showed a powerful result: that for a large class of systems, called finitely determined systems, the converse of Kolmogorov’s theorem is true – that is, if two such systems have the same entropy, then there is an isomorphism between them [22]. This construction cannot be “continuous” in general. In a more specific context, Keane and Smorodinsky [14] gave a finitary construction between two Bernoulli systems of the same entropy. We introduce the terminology below.

7
Definition 4.5. An isomorphism is called finitary if for almost every \(x \in X\) there exists a \(j \in \mathbb{N}\) such that for every \(x' \in X\), such that \(x[-j \ldots 0 \ldots j] = x'[-j \ldots 0 \ldots j]\), we have that \((\phi x)[0] = (\phi x')[0]\).

Note that this \(j\) exists only for a measure 1 subset of \(X\), and not necessarily for every point in it. Also, the \(j\) depends on the specific \(x\) that we choose. Keane and Smorodinsky proved that for Bernoulli systems, Ornstein’s construction can be made finitary.

Theorem 4.6. If \((X, \mathcal{B}, \mu, T)\) and \((Y, \mathcal{C}, \nu, S)\) are two Bernoulli systems with the same Kolmogorov-Sinai entropy, then there is a finitary isomorphism between \((X, \mathcal{B}, \mu, T)\) and \((Y, \mathcal{C}, \nu, S)\).

In our work, we show that the above construction can be utilized to construct a layerwise lower semicomputable isomorphism between the sets of algorithmically random objects of two computable mixing Markov dynamical systems. To introduce this strengthening, we now give an overview of the setting of algorithmic randomness.

4.2 Algorithmic Randomness and Layerwise Tests

One of the important applications of the theory of computing is in the definition of individual random objects, finite strings and infinite binary sequences in a mathematically robust way – first defined using constructive measure theory by Martin-Löf [17]. In this paper, we mention a recent generalization of the theory of algorithmic randomness to fairly general spaces, namely, computable metric spaces. Gács [4], and Gács, Hoyrup and Rojas, in a series of works [5], [10] have shown that there are universal tests of randomness in these general spaces. In this paper, we will deal with the Cantor space, where most of the general theory is not directly required. However, we need this theory for two specific purposes – first, we need the definition of a computable probability space. Second, the general theory of computable metric spaces is used to define the notion of layerwise computability [10], [9] which provides a more flexible way to determine whether an element of the space is algorithmically random. This theory plays a crucial role in our result.

Definition 4.7. A space \((X, d)\) is called a computable metric space if it satisfies the following.

1. \(X\) is separable - i.e., it has a countable dense subset \(S\).

2. \(S = \{s_i \mid i \in \mathbb{N}\}\) is a computably enumerable set.

3. For any \(s_i, s_j \in S\), \(d(s_i, s_j)\) are uniformly computable real numbers.

If \(x \in X\) and \(r > 0\), then the metric ball \(B(x, r)\) is the subset of \(X\) of points at less than \(r\) distance from \(x\). We consider a set of ideal balls \(\mathcal{B} = \{B(s, q) \mid s \in S, q \in \mathbb{Q}\}\). The set of ideal balls is associated with a canonical computably enumerable numbering \(\mathcal{B} = \{B_i \mid i \in \mathbb{N}\}\).
Example 4.8. The unit interval $[0, 1]$ endowed with the Euclidean metric, is a computable metric space. The set of dyadic rationals $\{\frac{m}{2^k} \mid m, k \in \mathbb{N}\}$ is a computably enumerable dense subset $S$. The set of canonical balls is then uniquely determined.

Pick any computable enumeration of the rationals. Then it is routine to utilize this to produce a canonical enumeration of the set of ideal balls.

Definition 4.9. An effectively open set is an open set $U$ such that there is a computably enumerable set of indices $E \subseteq \mathbb{N}$ with $\bigcup_{j \in E} B_j = U$.

An effectively compact set is a compact set $K$ such that the set

$$\{(i_1, \ldots, i_n) \mid n \in \mathbb{N} \text{ and } K \subseteq \bigcup_{j=1}^{n_i} B_{i_j}\}$$

is computably enumerable.

A set $V$ is effectively open in $K$ if there is an effective open set $U$ such that $V \cap K = U \cap K$. A set $V$ is decidable in $K$ if $V$ and $X - V$ are effectively open in $K$.

Thus effectively open sets are the analogues of computably enumerable sets. Similarly, we can define notions of computability on effective metric spaces. A function $f : X \to [-\infty, \infty]$ is lower semicomputable if the sets $f^{-1}(q, \infty]$ are uniformly effectively open. A function $f : X \to [-\infty, \infty]$ is upper semicomputable if $-f$ is lower semicomputable, and is computable if it is both upper and lower semicomputable.

Definition 4.10. Let $(X, d, S)$ be a computable metric space. A Borel probability measure $\mu$ on $X$ is computable if the probability of any finite union of canonical balls is computable.

In other words, there is a machine, which for every $\epsilon$ and every finite union of cylinders $C$, returns a rational number with $\epsilon$ of the probability of $C$.

Example 4.11. For the previous example, the Borel measure generated by specifying that $\mu((x, y]) = |y - x|$ is a computable probability measure.

Hoyrup and Rojas \cite{HoyrupRojas2011} prove an effective Prokhorov theorem for computable probability measures on computable metric spaces, which is the basis for their new definition of algorithmic randomness. For this, first we need the notion of a layerwise lower semicomputable function.

A Martin-Löf test $O$ is a sequence of uniformly effectively open sets $O_n$ such that for every $n \in \mathbb{N}$, $P(O_n) < \frac{1}{2^n}$. A point $x$ is said to be Martin-Löf random if for every Martin-Löf test $O$, $x \notin O_n$. If $P$ is a computable probability measure, then the set of Martin-Löf points has $P$ measure 1.

Every computable probability space $(X, P)$ also has a universal Martin-Löf test - that is, there is a Martin-Löf test $U$ such that $x \in X$ is Martin-Löf random if and only if for all $n \in \mathbb{N}$, $x \notin U_n$.\footnote{This is a more restricted notion than that considered in Hoyrup and Rojas \cite{HoyrupRojas2011}.}
Definition 4.12. [10], [8] Let $(X, P)$ be a computable probability space. Let $U$ be a universal Martin-Löf test for $P$. Then the sequence of compact sets $(K_n)_{n=0}^\infty$ where $K_n = X - U_n$ for every $n \in \mathbb{N}$, is defined as the layering of the space. For every $n \in \mathbb{N}$ is called the $n$th layer of the space.

Definition 4.13. A lowersemicomputable function $f : X \to \mathbb{R}$ is called layerwise lowersemicomputable if it is uniformly computable on $(K)_{n=1}^\infty$.

The layerwise lower semicomputable functions may be undefined on every point that is not Martin-Löf random. This is important since our construction diverges on many (but not necessarily all) nonrandom points.

Definition 4.14. A layerwise integrable test is a layerwise lower semicomputable function $t : X \to [0, \infty]$ such that $\int t d\mu$ is finite.

A point $x \in X$ is Martin-Löf random if for every layerwise integrable test $t$, we have $t(x) < \infty$.

The integrable function can be thought of as a martingale process. Thus a point is Martin-Löf random if no layerwise lowersemicomputable martingale can win unbounded money on it.

We will construct an isomorphism between two spaces which is layerwise lower semicomputable. Then we argue that the composition of the layerwise test on the domain and the isomorphism constitutes a layerwise test on the range.

5 Layerwise lower semicomputable isomorphism between processes with equal entropy

In this section, we provide the proof of Theorem 1. We will construct an isomorphism and establish that it is layerwise lower semicomputable. We will elaborate on the overview presented in section 3. The proof is structured in sections, as in the presentation of the Keane-Smorodinksy proof.

Now, we proceed with the stages of the algorithm.

5.1 An Effective Marker Lemma

We want to construct an isomorphism between two mixing Markov systems $A$ and $B$ with equal entropy. We will construct an intermediate mixing Markov system $C$ with a finite alphabet $\Sigma_C$ and computable probability. We designate two symbols 0 and 1. We aim to construct an intermediate system $C$ so that all cylinders $x \in \{0\}^*$ in system $A$ have the same probability as the cylinder $x$ in the system $C$ and similarly all cylinders $y \in \{1\}^*$ in the system $B$ have the same probability as the cylinder $y$ in system $C$.

If we can construct such an intermediate system, then we show that the problem reduces to mapping each cylinder in one system to a cylinder of same length in the other. Let $I_1$ be the layerwise lower semicomputable isomorphism between $A$ and $C$ and $I_2$ be the layerwise lower semicomputable isomorphism between $B$ and $C$. The final isomorphism between $A$ and $B$ is
Let \( c \in x \). Further, a layerwise lower semicomputable isomorphism between new systems trivially yields such an isomorphism preserves layerwise lower semicomputability.

Consequently, the final isomorphism also maps a cylinder to a cylinder of same length. We construct the intermediate system \( C \) in such a way that given a finite cylinder of a Martin-Löf random sequence in \( A \), we can determine the central coordinate of the image in a computably enumerable manner.

5.1.1 Construction of the intermediate system

Observe that we can assume without loss of generality that both \( \Sigma_A, \Sigma_B \) alphabets has at least 3 symbols and all symbols in \( \Sigma_A \) and \( \Sigma_B \) have probability less than \( 1/4 \).

To determine the intermediate system, first we need to fix an alphabet \( \Sigma_C \). Then we need to construct a Turing machine \( M_C \) which, on input \( x \in \Sigma_C^* \) and \( n \in \mathbb{N} \), returns the probability of cylinder \( x \) in \( C \) within error of \( \varepsilon_n \). We require that the probabilities thus computed satisfy the following.

1. For \( x \in \{0\}^* \), \( P_A(x) = P_C(x) \), and for \( y \in \{1\}^* \), \( P_B(y) = P_C(y) \).
2. \( H(C) = H(A) = H(B) \).
3. \( C \) is a mixing Markov process with memory \( m \).

First we determine the size of \( \Sigma_C \). Sort the symbols \( a \in \Sigma_A \) and \( b \in \Sigma_B \) in descending order with respect to \( P_A(a, 3) \) and \( P_B(b, 3) \) respectively. Designate a symbol in \( \Sigma_A \) with the largest probability as 0 and a symbol in \( \Sigma_B \) with the least probability as 1. Note that we are assured that \( P_C(\Sigma_C \setminus \{0, 1\}) > \frac{1}{4} \). Let \( c' \) be the smallest natural number so that \( \frac{1}{4} \log(c' - 2) \geq (H + 1) \).

Let \( c = \max(c', |\Sigma_A|, |\Sigma_B|) \). We define \( \Sigma_C \) to be \( \{0, 1, \ldots, c\} \).

Now we outline the algorithm \( M_C \) for computing the probability \( P_C \). Let the input string be \( x \in \Sigma_C^* \). We can easily satisfy condition 1: \( M_C(x, n) = M_A(x, n) \) when \( x \in \{0\}^* \) and returns \( M_C(x, n) = M_B(x, n) \) when \( x \in \{1\}^* \). For other inputs, we construct the probabilities by induction on the error bound.

If \( |x| \leq m + 1 \), then we do the following. Let, for every \( z \in \Sigma_C^{m+1} \), \( P(z) \) be \( M_C(z, n - 1) \). While either \( |H(P) - H(A)| \) or \( |H(P) - H(B)| \) is greater than \( \varepsilon_n \), we adjust, for strings \( z \in \Sigma_C^{m+1} \) in steps of \( P(z)\varepsilon_n \) so that \( \max\{|H(P) - H(A)|, |H(P) - H(B)|\} \) strictly decreases while maintaining that new probability is consistent with \( P(z, k) \) for \( 1 \leq k \leq n - 1 \). This is always possible – the maximum entropy of a distribution on \( \Sigma_C^{m+1} \) is

\[
(m + 1)c > \max\{4(m + 1)(H + 1), (m + 1) \log |A|, (m + 1) \log |B|\},
\]

\( ^5 \)Otherwise consider systems with alphabets \( \Sigma_A^n \) and \( \Sigma_B^n \) and ergodic transformations \( T_A^n \) and \( T_B^n \) for sufficiently large \( n \). It is easy to see that the new systems are also Markov and both the systems have same entropy \( nH \). Further, a layerwise lower semicomputable isomorphism between new systems trivially yields such an isomorphism between the original systems.
which is at least as large as the maximum possible entropies of $\Sigma_A^{m+1}$ and $\Sigma_B^{m+1}$, and the minimum possible entropy of a distribution on $\Sigma_C^{m+1}$ with $P_C(0)$ and $P_C(1)$ fixed appropriately, is $(m+1) \left[ -2 \times \frac{1}{4} \log_2 \frac{1}{4} - \frac{1}{2} \log_2 \frac{1}{2} \right]$, which matches the minimum possible entropy of $\Sigma_A^{m+1}$ and $\Sigma_B^{m+1}$. Since entropy is a continuous function of the probability vector, by making adjustments to the probabilities on $\Sigma_C^{m+1}$, it is possible to attain a probability vector $P$ on $\Sigma_C^{m+1}$ such that $|H(P) - H| < (m+1)\epsilon$. Although this $P$ might real valued, we are approximating within $\epsilon_n$ in $P_C(\cdot, n)$.

If the input string $x$ contains more than $m+1$ characters, we first assign, for every $z \in \Sigma_C^{m+1}$, $P(z) = M_C(z, n|\cdot)$. Now, for every symbol $a \in \Sigma_C$ and string $y \in \Sigma_C^m$, we let

$$P(a | y) = \frac{P(ya)}{\sum_{c \in \Sigma_C} P yc).$$

Now, compute

$$P(x[1 \ldots q]) = P(x[1 \ldots m]) \prod_{i=m+1}^{q} P(x[i] \mid x[i-m \ldots i-1]).$$

It is clear that the constructed process is a mixing Markov process with memory $m$.

From the construction, it is clear that the system $C$ thus constructed has a same entropy as that of $A$ and $B$. Also, it is clear from the construction that $C$ is a stationary mixing Markov with memory $m$. Note that we also compute $P_C(a|x, n)$ in the process of computing $P_C(\cdot, n)$.

Note that $P(z, n)$ for $z \in \Sigma_C^k$ is not a valid probability distribution – in particular “probabilities” may not add up to 1. But for the purpose of our construction it is immaterial.

5.2 An Effective Skeleton Lemma

We can now consider two systems $A$ and $C$ with $P_A(0) = P_C(0)$. We will now consider those pairs $(x, y) \in A \times C$ such that their patterns of zeroes are “similar”, and progressively restricting this set, we will finally ensure an isomorphism for every pair of random sequences. For this, we now introduce the notion of a skeleton. The skeleton of a finite string is the string we obtain by mapping any non-zero symbol in it to a special character, say $\omega$. Consequently, if the patterns of 0s in two finite strings $x \in \Sigma_A^r$ and $y \in \Sigma_C^r$ are identical, then their skeletons will be identical. In this subsection, we prove an effective version of Keane and Smorodinsky’s Skeleton Lemma [14] (see also Chapter 6, Lemma 5.3 in Petersen[24]). What goes in the blank spaces is called a filler.

The strategy that we will adopt in the isomorphism is to map sequences $x \in \Sigma_A^\infty$ to sequences $y \in \Sigma_C^\infty$ with identical skeletons. The first stage in the construction will be to identify the set of potential pairs of infinite sequences with identical skeletons. To this end, we now define the notion of a skeleton of rank $r$, $r \in \mathbb{N}$, and show that Martin-Löf random sequences in any system have skeletons of all ranks. Owing to the fact that we have only approximation of probabilities of mixing Markov systems, we consider a different setting for skeletons and later, their fillers, from the one considered in [14] for Bernoulli process.
Assume that we have a sequence of positive integers $N_0 < N_1 < \ldots$. (This sequence will be fixed when we discuss the filler lemma, where we establish that it can be computed layerwise.)

For a skeleton of rank $r$ centered at position $i$ in a sequence $x$, we look for the shortest substring centered at $x[i]$ starting and ending with $N_r$ (or more) consecutive 0s. We replace all non-zero symbols with blanks. We further replace the blocks of 0s shorter than $m$ with blanks.

**Definition 5.1.** Let $x \in A^Z$. A skeleton $S_{x,r,i}$ of rank $r$ in $x = [\ldots x_{-2} x_1 x_0 x_1 x_2 \ldots]$ is defined as follows. Starting from $x[i]$, pick the shortest string of the form $0^{n_0} \ell_1 0^{n_1} \ldots \ell_k 0^{n_k}$ such that the following hold.

- Each $\ell_i$ is at least 1, $(1 \leq i \leq k)$.
- Each $n_i$ is at least $m$, $(1 \leq i \leq k)$.
- $n_i < N_r$ for all $1 \leq i \leq k - 1$. Further, both $n_0$ and $n_m$ are greater than or equal to $N_r$.

Note that blocks of zeroes of length less than $m$ are converted to spaces. Thus, except for the extremities of the skeleton of rank $r$, there is no contiguous block of 0s longer than $N_r$. Also, it is routine to see that a rank-$r$ skeleton can be uniquely decomposed into skeletons of rank $r - 1$. [24]

We now show that the skeleton of every Martin-Löf random object in has skeletons of every rank $r$ (with respect to any predetermined sequence $N_1 < N_2 < \ldots$ of numbers) while having sufficiently many blanks in between. This is an effective version of the Skeleton Lemma in [4].

**Definition 5.2.** The length of the skeleton $S_{x,r,i}$, denoted $\ell(S_{x,r,i})$, is defined as follows.

$$\ell(S_{x,r,i}) = |\{i \mid x_i \neq 0, \quad i \in S_{x,r,i}\}|$$

**Lemma 5.3 (Layerwise Skeleton Lemma).** Let $\langle L_r \rangle_{r=1}^{\infty}$ be an increasing sequence of positive integers. Then there is a layering $\langle K'_r \rangle_{r=1}^{\infty}$ of $A$ and an increasing sequence of positive integers $\langle N_r \rangle_{r=0}^{\infty}$ uniformly computably enumerable in $\langle K'_r \rangle_{r=1}^{\infty}$ such that for every $r \in \mathbb{N}$ and every $x \in K'_r$, the following hold.

- There is a skeleton centered at $x[0]$ delimited by $N_r$ many zeroes.
- The central skeleton centered at $x[0]$ and delimited by $N_r$ many zeroes, has length at least $L_r$.

We prove the case when $m = 1$. The general case is similar.

**Proof.** Define $K'_r = \{x \in X \mid \ell(S_{x,r,0}) \geq L_r\}$. (Note that in this step, we choose $N_1, \ldots, N_r$ to determine the rank-$r$ skeleton.) Thus $K'_r$ contains all points $x$ such that their “central skeleton” of rank $r$ contains at least $L_r$ many spaces.
Consider
\[ K' = \bigcup_{n=1}^{\infty} \bigcap_{r=n}^{\infty} K'_r, \]
the set of points in \( X \) such that for large enough ranks \( r \), a skeleton of rank \( r \) contains at least \( L_r \) many non-zero symbols. We form a layerwise lower semicomputable function which attains infinity on each element in \( K'^c \).

Any \( x \) in \( K'^c \) has either of two properties – first, \( x \) does not have any skeleton of rank \( r \) (or above), and second, for every \( n \), there exists some rank \( r \geq n \) such that \( x \) has a central skeleton having less than \( L_r \) many spaces. We will form layerwise integrable functions which will attain \( \infty \) on \( x \) in either of these cases.

**Case I.** Suppose \( x \) has no central skeleton of rank \( r \) or more. By the pigeonhole principle, there is some rank \( r' < r \) such that a rank \( r' \) skeleton appears infinitely often in \( x \). Suppose \( r' \) is the highest rank which appears infinitely often in any skeleton of \( x \), including non-central skeletons.

Let the left zero extremity (analogously, the right zero extremity) of a string \( w \) be the longest block of zeroes at the left end (correspondingly the right end) of \( w \). (These may, of course be empty.) Let \( ZE : \Sigma_A^* \to \{0\}^* \) be the function which returns the shorter among the left zero extremity and right zero extremity.

Consider the following function defined on cylinders of \( A^Z \). The function \( f : A^* \to [0, \infty) \) is defined by
\[
f(a_1 w a_2) = \begin{cases} 
\frac{1}{(1-PA(00|w))}f(w) & \text{if } |ZE(w)| = N_r, \text{ and } a_1 a_2 \neq 00 \\
0 & \text{if } |ZE(w)| = N_r, \text{ and } a_1 a_2 = 00 \\
f(w) & \text{otherwise}
\end{cases}
\]

Define the function \( S : A^Z \to [0, \infty) \) by \( S(x) = \sup_n f(x[-n \ldots 0 \ldots n]) \). Since \( PA \) is computable, we can conclude that \( S \) is layerwise lower semicomputable.

For infinitely many \( n \), a skeleton of rank \( r' \) will appear as the extremities of \( x[-n \ldots 0 \ldots n] \). Hence the subsequent bits on the left and the right cannot both be 0. In this case, \( f(x[-n-1 \ldots 0 \ldots n+1]) > f(x[-n \ldots 0 \ldots n]) \). Thus, \( S(x) = \infty \).

We observe that \( \int f(\lambda) dPA = 1 \), viewing \( f(\lambda) \) as a constant function defined on every \( x \in X \). Similarly, on any cylinder \( w \), if \( w \) does not have extremities of the form \( 0^{N_r} \), then \( f(a_1 w a_2) = f(w) \), and we have
\[
\sum_{a_1 a_2 \in \Sigma_A^2} f(a_1 w a_2) PA(a_1 a_2 \mid w) = f(w) \sum_{a_1 a_2 \in \Sigma_A^2} PA(a_1 a_2 \mid w),
\]
which is \( f(w) \). If \( w \) ends in extremities of the form \( 0^{N_r} \), then
\[
\sum_{a_1 a_2 \in \Sigma_A^2 \setminus \{00\}} f(a_1 w a_2) PA(a_1 a_2 \mid w) = f(w) \frac{[1 - PA(00 \mid w)]}{1 - PA(00 \mid w)},
\]

14
which is \( f(w) \) as well. So we have that

\[
f(w)P_A(w) = \sum_{a_1a_2 \in \Sigma^2} f(a_1 w a_2)P_A(a_1 w a_2).
\]

Thus, we can see that

\[
\int S(x) dP_A = \int \limsup_n f(x[-n \ldots 0 \ldots n]) dP_A \leq \sup_n \int f(x[-n \ldots 0 \ldots n]) dP_A = 1,
\]

where the inequality follows by Fatou’s lemma.

**Case II.** Now suppose that for every \( n \), there is a central skeleton in \( x \) of rank \( r \geq n \) such that \( \ell(S_{x,r,0}) < L_r \). This implies that within at most \( L_r(N_r - 1) \) characters around \( x_0 \), the block \( 0^{N_r} \) will occur in \( x \).

Consider the function \( g_r : A^* \to [0, \infty) \) defined by

\[
g_r^k(\lambda) = \frac{1}{2^rL_r(N_r - 1)}
\]

\[
g_r^k(a_1 w a_2) = \begin{cases} \frac{1}{P_A(00|w)}g_r(w) & \text{if } k \leq |w| < N_r \text{ and } a_1a_2 = 00 \\ 0 & \text{if } k \leq |w| < N_r \text{ and } a_1a_2 = 00 \\ f(w) & \text{otherwise.} \end{cases}
\]

As in case I, we can verify that for all cylinders \( w \),

\[
g_r^k(w)P_A(w) = \sum_{a_1a_2 \in \Sigma^2} g_r^k(a_1 w a_2)P_A(a_1 w a_2).
\]

Consider the function \( g_r : A^* \to [0, \infty) \) defined by

\[
g_r = \sum_{k=1}^{L_r(N_r-1)} g_r^k.
\]

We know that if \( x \) has a deficient rank \( r \) at length \( k \), then

\[
g_r(x) \geq \frac{1}{2^rL_r(N_r - 1)PA(0)^{N_r}} \geq 1
\]

if we choose large \( N_r \) in a suitable manner.

Finally, consider the aggregate function \( S : A^Z \to [0, \infty) \) defined by \( S = \sup_n \sum_{r=1}^{\infty} g_r \). Then, as in case I, we see that \( S \) is layerwise lower semicomputable and integrable. Since by assumption \( x \) has infinitely many \( r \) for which \( g_r \) attains at least 1, we have that \( S(x) = \infty \).

We will now proceed to choose this sequence of \( L_r \)'s that is assumed in Lemma 5.3.
5.3 Effectively determining $L_r$ and Filler lemma

In the last subsection, we assumed that we have a sequence $L_0 < L_1 < \ldots$ of natural numbers. For every $i, r \in \mathbb{N}$ and $x \in \mathcal{A} \cup \mathcal{C}$, a skeleton in $x$ of rank $r$ at position $i$ was the shortest string centered at $x[i]$ and delimited by the earliest appearance of $N_r$ many zeroes and at least $L_r$ many spaces. We now see how we determine this sequence in a layerwise lower semicomputable manner.

We define a sequence of $\langle L_r \rangle_{r=1}^\infty$ for the lengths of the skeletons of rank $r$ inductively. We choose the sequence $\langle N_i \rangle_{i=1}^\infty$ such that a skeleton of rank $r$ has length at least $L_r$. We compute the lengths $L_r$ layerwise, in such a way that properties analogous to the asymptotic equipartition property hold for the skeletons of rank $r$ for every Martin-Löf random sequence. This will allow us to construct a provably isomorphic map between $\mathcal{A}$ and $\mathcal{C}$.

Let $\eta_r = \min_{i \in \{A,C\}} \min_{a \in \Sigma_i, b \in \Sigma_i} P_i(a \mid b, r)$. Similarly, let $\theta_r = \max_{i \in \{A,C\}} \max_{a \in \Sigma_i, b \in \Sigma_i} P_i(a \mid b, r)$. We want a strictly increasing sequence $\langle L_r \rangle_{r=1}^\infty$ such that

$$\lim_{r \to \infty} \frac{1}{2-\ell_r(\varepsilon_{r-1} - \varepsilon_r)} = 0.$$

Note that $L_r = 3 + r$ is such a sequence – since $\frac{1}{\eta_r} \leq \frac{1}{\eta (1-\varepsilon_r)}$ where $\eta$ is minimum non-zero conditional probability and $\varepsilon_{r-1} - \varepsilon_r = \varepsilon_r$. In the following, we assume $L_r = 3 + r$. We use this specific value in lemma 5.3.

Let $S$ be a skeleton of length $\ell$ and rank $r$. So, the filler for $S$ in $\mathcal{A}$ is $\Sigma_A^\ell$ and in $\mathcal{C}$ is $\Sigma_C^\ell$. For the moment let us only consider the system $\mathcal{A}$ – similar is done for $\mathcal{C}$ in order to construct the isomorphism.

Let $S$ be a skeleton of length $\ell$ and rank $r$. Let $\mathcal{F}(S) \subseteq \Sigma_A^\ell$ denote the set of fillers for $S$ in $\mathcal{A}$. Let $Z_S$ denote the indices of 0s in $S$ and let the blanks be in positions $s_1, s_2, \ldots, s_\ell$. Note that $Z_S$ does not contain indices of a 0 block shorter than $m$.

Given a filler $F \in \mathcal{F}(S)$ and an index set $I \subseteq \{s_1, s_2, \ldots, s_\ell\}$ over the blank places of $S$, let $\langle I, F, S \rangle$ denote the cylinder generated by setting 0s from $S$ and setting $i^{th}$ position for $i \in I$ with the corresponding symbol in the filler $F$.

Now, for an $\in \mathbb{N}, n \geq r$ we define an equivalence relation $\sim_n$ for error bound $\varepsilon_n$ on $\mathcal{F}(S)$ and denote equivalence class of $F$ by $\tilde{F}_n$. We define a subset of places $J(F, n) \subseteq \{s_1, s_2, \ldots, s_\ell\}$ for each $F$ and declare $F \sim_n F'$ if $J(F, n) = J(F', n)$ and $F$ agrees with $F'$ on $J(F, n)$.

For a fixed $n \geq r$ and $F$, we define $J$ inductively on the rank of the skeleton.

For a skeleton $S$ of rank 1 and length $\ell$, we proceed as follows. Pick the largest positive integer $k$, $k \leq \ell$ such that

$$P_A(\langle \{s_1, s_2, \ldots, s_k\}, F, S \rangle, n) \geq \frac{3}{2\eta_1} 2^{-\ell + |Z_S|(H-\varepsilon_1)}.$$

\[^{a}\text{Note that here we deviate from the original construction. Without loss of generality, } \eta_r > 0 \text{ – since if a } m + 1 \text{ length sequence of alphabets has probability 0, we can safely ignore those strings as we are required to succeed only in a measure 1 subset.}
\[^{b}\text{In contrast with [14], we include 0s in our filler alphabets.}
\[^{c}\text{We define } J(F, n) \text{ only when } n \geq r\]
Then, let $J(F, n) = Z_S \cup \{s_1, s_2, \ldots, s_k\}$.

Now, for a rank $r \geq 2$ skeleton $S$ and $F \in \mathcal{F}(S)$, we do the following: Let us assume that $S = S_1 \times S_2 \times \cdots \times S_i$ is the skeleton decomposition of $S$ where each $S_i$ is of rank $r - 1$. Also, let $F_1, F_2, \ldots, F_i$ are the corresponding fillers which coincide with $F$. We assume that we have determined $J(F_i, n \log 3t)$ inductively for each $F_i$.

Let $J_0(F, n) = \cup_{i=1}^{t} J(F_i, n \log 3t)$ These are the positions in $S$ which have already been determined in the previous rank.

Also, let $\{s_1, \ldots, s_i\} \setminus J_0(F, n) = \{t_1 \ldots t_u\}$, where $t_1 < \cdots < t_u$. These are the positions in the skeleton $S$ which have not been fixed by any rank $r - 1$ subskeletons. Then, we set $J(F, n) = Z_S \cup J_0(F, n) \cup \{t_1, \ldots, t_k\}$, where $k \leq u$ is the largest positive integer such that

$$P_A(\langle t_1, \ldots, t_{k-1} \rangle \cup J_0(F, n), F, S), n ) \geq \frac{(1 + \varepsilon_r)}{\eta_r} 2^{-(l + |Z_S|)(H - \varepsilon_r)}.$$

Note that we allow $J(F, n)$ to be $Z_S \cup \{s_1, s_2, \ldots, s_l\}$ and $\eta_r/(1+\varepsilon_r)$ is a pessimistic approximation of true minimum conditional probability of an alphabet. Let $x = 0^{\ell_1} x_1 0^{\ell_2} x_2 0^{\ell_3} \cdots 0^{\ell_i} x_i 0^{\ell_{i+1}}$ be a string where $\ell_i > m$ for all $1 \leq i \leq t + 1$. Let

$$P_A(x, n) = \frac{\prod_{i=1}^{t} P_A(0^{\ell_i} x_i 0^{\ell_{i+1}}, n \log 3t)}{\prod_{i=2}^{t} P_A(0^{\ell_i})}.$$

By the Markov property, we have that

$$P_A(x) = P_A \left(0^{\ell_1} x_1 0^{\ell_2} \right) \prod_{i=2}^{t} P \left(x_i \left| 0^{\ell_1} x_1 \cdots 0^{\ell_{i-1}} \right. \right) = P_A \left(0^{\ell_1} x_1 0^{\ell_2} \right) \prod_{i=2}^{t} P \left(x_i \left| 0^{\ell_1} x_1 \cdots 0^{\ell_{i-1}} \right. \right).$$

Hence,

$$|P_A(x) - P_A'(x, n)| \leq \varepsilon_n P_A'(x, n). \quad (2)$$

So, $P_A'(x, n)$ can be used in place of $P_A(x, n)$ but for the fact we cannot compute $P_A(0^{\ell_i})$ exactly. But we use the essentially multiplicative nature of $P_A$ and that it approximates $P_A()$ in the proof of lemma 5.9. The approximation is as follows: $|P_A(x, n) - P_A'(x, n)| \leq 2\varepsilon_r P_A(x, n)$. Also, we note that for a given $F \in \mathcal{F}(S)$ and an integer $n$, $J(F, n) \subseteq J(F, n + 1)$. In other words, if we decrease the error bound in estimation of probability the equivalence relation can only get finer.

Similar relations holds for $C$.

Note that $\theta_r$ and $\eta_r$ is bounded from below and above with respect to $r$ for a mixing Markov process. Then we have the following bounds.

**Lemma 5.4** (Filler Lemma). There is a layering $(K''_p)_{p=1}^{\infty}$ such that or every $n$, there is a large enough $r$ such that for every skeleton $S$ of rank $r$ and length $\ell$ corresponding to $x \in K''_p$, we have:

---

9The purpose of $n \log 3t$ will be clear in lemma 5.9
1. For all \( F \in \mathcal{F}(S) \), \( P_A(\bar{F}_r, n) \geq (1 - \varepsilon_n)2^{-L(H-\varepsilon_r)} \)

2. For all \( F \in \mathcal{F}(S) \) except maybe on a set of measure \( \varepsilon_n \):
   
   \( a \) \( P_A(\bar{F}_r, n) < \frac{1 + \varepsilon_n}{1 - \varepsilon_r} \frac{1}{\eta_r} 2^{-L(H-\varepsilon_r)} \)
   
   \( b \) \( \frac{1}{n} \left| J(F, r) \right| > 1 - \frac{3}{|\log_2 n|} \varepsilon_r \)

where \( L = \ell + |Z_S| \).

**Proof.** From the effective Shannon-McMillan-Breiman theorem, we know that there is a layering \( \langle K''_{p=1} \rangle \) such that for all \( p \) there is a \( k_p \) so that for all \( k \geq k_p \), \( \Sigma_A = K''_p \cup (K''_p)^c \), with the following properties:

- \( P_A(K''_p) \geq 1 - \varepsilon_p \)
- For each \( x \in K''_p \) we have
  \[ \frac{1 + \varepsilon_p 2^{-k(H+\varepsilon_p)}}{\eta_p} < P_A(x, p) < \frac{1 - \varepsilon_p 2^{-k(H-\varepsilon_p)}}{\eta_p}. \]

Now given an \( n \), let \( r \geq n + 1 \) be such that \( L_r \geq k_n \). Such an \( r \) exists, since \( \{L_r\} \) is an increasing sequence.

1. Let \( J(F, r) = Z_S \cup J_0(F) \cup \{t_1, t_2, \ldots, t_w\} \). Then
   \[ P_A(\bar{F}_r, n) \geq (1 - \varepsilon_n)P_A(\langle J_0(F) \cup \{t_1, t_2, \ldots, t_w\}, F, S \rangle) \]
   \[ \geq \frac{1 - \varepsilon_n}{1 + \varepsilon_r} P_A(\langle J_0(F) \cup \{t_1, \ldots, t_w\}, F, S, r \rangle) \]
   \[ \geq \frac{1 - \varepsilon_n}{1 + \varepsilon_r} P_A(\langle J_0(F) \cup \{t_1, \ldots, t_{w-1}\}, F, S, r \rangle) \times P_A(F[t_w]) \langle J_0(F) \cup \{t_1, \ldots, t_{w-1}\}, F, S, r \rangle) \]
   \[ \geq \frac{1 - \varepsilon_n}{1 + \varepsilon_r} \frac{1 + \varepsilon_r}{\eta_r} 2^{-L(H-\varepsilon_r)} \times P_A(F[t_w]) \langle J_0(F) \cup \{t_1, \ldots, t_{w-1}\}, F, S, r \rangle \]
   \[ \geq (1 - \varepsilon_n)2^{-L(H-\varepsilon_r)}. \]

where the inequality before the last follows from the definition of \( J(F, r) \).

2. (a) If \( |J(F, r)| < L \), then by the definition of \( J(F, r) \), we have
   \[ P_A(\bar{F}_r, n) \leq (1 + \varepsilon_n)P_A(\bar{F}_r) \leq \frac{1 + \varepsilon_n}{1 - \varepsilon_r} P_A(\bar{F}_r, r) < \frac{1 + \varepsilon_n}{1 - \varepsilon_r} \frac{1 + \varepsilon_r}{\eta_r} 2^{-L(H-\varepsilon_r)} < \frac{1 + \varepsilon_n}{1 - \varepsilon_r} \frac{1}{(1 - \varepsilon_r)2^{-L(H-\varepsilon_r)}} \]

If \( |J(F, r)| = L \), then \( \bar{F}_r = F \). But \( |F| = L \geq L_n \geq k_n \) and hence
\[ P_A(F, n) \leq (1 + \varepsilon_n)P_A(F) \leq \frac{1 + \varepsilon_n}{1 - \varepsilon_r} P_A(F, r) < \frac{1 - \varepsilon_r}{\eta_r} 2^{-L(H-\varepsilon_r)} \]

unless \( F \in (K''_p)^c \) and \( P_A((K''_p)^c) < \varepsilon_r < \varepsilon_n \).
(b) Without loss of generality, assume that (a) holds. (Otherwise we already have that such $F$ has to be in $\varepsilon_r$ measure set.) Let $L = |J(F, r)| \geq 3L\varepsilon_r/|\log_2 \theta_r|$. Then,

\[
P_A(F, n) \leq \frac{1 + \varepsilon_n}{1 - \varepsilon_r} P_A(F, r)
\]

\[
\leq \frac{1 + \varepsilon_n}{1 - \varepsilon_r} P_A(\tilde{F}_r, r) \prod_{i \notin J(F, r)} P_A(F[i]|(J_0(F) \cup \{t_1, \ldots, t_{w-1}\}, F, S), r)
\]

\[
\leq \frac{1 + \varepsilon_r}{1 - \varepsilon_n} P_A(\tilde{F}_r, r) \cdot \theta_r^{3L\varepsilon_r/|\log_2 \theta_r|} \leq \frac{1}{(1 - \varepsilon_n)^2(1 - \varepsilon_r)^2} \frac{1 + \varepsilon_r}{\eta_r} 2^{-L(H-\varepsilon_r)2-3L\varepsilon_r}
\]

\[
\leq \frac{2^{-L\varepsilon_r}}{(1 - \varepsilon_n)^2(1 - \varepsilon_r)^2} \frac{1 + \varepsilon_r}{\eta_r} 2^{-L(H+\varepsilon_r)} \leq \frac{1 + \varepsilon_r}{\eta_r} 2^{-L(H+\varepsilon_r)}.
\]

Here the last inequality follows from $L_r = 3 + r$. We also use the inequalities $\frac{1 + \varepsilon_n}{1 - \varepsilon_r} < \frac{1 + \varepsilon_n}{1 - \varepsilon_n} < 2^{-1}$. In this case $F$ must belong to the set $(K_i')^c$ of measure less than $\varepsilon_r$. Hence the set on which $L$ can violate the bound has measure $< 2\varepsilon_r \leq \varepsilon_n$.

\[\square\]

### 5.4 Societies and Marriage Lemma

Once we have determined the filler alphabets and filler probabilities for $A$ and $C$, we are now in a position to start building the isomorphism between cylinders from $A$ and $C$ which have identical skeletons. Each cylinder in $A$ has multiple possible matches in $C$ and conversely. We model this as a bipartite graph with the filled-in skeletons from $A$ forming the left set of vertices, and those from $C$ forming the right set. The presence of an edge represents a potential match between the corresponding vertices. We obtain this by a minor variant of Keane and Smorodinsky’s marriage lemma, where the variation is forced by the fact that we have only an approximation of the actual probabilities of the vertices.

Let us assume we are given two probability space $(\Omega_1, \mu_1), (\Omega_2, \mu_2)$, with both $\Omega_1$ and $\Omega_2$ finite. A **society** or a knowledge relationship is a map $f : \Omega_1 \rightarrow 2^{\Omega_2}$ so that for all $X \subseteq \Omega_1$, we have $\mu_1(X) \leq \mu_2(f(X))$ where $f(X)$ is defined in the natural way. When the underlying probabilities are clear from context, we denote a society as $f : \Omega_1 \leadsto \Omega_2$. Now consider the undirected knowledge graph constructed out of the knowledge relationship, with vertices set $\Omega_1 \cup \Omega_2$ and edge set $E = E_1 \cup E_1^{-1}$ where $E_1 = \{(a, b) \in \Omega_1 \times \Omega_2 : b \in f(a)\}$. Note that the knowledge graph is bipartite by definition. Now we define a couple of notions which provides us with the tools necessary for defining isomorphism:

**Definition 5.5 (Join of societies).** Given societies $f_i : \Omega_{i,1} \leadsto \Omega_{i,2}$ for $1 \leq i \leq j$, we define their **join** $f : \Omega_{1,1} \times \Omega_{2,1} \times \ldots \times \Omega_{j,1} \prod_{\prod} \Omega_{1,2} \times \Omega_{2,2} \times \ldots \Omega_{j,2}$ as a map $f : \Omega_{1,1} \times \Omega_{2,1} \times \ldots \Omega_{j,1} \rightarrow 2^{\Omega_{1,2} \times \Omega_{2,2} \times \ldots \Omega_{j,2}}$ where $(\omega_1, \omega_2, \ldots, \omega_j) \in f(\nu_1, \nu_2, \ldots, \nu_j)$ for $\omega_i \in \Omega_{i,2}, \nu_i \in \Omega_{i,1}$ iff $\omega_i \in f_i(\nu_i)$. 

19
**Definition 5.6** (ε-robust). Consider a society \( f \) between probability spaces \( (\Omega_1, \mu_1), (\Omega_2, \mu_2) \). Consider the undirected knowledge graph \( G = V_1 \cup V_2 \cup \cdots \cup V_w \) where \( V_i \)'s are connected components of \( G \). Given an \( \varepsilon > 0 \), society \( f \) is called ε-robust if for all \( 1 \leq i \leq m \), for all \( X \subset V_i \cap \Omega_1 \) and for all \( Y \subset V_i \cap \Omega_2 \), we have:

\[
\mu_1(X)(1 + \varepsilon) \leq \mu_2(f(X))(1 - \varepsilon) \\
\mu_2(Y)(1 + \varepsilon) \leq \mu_1(f^{-1}(Y))(1 - \varepsilon).
\]

It is easy to see that for \( \varepsilon > 0 \), an ε-robust society is a society.

Note that we only consider proper subsets \( X \) and \( Y \) in the above definition, since \( \mu_1(V_i \cap \Omega_1) = \mu_2(V_i \cap \Omega_2) \). This easily follows from the fact that \( f \) and \( f^{-1} \) are societies. Also note that a society \( f \) is ε-robust iff the dual of the society \( f^{-1} \) is ε-robust.

A society is minimal if the removal of any edge from it will violate the condition for a society. In the construction of an isomorphism, we consider various minimal sub-societies of given societies. Now since we only have some approximation of probabilities of vertices, we have to be careful while removing edges from knowledge graph to construct minimal sub-society. The next lemma shows that it is enough to consider only ε-robust minimal societies for our purpose.

**Lemma 5.7.** Given a society \( f \) between probability spaces \( (\Omega_1, \mu_1), (\Omega_2, \mu_2) \) and a minimal sub-society \( g \), there is an \( \varepsilon > 0 \) so that \( g \) is ε-robust.

**Proof.** We know that the minimal sub-society \( g \) is generated by a joining\(^\text{10}\), say \( \mu \) - that is, a joint distribution \( \mu \) on \( \Omega_1 \times \Omega_2 \) such that \( \mu_1 \) and \( \mu_2 \) are its marginals (see Chapter 6 of [24]). Consider the knowledge graph \( G \) for the society \( g \). Note that \( G \) is a finite graph. Let \( G = V_1 \cup V_2 \cup \cdots \cup V_w \), where \( V_i \)'s are the connected components. Consider any arbitrary component \( V_i \). Let \( X \subset V_i \cap \Omega_1 \). Now \( X \subset g^{-1}(g(A)) \). So,

\[
\mu_1(X) = \sum_{a \in X} \mu_1(a) = \sum_{a \in X} \sum_{b \in g(X)} \mu(a, b) < \sum_{a \in g^{-1}(g(X))} \sum_{b \in g(X)} \mu(a, b) = \sum_{b \in g(X)} \sum_{a \in g^{-1}(g(X))} \mu(a, b) = \mu_2(g(X))
\]

Using a similar argument, we can show that for \( Y \subset V_i \cap \Omega_2 \), \( \mu_2(Y) < \mu_1(g^{-1}(Y)) \). So there is an \( \varepsilon' > 0 \) so that \( \mu_1(X)(1 + \varepsilon) \leq \mu_2(f(X))(1 - \varepsilon) \) and \( \mu_2(Y)(1 + \varepsilon) \leq \mu_1(f^{-1}(Y))(1 - \varepsilon) \). Let \( \varepsilon \) be minimum of all such \( \varepsilon' \) where minimum is taken over all \( i, X \) and \( Y \). \( \square \)

Now we quote a variant of the Marriage Lemma.

**Lemma 5.8** (Marriage Lemma). For any given society \( S \) between \( (\Omega_1, \mu_1) \) and \( (\Omega_2, \mu_2) \), any minimal sub-society \( R \) has the property that \( |\Omega_2| > |\{ w \in \Omega_2 : (\exists w_1, w_2 \in \Omega_1)(w_1 \neq w_2 \wedge w_1 Rw \wedge w_2 Rw) \}| \).

\(^{10}\)In the literature, the joining operation is also known as coupling.
The proof is exactly analogous to [14], see Chapter 6 of [24].

During the construction of isomorphism, we compute various minimal subsocieties. There can be many such minimal subsocieties and “inconsistent” choices in different stages may break the construction. In the following subsections, we describe a way of choosing the minimal subsocieties such that the construction goes through.

5.5 Construction of the isomorphism

We now have a skeleton $S$ common to two sequences $x \in \Sigma_A^\infty$ and $y \in \Sigma_C^\infty$, and have defined an equivalence relation on the fillers for $S$ in $\Sigma_A^*$ and $\Sigma_C^*$, for a desired level of error. We now inductively build societies between equivalence classes of fillers of $A$ and of $C$ and use the marriage lemma from the preceding section to define an isomorphism between $A$ and $C$.

Given a skeleton $S$ of rank $r$, $r \geq 1$, and length $\ell$, let $F(S)$ and $G(S)$ denote the set of its fillers in $A$ and $C$. Given $n \geq r$, let $\tilde{F}(S,n)$ and $\tilde{G}(S,n)$ denote the set of equivalence classes with respect to the equivalence relation $\sim_n$ (i.e., $\tilde{F}(S,n) = \{ \tilde{F}_n : F \in F(S) \}$ and $\tilde{G}(S,n) = \{ \tilde{G}_n : G \in G(S) \}$). We denote the $\varepsilon_n$-robust societies between $\tilde{F}(S,n)$ and $\tilde{G}(S,n)$ by induction on $r$:

$$R_{S,n} : \tilde{F}(S,n) \rightsquigarrow \tilde{G}(S,n) \text{ if } r \text{ is odd, and } R_{S,n} : \tilde{G}(S,n) \rightsquigarrow \tilde{F}(S,n) \text{ otherwise.}$$

The measure for every $F \in \tilde{F}(S,n)$ is $P_A(F,n)$ and that for every $G \in \tilde{G}(S,n)$ is $P_C(G,n)$.

Fix an $n$. For $r = 1$, build a trivial society where each of $\tilde{F}(S,n)$ knows each of $\tilde{G}(S,n)$. Construct a minimal $\varepsilon_n$-robust sub-society $R_{S,n}$ of the trivial society.

Now construct the inductive construction: let $r > 1$ be even. Let $S$ be a skeleton of rank $r$. Let $S = S_1 \times S_2 \times \cdots \times S_t$ be a rank $r - 1$ skeleton decomposition of $S$. Assume that we have a procedure to define societies for all $S_i$ ranks at most $r - 1$ skeleton and any desired precision. Let us consider $R_{S,n,\log 3t} : \tilde{F}(S_1,n \log 3t) \rightsquigarrow \tilde{G}(S_1,n \log 3t)$ for $i = 1, 2, \ldots, t$. Note that we are using induction only on $r$ and not $n$ – for a larger $n$, we repeat the induction procedure from scratch.

Consider their duals $R_{S,n,\log 3t}^* : \tilde{G}(S_1,n \log 3t) \rightsquigarrow \tilde{F}(S_1,n \log 3t)$. Construct the join of societies $R : \tilde{G}(S_1,n \log 3t) \times \tilde{G}(S_2,n \log 3t) \times \cdots \times \tilde{G}(S_t,n \log 3t) \xrightarrow{\text{prod}} \tilde{F}(S_1,n \log 3t) \times \tilde{F}(S_2,n \log 3t) \times \cdots \times \tilde{F}(S_t,n \log 3t)$.

Let $\tilde{F}(S,n) = \tilde{F}(S_1,n \log 3t) \times \cdots \times \tilde{F}(S_t,n \log 3t)$ and $\tilde{G}(S,n) = \tilde{G}(S_1,n \log 3t) \times \cdots \times \tilde{G}(S_t,n \log 3t)$. So, $R : \tilde{G}(S,n) \rightsquigarrow \tilde{F}(S,n)$.

Lemma 5.9. The $R$ constructed above is $\varepsilon_n$-robust with respect to measure $P_C(\cdot, n)$ and $P_A(\cdot, n)$.

We omit the proof – it is routine to verify the conditions of robust society hold when we approximate $P_A(\cdot)$ and $P_C(\cdot)$ with $P_A'(\cdot)$ and $P_C'(\cdot)$ and use equation 2.

Since $\tilde{F}(S,n)$ is determined by $J_0(F,n)$ and $\tilde{F}(S,n)$ is determined by $J(F,n)$, the latter is finer equivalence class. So we may consider $R : \tilde{G}(S,n) \rightsquigarrow \tilde{F}(S,n)$, where each $\tilde{F}(S,n)$ is split into multiple $\tilde{F}(S,n)$s and the knowledge relation is extended accordingly. Construct the minimal $\varepsilon_n$-robust sub-society $U : \tilde{G}(S,n) \rightsquigarrow \tilde{F}(S,n)$. From $U$, construct $R_{S,n} : \tilde{G}(S,n) \rightsquigarrow \tilde{F}(S,n)$ such that $R_{S,n}(\tilde{G}(S,n)) = U(\tilde{G}(S,n))$ where $\tilde{G}(S,n)$ is uniquely determined by finer equivalence class $\tilde{G}(S,n)$. 

21
Now, we elaborate on how we construct the minimal $\varepsilon_n$-robust sub-societies in Algorithm 1.

**Algorithm 1** Computing the minimal robust sub-society

**Input:** $n, R : \mathcal{G}(S,n) \rightsquigarrow \mathcal{F}(S,n)$.

1: for $i = 1; i \leq n; i \leftarrow i + 1$ do
2: Construct $E : \mathcal{G}(S,i) \rightsquigarrow \mathcal{F}(S,i)$ from $R$ with measures $P_j(\cdot, i)$ for $j \in \{A, C\}$ by merging various equivalence class in both sides of the society
3: while some edge can be removed do
4: cycle through edges of $E$ in dictionary ordering
5: if an edge $e$ can be removed while the remaining graph remains $\varepsilon_i$-robust then
6: remove an edge $e$.
7: update $R$ by removing all the edges corresponding to $e$.
8: end if
9: end while
10: end for
11: return $R$

Note that the society constructed in step 2 of the algorithm gives us a $\varepsilon_i$-robust society.

For odd $r$, we switch the role of $F$ and $G$.

Now let us describe the construction of the isomorphism: For a $x \in K'_r \cap K''_r$, let $S_{x,r,i_r}$ denote the skeleton of rank $r$ which occurs in $x$ and contains the current central co-ordinate. Let the position of current central co-ordinate in $S_{x,r,i_r}$ be denoted by $i_r$. Given an $n \in \mathbb{N}$, let $\mathcal{F}_r(x,n)$ denote the filler equivalence class of the filler that occur in $x$ corresponding to $S_r(x)$ with respect to the equivalence relation $\sim_n$. We use similar notation for $C$ where $G$ replaces $F$.

For $x \in A$ such that $x \in K'_r \cap K''_r$, we find large enough even $r$ (this $r$ is computable from $r'$) such that $\forall G_r(x,r) \in R_{S_r(x),r}^{-1}(\mathcal{F}_r(x,r))$, we have that $\mathcal{G}_r(x,r)[i_r]$ is defined (it stabilizes thenceforth). Now the shift preserving map $\phi$ is so defined that

$$(\phi(x))[0] = \begin{cases} 0 & \text{if the block of 0 containing } i_r \text{ is longer than } m \\ \mathcal{G}_r(x,r)[i_r] & \text{otherwise.} \end{cases}$$

Note that this definition specifies all the co-ordinates of $\phi(x)$ because we want it to be shift preserving. It is known that if a measure-preserving shift applied to a Martin-Löf random point yields a Martin-Löf random point [18, 0, 26].

This concludes the description of the algorithm for constructing $\phi$. We now show that algorithm is layerwise lower semicomputable.

### 5.6 Proof that $\phi$ is a layerwise lower semicomputable isomorphism

Now we show that $\phi$ is indeed an isomorphism and well-defined n every Martin-Löf random element in $A$, and similarly, that $\phi^{-1}$ is well-defined for every Martin-Löf random element in $C$.
argue about the measures. The assignment implies for all $x \in x_{a - 1}$ to $x_{1}$ in positions $\phi$ where we use the estimates given by lemma 5.4.

We show that we can always find an $r$ sufficiently large to stabilize the construction of the society. Let us consider the case when for a given $r$-rank skeleton $S$, $n$ is so large that $R_{S,n}$ stabilizes (i.e., it remains unchanged for any larger $n$) – such a $n$ exists due to Lemma 5.7 and the fact that $J(\cdot, n)$s are non-decreasing in $n$ and bounded above. Call such stabilized society $R_S : \mathcal{G}(S) \sim \mathcal{F}(S)$. Then the following result holds.

**Lemma 5.10 (Assignment Lemma).** If $x \in A$ such that $x \in G_{r'} \cap G_{r''}$ with $x[0]$ not contained in a block of 0 longer than $m$, then there is an even $r$, computable from $r'$, such that

1. With respect to the society $R_{S_i}(x) : \mathcal{G}(S_i(x)) \sim \mathcal{F}(S_i(x))$, $R_{S_i}(x)^{-1}(\mathcal{F}(S_i(x)))$ is a singleton, say, $\mathcal{G}(S_i(x))$.

2. $i_r(x) \in J_0(\mathcal{G}(S_i(x)))$.

We omit the proof of this lemma – it is similar to the proof of the assignment lemma given in 23 where we use the estimates given by lemma 5.4.

Now we show how the above lemma ensures the existence of the map $\phi$ for every $x \in G_{r'} \cap G_{r''}$. If the co-ordinate $i_r$ is part of a block of 0 of length at least $m + 1$, then we are done. Otherwise, the above lemma shows that for each $x \in G_{r'} \cap G_{r''}$, there is a sufficiently large $r$, computable from $r'$ such that for all sufficiently large $n$, $(\mathcal{G}_r(x))_{i_r}$ becomes fixed – this is defined to be $\phi(x)[0]$. Let $r_1$ be greater than $n$ and $r$. Since $R_{S_r}(x)$, $R_{S_r}(x)$ is derived from $R_{S_r}(x,n)$ (via the construction of consistent minimal sub-society), we have that all $\mathcal{G}_R(x,r_1)$ which know some $\mathcal{F}_R(x,r_1)$ have the coordinate $i_r$ fixed with same symbol $(\mathcal{G}_r(x))_{i_r}$. Hence at this $r_1$ we can level of the inductive construction we can compute $\phi(x)[0]$.

Finally we show that $\phi$ is indeed an isomorphism. The map is by construction measurable, and shift-invariant. We only need to show that it is measure-preserving. We use a similar technique as in the original proof. Consider $x \in \mathcal{C}$ specified by fixing $z$ consecutive co-ordinates for some $z$. We show that $P_A(\phi^{-1}(Y)) \geq P_C(Y)$. This holds for all $Y \in \Sigma^\infty_C$. Hence we have that $\phi$ is measure-preserving on the algebra $\Sigma^\infty_C$. This is sufficient, since elements of $\Sigma^\infty_C$ over all $z$ generate the $\sigma$-algebra $\Sigma^\infty_C$.

Let $X = \{ x \in \Sigma^\infty_A \mid x[k \ldots z + k] = c_k c_{k+1} \ldots c_{z} \}$. Since both $A$ and $C$ are stationary and $\phi$ is shift preserving, we can assume that $k = 0$. Consider $x \in T_{C}^{(k+1)} X$, i.e., the symbols in positions $-k - 1$ to $-1$ match those in corresponding places of $C$. Now, consider a cylinder $xa$ for $a \in \Sigma_C$. Clearly, $X = \cup_{a \in C} X_a$. Now we use the assignment lemma on cylinder $Xa$ to argue about the measures. The assignment implies for all $x \in G_{r'} \cap G_{r''}$ there is an $r_1$ such that
\((G_{r'} \cap G'_{r'})^c\) has measure \(\delta_{r'}\) and we can find the assignment for \(\phi(x)[0]\) in the \(r_1\) level of the inductive construction. Note, \(\delta_{r'} \rightarrow 0\) as \(r' \rightarrow \infty\). So,

\[
P_C(x) = \sum_{a \in \Sigma_C} P_C(Xa) \leq \sum_{\bar{x}_1(xa) \in R_{r_1}^{L_{1,r}}(Xa), \beta_{r_1}(Xa)} P_A(\bar{x}_1(Xa), r_1) + \delta_{r'} \leq P_A(\phi^{-1}(X))(1 - \varepsilon_{r_1}) + \delta_{r'},
\]

since the map \(\phi\) respects society and we consider \(\varepsilon_{r_1}\) robust societies in level \(r_1\). Since \(x \in G_s \cap G'_s\) for all \(s \geq r'\), we have \(P_C(X) \leq P_A(\phi^{-1}(X))\).

### 5.6.2 Layerwise Lower Semicomputability of \(\phi\)

In this section, we recapitulate the major steps in the construction of the isomorphic map \(\phi\) and show that it is layerwise lower semicomputable. This yields, as a corollary, that it is defined for every Martin-Löf random sequence \(x \in \mathcal{A}\). We conclude by proving that \(\phi(x) \in \mathcal{C}\) is a Martin-Löf random as well.

We show that there is a layering \(\langle K^A_r \rangle_{r=1}^{\infty}\) of \(\mathcal{A}\) such that the following holds. For every \(x \in K^A_r\), there is a central cylinder \([m_r + 1 \ldots m_r + 1]\) that maps to a central cylinder \([m_r + 1 \ldots m_r + 1]\) such that \(P_A(x[-m_r + 1 \ldots m_r - 1])\) is approximately \(P_C(y[-m_r + 1 \ldots m_r - 1])\).

To see this, note that the layering \(\langle K'_r \cap K''_r \rangle_{r=1}^{\infty}\) of \(\mathcal{A}\), where \(\langle K'_r \rangle_{r=1}^{\infty}\) is the layering of \(\mathcal{A}\) in the Skeleton Lemma and \(\langle K''_r \rangle_{r=1}^{\infty}\) is its layering in the Filler Lemma, has the following property. For every \(r \in \mathbb{N}\) and \(x \in G_r \cap G'_{r'}\), there is a central skeleton of \(x\) of rank \(r\) and length \(L_r\), for which every filler \(F \in \Sigma^{L_r}_{\mathcal{A}}\) obeys the probability bounds in the filler lemma.

Similarly, there is a layering of \(\mathcal{C}\) which has the following property. For every \(r \in \mathbb{N}\) and \(y\) in the \(r\)th layer, there is a central skeleton of \(y\) of rank \(r\) and length \(L_r\), for which every filler \(G \in \Sigma^{L_r}_{\mathcal{C}}\) obeys the probability bounds in the filler lemma.

Then we create a bipartite graph among the equivalence classes \(\bar{F}_n\) and \(\bar{G}_n\) of fillers in \(\Sigma^{L_r}\) and \(\Sigma^{L_{r'}}_{\mathcal{C}}\), and build an \(\epsilon_n\)-robust minimal subsociety according to Algorithm \(\Box\). This is a computable process, since the societies are finite. The assignment lemma yields us a layerwise lower semicomputability of the central co-ordinate \(\phi(x)[0]\).

Let \(T_A\) and \(T_C\) be the shifts associated with \(\mathcal{A}\) and \(\mathcal{C}\), respectively. If \(x\) is Martin-Löf random in \(\mathcal{A}\), the computability and measure-preservation of \(T\) ensure that \(T^ix, i \in \mathbb{Z}\) is also Martin-Löf random in \(\mathcal{A}\). Hence for all large enough ranks \(r'\), \(T^ix \in K^A_r\). Noting that \(x[i] = (T^i x)[0]\) and that \(\phi\) is a factor map, we see that

\[
(\phi \circ T^i_A(x))[0] = (T^i_C \circ \phi(x))[0] = (\phi(x))[i],
\]

we see that all co-ordinates \(\phi(x)[-m + 1 \ldots m + 1]\) will be fixed for all large enough ranks \(K^A_r\). This is an iteration over a layerwise lower semicomputable function, hence is layerwise lower semicomputable.

Hence the maps \(\phi\) and \(\phi^{-1}\) thus constructed are layerwise lower semicomputable and can be computed for all Martin-Löf random points.
Lemma 5.11. Let \( t_A : \Sigma^\infty \rightarrow [0, \infty] \) be a layerwise \( P_A \)-integrable test. Then \( t'_C = t_A \circ \phi^{-1} \) is a layerwise \( P_C \)-integrable test. Conversely, if \( t_C : \Sigma^\infty \rightarrow [0, \infty] \) be a layerwise \( P_C \)-integrable test. Then \( t'_A = \phi \circ t_A \) is a layerwise \( P_A \)-integrable test.

Proof. Let \( y \in C \) be such that \( \phi^{-1}(y) \in A \) is defined. Consider the function \( t'_C(y) = t_A(\phi^{-1}y) \) is layerwise lower semicomputable. Also, \( \int t'_C dP_C = \int t_A \circ \phi^{-1} dP_A \), since \( \phi \) is a measure-preserving isomorphism. Hence \( \int t'_C dP_C \) is finite. Thus \( t'_C \) is a layerwise \( P_C \)-integrable test.

The proof in the converse direction is similar. \( \square \)

Corollary 5.12. \( x \in A \) is Martin-Löf random if and only if \( \phi(x) \in C \) is Martin-Löf random, and \( y \in C \) is Martin-Löf random if and only if \( \phi^{-1}(y) \in A \) is Martin-Löf random.

Proof. Let \( t_A, t'_A, t_C \) and \( t'_C \) be as in the previous lemma. If \( t_A(\phi^{-1}(y)) = \infty \), then \( t'_C(y) = \infty \) implying that \( y \) is not Martin-Löf random in \( C \).

Conversely, by a similar argument, we see that for \( x \in A \) such that \( \phi(x) \in C \) is defined, if \( \phi(x) \) is not Martin-Löf random in \( C \), then \( x \) is not Martin-Löf random in \( A \).

\( \square \)

6 Computable isomorphisms

In this section, we show that only trivial computable isomorphisms exist between two computable dynamical systems. This partly justifies layerwise lower semicomputability as a notion of appropriate power for constructing the isomorphism between the systems.

Recall that a homeomorphism is a continuous bijection whose inverse is also continuous.

Lemma 6.1. Suppose \((X, B, \mu, T)\) and \((Y, C, \nu, S)\) be two Bernoulli systems with the same entropy. If \( \phi : X \rightarrow Y \) is a measure-preserving homeomorphism, then \( \mu \) and \( \nu \) are permutations of each other.

Proof. Let \( \alpha = \{A_1, A_2, \ldots, A_n\} \) be a time-0 partition of \( X \) which generates \( F \). Since \( \phi \) is measure-preserving, we have that for \( \nu(\phi(A_i)) = \mu(A_i) \) for all \( i = 1, 2, \ldots, n \). Since \( \phi^{-1}(B_i \cap B_j) = \phi^{-1}(B_i) \cap \phi^{-1}(B_j) = A_m \cap A_n \) which has \( \mu \) measure 0, we know that the sets \( B_1, \ldots, B_n \) are almost disjoint. Since \( \alpha \) covers \( X \) except possibly for a \( \mu \)-measure 0 set, we know that \( \phi\alpha \) covers \( Y \) except possibly for a \( \nu \)-measure 0 set.

Since \( \phi \) and \( \phi^{-1} \) are both continuous, we have that the image of a cylinder \( A_i \) in the time-0 in \( \alpha \) is a cylinder \( B_i \) in the time-0 partition of \( Y \). Also, for any natural numbers \( m \) and \( n \), we have:

\[
S^{-m}B_i \cap S^{-n}B_j = S^{-m}\phi A_i \cap S^{-n}\phi A_j = \phi T^{-m}A_i \cap \phi T^{-n}A_j = \phi(T^{-m}A_i \cap T^{-n}A_j),
\]

where the last equality follows from the fact that \( \phi \) is a bijection. From this, we conclude that \( \phi\alpha \) is a generator for \((Y, C, \nu, S)\). Thus \((X, B, \mu, T)\) and \((Y, C, \nu, S)\) are generated by generators of the same cardinality.

The entropy for a finite alphabet is a symmetric, strictly convex function. Hence for two probability distributions of dimension \( n \), their entropies are the same only if they are permutations of each other. This concludes our proof. \( \square \)
Since total computable functions are continuous, we have the following corollary.

**Corollary 6.2.** Suppose $(X, \mathcal{B}, \mu, T)$ and $(Y, \mathcal{C}, \nu, S)$ be two Bernoulli systems with the same entropy. If $\phi : X \to Y$ is a measure-preserving computable isomorphism, then $\mu$ and $\nu$ are permutations of each other.

7 Comparison of the results

Ornstein showed that a process satisfying a weaker condition, viz. a finitely determined system with entropy $H$ is isomorphic to some Bernoulli process with entropy $H$. Thus elements of a much broader class of processes are isomorphic to Bernoulli systems of equal entropy, the latter being intuitively the most random systems possible. Several “deterministic” dynamical systems have been shown to be finitely determined (for a survey, see Ornstein[20]), leading to the interpretation that all such systems are, intuitively, encodings of the most random possible systems. However, to demonstrate this, we need isomorphic maps which are termed stationary codes. [27] Rudolph has proved a characterization of systems finitarily isomorphic to each other [25], showing that if we restrict our codes to finitary codes, there are weakly Bernoulli systems and finitely determined systems which cannot be isomorphic to any Bernoulli system with the same entropy.

We show that computable mixing Markov systems of equal entropy have a layerwise lower semicomputable isomorphism. Thus the targets of our isomorphisms are not intuitively as random as that of the Ornstein construction. However, our code has a stronger computability property than Ornstein’s original construction and the maps in Rudolph’s characterization of finitary isomorphism.

Rudolph’s characterization of systems finitarily isomorphic to Bernoulli systems uses the notion of conditional block independence. We leave open whether there is a similar characterization of computable systems which are layerwise isomorphic to a computable mixing Markov system.

Acknowledgments

The authors thank Mathieu Hoyrup for valuable discussions and anonymous referees for their suggestions.

References

[1] Laurent Bienvenu, Adam R. Day, Mathieu Hoyrup, Ilya Mezhirov, and Alexander Shen. A constructive version of birkhoff’s ergodic theorem for martin-löf random points. *Inf. Comput.*, 210:21–30, 2012.

[2] Patrick Billingsley. *Ergodic Theory and Information*. John Wiley and Sons, New York, 1965.
[3] M. Braverman. On the complexity of real functions. In Proceedings of the forty sixth IEEE Annual Symposium on the Foundations of Computer Science, pages 155–164, 2005.

[4] P. Gács. Uniform test of algorithmic randomness over a general space. Theoretical Computer Science, 341:91–137, 2005.

[5] P. Gács, Mathieu Hoyrup, and Cristóbal Rojas. Randomness on computable probability spaces - a dynamical point of view. Theory of Computing Systems, 48(3):465–485, April 2011.

[6] Stefano Galatolo, Mathieu Hoyrup, and Cristóbal Rojas. A constructive Borel-Cantelli lemma, constructing orbits with required statistical properties. Theoretical Computer Science, 410(21-23):2207–2222, 2009.

[7] Michael Hochman. Upcrossing inequalities for stationary sequences and applications. Annals of Probability, 37(6):2135–2149, 2009.

[8] M. Hoyrup and C. Rojas. An application of martin-löf randomness to effective probability theory. In Proceedings of the 5th conference on Computability in Europe, 2009.

[9] M. Hoyrup and C. Rojas. Applications of effective probability theory to Martin-Löf randomness. In Proceedings of the 36th International Colloquium on Automata, Logic and Programming, volume 5555/2009, pages 549–561. Springer, 2009.

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

[11] Mathieu Hoyrup. The dimension of ergodic random sequences. In Symposium on Theoretical Aspects of Computer Science, pages 567–576, 2012.

[12] M. Keane and M. Smorodinsky. The finitary isomorphism theorem for Markov shifts. Bull. Am. Math. Soc., New Ser., 1:436–438, 1979.

[13] M. Keane and Meier Smorodinsky. A class of finitary codes. Israel Journal of Mathematics, 6(3-4):352–371, 1977.

[14] Michael Keane and Meier Smorodinsky. Bernoulli schemes of the same entropy are finitarily isomorphic. Annals of Mathematics, 109(2):397–406, 1979.

[15] A. N Kolmogorov. New metric invariants of transitive dynamical systems and automorphisms. Doklady Akad. Nauk SSSR, 93:763–766, 1953.

[16] A. N. Kolmogorov. A new invariant for transitive dynamical systems. Doklady Akad. Nauk SSSR, 119:861–864, 1958.

27
[17] P. Martin-Löf. The definition of random sequences. *Information and Control*, 9(6):602–619, 1966.

[18] S. Nandakumar. An effective ergodic theorem and some applications. In *Proceedings of the 40th Annual Symposium on the Theory of Computing*, pages 39–44, 2008.

[19] Donald Ornstein. Bernoulli shifts with the same entropy are isomorphic. *Advances in Mathematics*, 4(3):337 – 352, 1970.

[20] Donald Ornstein. Newton’s laws and coin tossing. *Notices of the American Mathematical Society*, 60(4):450–459, 2013.

[21] Donald Ornstein and Benjamin Weiss. Entropy is the only finitely observable invariant. *Journal of Modern Dynamics*, 1(1):93–105, 2007.

[22] D.S. Ornstein. Imbedding bernoulli shifts in flows. In *Contributions to Ergodic Theory and Probability*, volume 160 of *Lecture Notes in Mathematics*, pages 178–218. Springer Berlin Heidelberg, 1970.

[23] Cornfield I. P., Fomin S. V., and Sinai Ya. G. *Ergodic Theory*. Springer Verlag, 1982.

[24] Karl Petersen. *Ergodic Theory*. Cambridge Studies in Advanced Mathematics, 1989.

[25] Daniel J. Rudolph. A characterization of those processes finitarily isomorphic to a Bernoulli shift. *Ergodic Theory and Dynamical Systems: Proceedings, special year, Maryland 1979-1980. Edited by A. Katok*, pages 1–64, 1981.

[26] A. Kh. Shen’. The frequency approach to the definition of a random sequence. *Semiotika i Informatika*, 18:14–42, 1982. (In Russian.).

[27] Paul Shields. The interactions between ergodic theory and information theory. *IEEE Transactions on Information Theory*, 44(6):2079–2093, 1998.

[28] Ya. G. Sinai. The notion of entropy of a dynamical system. *Doklady Akad. Nauk SSSR*, 125:768–771, 1959.

[29] M. van Lambalgen. *Random Sequences*. Academish Proefschri’t, Amsterdam, 1987.

[30] V. Vovk. The law of the iterated logarithm for random kolmogorov, or chaotic, sequences. *Theory of Probability and Applications*, 32:413–425, 1987.

[31] V. V. V’yugin. Ergodic theorems for individual random sequences. *Theoretical Computer Science*, 207:343–361, 1998.