Randomness and the ergodic decomposition

Mathieu Hoyrup

Abstract

The interaction between algorithmic randomness and ergodic theory is a rich field of investigation. In this paper we study the particular case of the ergodic decomposition. We give several positive partial answers, leaving the general problem open. We shortly illustrate how the effectivity of the ergodic decomposition allows one to easily extend results from the ergodic case to the non-ergodic one (namely Poincaré recurrence theorem). We also show that in some cases the ergodic measures can be computed from the typical realizations of the process.

1 Introduction

The goal of the paper is to study the interaction between the theory of algorithmic randomness, started by Martin-Löf [ML66], and ergodic theory (i.e. restricting to shift-invariant measures). The first results in this direction were obtained by V'yugin [V'y98], who proved that Birkhoff's ergodic theorem and a weak form of Shannon-McMillan-Breiman theorem hold for each Martin-Löf random sequence. Recently several improvements of the first result have been achieved [Nan08, HR09a, BDMS10].

A classical result from ergodic theory, called the *ergodic decomposition*, states that given a stationary process, almost every realization is actually a typical realization of an *ergodic* process. The full process can be decomposed as the combination of a collection of ergodic processes. It is natural to ask the question whether every Martin-Löf random sequence (with respect to the stationary measure) statistically induces an ergodic measure, and if the sequence is Martin-Löf random with respect to it.

We give three orthogonal cases in which we can give a positive answer: (i) when the decomposition of the measure is computable, (ii) when the decomposition of the measure is supported on an effective compact class of ergodic measures, (iii) when the decomposition of the measure is finite. Observe that the three cases are mutually incomparable. We leave the general problem open.

As a side result, we give sufficient conditions to infer the statistics of the system from the observation; formally we give a sufficient condition on an ergodic measure to be computable relative to its random elements.

In Section 2 we give the necessary background on computability and randomness. In Section 3 we develop results about randomness and combinations of measures that will be applied in the sequel, but are of independent interest (outside ergodic theory). We start Section 4 with a reminder on the ergodic decomposition and then we present our results relating it to randomness. We finish in Section 5 by a side result on the inference of ergodic measures from their random elements.

2 Preliminaries

We assume familiarity with algorithmic randomness and computability theory.

All the results stated in this paper hold on effectively compact computable metric spaces X and for computable maps $T : X \to X$ (as defined in computable analysis, see [Wei00]), but for the sake of simplicity we formulate them only on the Cantor space $X = \{0,1\}^{\mathbb{N}}$ and for the shift transformation $T : X \to X$ defined by $T(x_0x_1x_2\dots) = x_1x_2x_3\dots$. The Cantor space is endowed with the product topology, generated by the cylinders $[w]$, $w \in \{0,1\}^*$. Implicitly, measures are probability measures defined on the Borel σ-algebra, and ergodic measures are stationary (i.e., shift-invariant) ergodic Borel probability measures. The set $\mathcal{P}(X)$ of probability measures over X is endowed with the weak* topology, given by the notion of weak convergence: measures P_n converge to P if for every $w \in \{0,1\}^*$, $P_n[w] \to P[w]$.

A name for a real number r is an infinite binary sequence encoding, in some canonical effective way, a sequence of rational numbers q_n such that $|q_n - r| < 2^{-n}$ for all n. A name for a probability measure P is the interleaving, in some canonical effective way, of names for the numbers $P[w]$, $w \in \{0,1\}^*$. A computable probability measure is a measure admitting a computable name: in other words, the numbers $P[w]$ are uniformly computable.

Let X, Y be any spaces among $\{0,1\}^{\mathbb{N}}$, $\mathbb{R}$ and $\mathcal{P}(X)$. A function $f : X \to Y$ is computable if there is an oracle machine that, given a name of $x \in X$ as an oracle, outputs a name of $f(x)$ (the computation never halts). Computable functions are continuous. f is computable on a set $A \subseteq X$ if the same holds for all $x \in A$ (nothing is required to the machine when $x \notin A$). An object y is ***computable relative to*** an object x if the function $x \mapsto y$ is computable on $\{x\}$, i.e. if there is an oracle machine that on any name of x as oracle, produces a name of y.

An open subset U of the Cantor space is effective if there is a (partial) computable function $\varphi : \mathbb{N} \to \{0,1\}^*$ such that $U = \bigcup_{n\in\mathbb{N}}[\varphi(n)]$. An effective compact set is the complement of an effective open set. Let $K \subseteq X$ be an effective compact set and $f : K \to Y$ a function computable on K.

Fact 1 (Folklore). $f(K)$ is an effective compact set.

Fact 2 (Folklore). If f is moreover one-to-one then $f^{-1} : f(K) \to K$ is computable on $f(K)$.

Proof. For the sake of clarity, we denote f^{-1} by g. Here we use the following characterization: g is computable on $f(K)$ if and only if the preimage of any effective open set U is an effective open set on $f(K)$, uniformly, i.e. there is an effective open set V such that $g^{-1}(U) = f(K) \cap V$.

Let U be an effective open set. We have to prove The set $C := K \setminus U$ is an effective compact set. $g^{-1}(U) = g^{-1}(K \setminus C) = g^{-1}(K) \setminus g^{-1}(C) = f(K) \setminus f(C)$. As C is an effective compact set, its complement is an effective open set (everything is uniform here). $\square$

The product of two computable metric spaces has a natural structure of computable metric space.

Fact 3 (Folklore). If $K \subseteq X$ is an effective compact set and $f : K \times Y \to \overline{\mathbb{R}}$ is lower semi-computable, then the function $g : Y \to \overline{\mathbb{R}}$ defined by $g(y) = \inf_{x\in K} f(x, y)$ is lower semi-computable.

Proof. Let us prove that $g^{-1}(q,+\infty] = \{y : K \times \{y\} \subseteq f^{-1}(q,+\infty]\}$ is an effective open set, uniformly in q. Let q be some fixed rational number. The effective open set $U_q = f^{-1}(q,+\infty]$ can be expressed as an effective union of product balls $U_q = \bigcup_{i\in\mathbb{N}} B_i^X \times B_i^Y$. The set $E_q = \{(i_1,\dots,i_k) : K \subseteq B_{i_1}^X \cup \dots \cup B_{i_k}^X\}$ is r.e. and it is easy to prove that $g^{-1}(q,+\infty] = \bigcup_{(i_1,\dots,i_k)\in E_q} B_{i_1}^Y \cap \dots \cap B_{i_k}^Y$, which is an effective open set. The argument is uniform in q. □

If f, g are real-valued functions, $f \overset{*}{<} g$ means that there exists $c \geq 0$ such that $f \leq cg$. $f \overset{*}{=} g$ means that $f \overset{*}{<} g$ and $g \overset{*}{<} f$.

2.1 Effective randomness

Martin-Löf [ML66] was the first one to define a sound individual notion of random infinite binary sequence. He developed his theory for any computable probability measure on the Cantor space. This theory was then extended to non-computable measures by Levin [Lev73], and later by [Gác05, HR09b] on general spaces ([HW03] was an extension to topological spaces, but for computable measures).

We will use the most general theory: we will be interested in randomness on the Cantor space and on the space of Borel probability measures over the Cantor space, for arbitrary (i.e. not necessarily computable) probability measures. In particular, we will use the notion of uniform test of randomness, introduced by Levin [Lev73] and further developed in [Gác05, Gác08, HR09b].

On a computable metric space X endowed with a probability measure P, there is a set $\mathcal{R}_P$ of P-random elements satisfying $P(\mathcal{R}_P) = 1$, together with a canonical decomposition (coming from the universal P-test) $\mathcal{R}_P = \bigcup_n \mathcal{R}_P^n$ where $\mathcal{R}_P^n$ are uniformly effective compact sets relative to P, $\mathcal{R}_P^n \subseteq \mathcal{R}_P^{n+1}$ and $P(\mathcal{R}_P^n) > 1-2^{-n}$. The sets $X \setminus \mathcal{R}_P^n$ constitute a universal Martin-Löf test. A P-test is a function $t : X \to [0,+\infty]$ which is lower semi-computable relative to P, such that $\int t \, \mathrm{d}P \leq 1$.

A function $f : X \to Y$ is ***P-layerwise computable*** if there is an oracle machine that, given n as input and a name of $x \in \mathcal{R}_P^n$ as an oracle, outputs a name of $f(x)$ (the computation never halts). Nothing is required to the machine when x is not P-random. When f is P-layerwise computable, for every P-random x, $f(x)$ is computable relative to x in a way that is not fully uniform, but uniform on each set $\mathcal{R}_P^n$.

Such a machine can be thought of as a probabilistic algorithm, but here the randomness is not part of the algorithm but of the input. Formally, it is the same notion, but usually, "succeeding with high probability" means that if we run the program on a given input several times, independently, it will succeed most of the times; here, the algorithm is deterministic and it will succeed on most inputs.

Lemma 2.1. *Let P be a computable measure, $f : X \to Y$ a P-layerwise computable function and $Q = f_*P$.*

1. *Q is computable and $f : \mathcal{R}_P \to \mathcal{R}_Q$ is onto.*

2. *If $f : X \to Y$ is moreover one-to-one then $f : \mathcal{R}_P \to \mathcal{R}_Q$ is one-to-one and f^{-1} is Q-layerwise computable.*

Proof. We only prove that f^{-1} is Q-layerwise computable, the other statements are proved in [HR09a]. There is $c \in \mathbb{N}$ such that $\mathcal{R}_Q^n \subseteq f(\mathcal{R}_P^{n+c})$ for all n. Let $n \in \mathbb{N}$. $f : \mathcal{R}_P^{n+c} \to Y$ is one-to-one and computable so by Fact 2, $f^{-1} : f(\mathcal{R}_P^{n+c}) \to X$ is computable. As $\mathcal{R}_Q^n \subseteq f(\mathcal{R}_P^{n+c})$, $f^{-1} : \mathcal{R}_Q^n \to X$ is computable. This is uniform in n. □

3 Randomness and continuous combination of measures

The material developed here will be used to investigate the algorithmic content of the ergodic decomposition.

Given a *countable* class of probability measures P_i and real numbers $\alpha_i \in [0, 1]$ such that $\sum_i \alpha_i = 1$, the convex combination $P = \sum_i \alpha_i P_i$ is again a probability measure. This can be generalized to *continuous* classes of measures, as we briefly remind now.

Let m be a probability measure over $\mathcal{P}(X)$. The set function P defined by $P(A) = \int Q(A)\,\mathrm{d}m(Q)$ for measurable sets A is a probability measure over X, called the barycenter of m. It satisfies $\int f\,\mathrm{d}P = \int(\int f\,\mathrm{d}Q)\,\mathrm{d}m(Q)$ for $f \in L^1(X, P)$. When m is computable, so is P. We can think of P as the measure describing the following process: first pick some measure Q at random according to m; then run the process with distribution Q.

Probabilistically, picking a sequence according to P or decomposing into these two steps are equivalent. We are interested in whether the algorithmic theory of randomness fits well with this intuition: are the P-random sequences the same as the sequences that are Q-random for some m-random Q?

Remark 3.1. Let $f : X \to [0, +\infty]$ be a lower semi-computable function. Let $F : \mathcal{P}(X) \to [0, +\infty]$ be defined by $F(Q) = \int f\,\mathrm{d}Q$. F is lower semi-computable and $\int F\,\mathrm{d}m = \int f\,\mathrm{d}P$. As a result, F is a m-test if and only if f is a P-test.

Theorem 3.1. *Let $m \in \mathcal{P}(\mathcal{P}(X))$ be computable, and P be the barycenter of m. For $x \in X$, the following are equivalent:*

1. *x is P-random,*
2. *x is Q-random for some m-random Q.*

In other words,

$$\mathcal{R}_P = \bigcup_{Q \in \mathcal{R}_m} \mathcal{R}_Q.$$

Proof.

Let $f(x) = \inf_Q t_m(Q).t_Q(x)$. f is lower semi-computable by Fact 3 ($\mathcal{P}(X)$ is effectively compact). As $\int f\,\mathrm{d}P = \int(\int f\,\mathrm{d}Q)\,\mathrm{d}m(Q) \leq \int t_m(Q)(\int t_Q\,\mathrm{d}Q)\,\mathrm{d}m(Q) \leq 1$, f is a P-test, so if x is P-random then it is Q-random for some m-random measure Q.

Conversely, let $T_P(Q) = \int t_P\,\mathrm{d}Q$ where t_P is a universal P-test. By Remark 3.1, T_P is a m-test so if Q is m-random then $T_P(Q) < \infty$, so t_P is a (multiple of a) Q-test. As a result, $\mathcal{R}_Q \subseteq \mathcal{R}_P$.

□

4 Randomness and ergodic decomposition

4.1 Background from ergodic theory

A sequence $x \in \{0,1\}^{\mathbb{N}}$ is ***generic*** if for each $w \in \{0,1\}^*$, the frequency of occurrences of w in x converges. If x is generic, we denote by Q_x the set function which maps each cylinder $[w]$ to the limit frequency of occurrences of w in x. Q_x extends to a probability measure over the Cantor space, which we also denote by Q_x. If x is generic then Q_x is stationary, i.e. $Q_x(A) = Q_x(T^{-1}(A))$ for every measurable set A. Birkhoff's ergodic theorem states that given a stationary measure P, P-almost every x is generic. A stationary measure P is ergodic if the only invariant sets have measure 0 or 1. Formally, if $T^{-1}(A) = A$ then $P(A) = 0$ or 1, for every measurable set A. If P is stationary ergodic then $Q_x = P$ for P-almost every x.

The ergodic decomposition theorem states that given a stationary probability measure P, the measure Q_x is ergodic for P-almost every x. There are mainly two proofs of this fact. One of them uses Choquet theorem from convex analysis: the set of stationary probability measures is a convex compact set whose extreme points are exactly the ergodic measures. Then any point in that set, i.e. any invariant measure, can be expressed as a barycenter over the ergodic measures. More precisely, for any invariant measure P there is a unique probability measure m_P over $\mathcal{P}(X)$ which gives full weight to the ergodic measures, and such that $P(A) = \int Q(A)\,\mathrm{d}m_P(Q)$ for every Borel set A. We will call m_P the ***Choquet measure*** associated to P.

4.2 Randomness and ergodic theorems

An algorithmic version of Birkhoff's ergodic theorem was eventually proved by V'yugin [V'y97]: given a shift-invariant probability measure P, every P-random sequence is generic, and if P is moreover ergodic then $Q_x = P$ for every P-random sequence x (it was proved for computable measures, but it still works for non-computable measures). The proof was not immediate to obtain from the classical proof of Birkhoff's theorem, which is in a sense not constructive. In this paper we are interested in an algorithmic version of the ergodic decomposition theorem, which again cannot be proved directly.

More precisely, given a stationary measure P, we are interested in the following questions:

- if x is P-random, is Q_x ergodic?
- if x is P-random, is x also Q_x-random?
- if x is P-random, is Q_x an m_P-random measure?
- does any converse implication hold?

We give positive partial answers to these questions, leaving the general problem open. We will use the following lemmas (the first one was proved in [V'y97]).

Lemma 4.1. *Let P be an ergodic stationary probability measure. For every $x \in \mathcal{R}_P$, $Q_x = P$.*

Lemma 4.2. *Let P be a stationary probability measure and m_P the associated Choquet measure. Every m_P-random measure is ergodic and stationary.*

Proof. It is known that the set of ergodic stationary measure is a G_δ-set. It is moreover effective, i.e. it is an intersection of effective open sets. As it has m_P-measure one, it contains $\mathcal{R}_{m_P}$. $\square$

4.3 First answer: effective decomposition

A stationary probability measure P is always computable relative to its associated Choquet measure m_P. The converse does not always hold (see Section 4.4 for a counter-example).

Definition 4.1. A stationary probability measure P is ***effectively decomposable*** if its Choquet measure is computable relative to P.

4.3.1 When P is computable

As an application of Theorem 3.1, we directly get a result when P is computable and effectively decomposable (i.e. when $m := m_P$ is computable).

Corollary 4.1. *Let P be a computable stationary probability measure that is effectively decomposable. For $x \in X$, the following are equivalent:*

1. *x is P-random,*
2. *x is Q-random for some m-random Q.*

In other words, the following are equivalent:

1. *x is P-random,*
2. *x is generic, Q_x-random and Q_x is m-random.*

We also have the following characterization.

Theorem 4.1. *Let P be a computable stationary probability measure. The following are equivalent.*

1. *P is effectively decomposable,*
2. *the function $X \to \mathcal{P}(X), x \mapsto Q_x$ is P-layerwise computable.*

Proof. $1 \Rightarrow 2$. In any probability space (Y, μ) with random elements $\mathcal{R}_\mu = \bigcup_n \mathcal{R}_\mu^n$, we define $d_\mu(y) = \min\{n : y \in \mathcal{R}_\mu^n\}$ ($d_\mu(y) = +\infty$ if y is not μ-random). $d : \mathcal{P}(Y) \times Y \to [0, +\infty]$ which maps (μ, y) to $d_\mu(y)$ is lower semi-computable.

Let $C_n = \{(Q, x) : d_m(Q) \leq n \text{ and } d_Q(x) \leq n\}$. The second projection $\pi_2 : \bigcup_n C_n \to X$ is one-to-one. Indeed, if $(Q_i, x_i) \in \bigcup_n C_n$, $i = 1, 2$ and $\pi_2(Q_1, x_1) = \pi_2(Q_2, x_2)$ then (i) $x_1 = x_2$, (ii) Q_1, Q_2 are m-random hence ergodic, (iii) x_i is Q_i-random so $Q_{x_i} = Q_i$; as a result, $Q_1 = Q_{x_1} = Q_{x_2} = Q_2$. C_n is effectively compact so π_2^{-1} is computable on each $\pi_2(C_n)$ (uniformly in n) by Fact 2.

We know from the proof of Theorem 3.1 that there exists a constant c such that for all n and all $x \in \mathcal{R}_P^n$, $(Q_x, x) \in C_{n+c}$, hence $\mathcal{R}_P^n \subseteq \pi_2(C_{n+c})$. It implies that π_2^{-1} is computable on each $\mathcal{R}_P^n$, uniformly in n, i.e. π_2^{-1} is P-layerwise computable. Finally, $\pi_1 \circ \pi_2^{-1}$, which maps $x \in \mathcal{R}_P$ to Q_x is P-layerwise computable.

$2 \Rightarrow 1$. Conversely, if $\psi : x \mapsto Q_x$ is P-layerwise computable, then $m = \psi_* P$ is the push-forward of P under ψ, so it is computable by Lemma 2.1, item 1. □

Remark 4.1. For $f \in L^1(X, P)$, we denote by f^* the limit of the Birkhoff averages of f. One can also prove that if P is computable then P is effectively decomposable if and only if the function

$$\begin{array}{rcl} L^1(X,P) & \to & L^1(X,P) \\ f & \mapsto & f^* \end{array}$$

is computable.

Remark 4.2. We can use Lemma 2.1, item 2 to prove that $\pi_2^{-1} : x \mapsto (Q_x, x)$ is P-layerwise computable.

It can be proved that there is a probability measure μ over $\mathcal{P}(X) \times X$ such that $\mu(\mathcal{A} \times B) = \int_{Q \in \mathcal{A}} Q(B) \, \mathrm{d}m(Q)$ for all measurable sets $\mathcal{A} \subseteq \mathcal{P}(X), B \subseteq X$. μ satisfies $\int f \, \mathrm{d}\mu = \int\int f(Q, .) \, \mathrm{d}Q \, \mathrm{d}m$ for all $f \in L^1(\mathcal{P}(X) \times X, \mu)$. It can be proved that μ is computable and that (Q, x) is μ-random if and only if Q is m-random and x is Q-random.

The projection $\pi_2 : \mathcal{P}(X) \times X \to X$ is computable and one-to-one on $\mathcal{R}_\mu$. As $(\pi_2)_* \mu = P$, π_2^{-1} is P-layerwise computable by Lemma 2.1, item 2.

4.3.2 When P is not computable.

If P is not computable, but still effectively decomposable, one implication in Corollary 4.1 remains.

Theorem 4.2. *Let P be a stationary probability measure that is effectively decomposable. For every P-random x, Q_x is m_P-random, hence ergodic, and x is Q_x-random.*

The converse implication does not hold in general, as the following counter-example proves. Let $p_1, p_2 \in (0, 1)$ be two different real numbers and P_1, P_2 the Bernoulli measures with parameters p_1, p_2 respectively. Let x be P_1-random and let $\alpha \in (0, 1)$ be the real number whose binary expansion is $0.x$. Let $P = \alpha P_1 + (1 - \alpha) P_2$. P is effectively decomposable, as $m_P = \alpha \delta_{P_1} + (1 - \alpha) \delta_{P_2}$ is computable relative to P. As x is computable relative to P, x is not P-random but x is P_1-random and P_1 is m_P-random.

The effectivity of the ergodic decomposition enables one to extend results from ergodic systems to non-ergodic ones. Let us illustrate it. It was proved in [BDMS10] that when P is an ergodic measure, every P-random sequence eventually visits every effective compact set of positive measure under shift iterations. When the decomposition is effective, this theorem can be generalized to non-ergodic measures, giving a version of Poincaré recurrence theorem for random sequences.

Corollary 4.2. *Let P be a stationary measure that is effectively decomposable. Let F be an effective compact set such that $P(F) > 0$. Every P-random $x \in F$ falls infinitely often in F under shift iterations.*

Proof. x is Q_x-random and Q_x is ergodic. As all random sequences belong to effective open sets of measure one and $x \in F$, $Q_x(F) > 0$. Hence we can apply the result in [BDMS10] to the ergodic measure Q_x (strictly speaking their result was proved for *computable* ergodic measures, but it can be relativized without difficulty). □

The result actually holds as soon as for every P-random x, Q_x is ergodic and x is Q_x-random.

4.4 V'yugin's example

In [V'y97], V'yugin constructed a computable stationary measure for which the convergence of Birkhoff's average is not effective. We give a simpler construction and show that this measure is not effectively decomposable.

Let M_i be some effective enumeration of the Turing machines. For each i, let $p_i = 2^{-t_i}$ is M_i halts in time t_i, $p_i = 0$ if t_i does not halt. The real numbers p_i are computable uniformly in i (while they are not uniformly computable as rational numbers). Let P_i be the Markovian stationary measure defined by $P_i[0] = P_i[1] = \frac{1}{2}$ and $\frac{P_i[w01]}{P_i[w0]} = \frac{P_i[w10]}{P_i[w1]} = p_i$ for all $w \in \{0,1\}^*$ (the probability of changing between states 0 and 1 is p_i). Let $P = \sum_i 2^{-i} P_i$. P is computable. We now show that $x \mapsto Q_x$ is not P-layerwise computable (which will imply that P is not effectively decomposable by Theorem 4.1). Let $f = \chi_{[1]}$. Let $\alpha = \sum_{i:M_i \text{ halts}} 2^{-i}$. $f^*(x) = 0$ for $x = 0^{\mathbb{N}}$, $f^*(x) = 1$ for $x = 1^{\mathbb{N}}$ and $f^*(x) = \frac{1}{2}$ for P-almost all $x \notin \{0^{\mathbb{N}}, 1^{\mathbb{N}}\}$. By definition of Q_x, $f^*(x) = Q_x[1]$ for every P-random x. If f^* were P-layerwise computable, then $P(f^{*-1}[0, 1/4)) = (1-\alpha)/2$ would be lower semi-computable by basic properties of layerwise computable functions (see [HR09a]).

While P is not effectively decomposable, we can still get a result about random elements.

Proposition 4.1. *For every P-random x, Q_x is ergodic and x is Q_x-random.*

Proof. The decomposition $P = \sum_i 2^{-i} P_i$ is partial in the sense that some P_i are not ergodic (when M_i does not halt). However we can apply Theorem 3.1 to this decomposition: P is the barycenter of the computable measure $m' = \sum_i 2^{-i} \delta_{P_i}$, so every P-random x is random for some P_i. (i) If M_i halts, then P_i is ergodic. (ii) If M_i does not halt then $P_i = \frac{1}{2}(\delta_0 + \delta_1)$ (where δ_0 is the measure concentrated on $0^{\mathbb{N}}$, δ_1 on $1^{\mathbb{N}}$). In turn, P_i, which is non-ergodic is effectively decomposable. Hence as x is P_i-random, $Q_x = \delta_0$ or δ_1 and x is Q_x-random. □

As a result, Corollary 4.2 also holds for the measure P.

4.5 A particular case: effective compact classes of ergodic measures

Proposition 4.2. *Let P be a stationary probability measure. If m_P is supported on an effective compact class of ergodic measures, then P is effectively decomposable.*

Proof. Let $\mathscr{C}$ be an effective compact class of stationary ergodic probability measures. Let $\mathcal{P}(\mathscr{C})$ be the set of probability measures m over $\mathcal{P}(X)$ such that $m(\mathscr{C}) = 1$. $\mathcal{P}(\mathscr{C})$ is an effective compact subset of $\mathcal{P}(X)$: indeed, it is the pre-image of $[1, +\infty]$ under the upper semi-computable function $m \mapsto m(\mathscr{C})$. If $m \in \mathcal{P}(\mathscr{C})$, the barycenter P of m is defined by $P(A) = \int Q(A)\,\mathrm{d}m(Q)$ for every measurable set A. The function ψ which maps m to P is computable. Let $\mathcal{I}_{\mathscr{C}}$ be the class of invariant measures that are barycenters of $\mathscr{C}$, i.e. the image of $\mathcal{P}(\mathscr{C})$ under ψ: $\mathcal{I}_{\mathscr{C}}$ is an effective compact class too. By existence and uniqueness of the ergodic decomposition, $\psi : \mathcal{P}(\mathscr{C}) \to \mathcal{I}_{\mathscr{C}}$ is onto and one-to-one; as it is computable and $\mathcal{P}(\mathscr{C})$ is an effective compact set, its inverse is also computable by Fact 2. $\square$

Example 1. Let m be a probability measure over the real interval $[0, 1]$. Pick a real number p at random according to m, and then generate an infinite sequence of $0, 1$ tossing a coin with probability of heads p. As an application of the preceding proposition, we get that the function which maps a random sequence generated by the process to the number p that was picked is P-layerwise computable: it can be computed from the observed outcomes with high probability.

We also learn that the algorithmic theory of randomness fits well with this example: obviously, we expect a random sequence for the whole process to be random for some Bernoulli measure B_p, which is not immediate.

In Section 2.1, we define P-layerwise computable function when P is a computable probability measure. This can be extended straightforwardly to any effective compact class of measures $\mathscr{C}$. The universal $\mathscr{C}$-test induces a decomposition $\bigcup_n \mathcal{R}_n^{\mathscr{C}}$ of the sequences that are random for some measure in $\mathscr{C}$. A function $f : X \to Y$ is $\mathscr{C}$-layerwise computable if it is computable on each $\mathcal{R}_n^{\mathscr{C}}$, uniformly in n. It means that one can compute $f(x)$ if x is random for some measure $P \in \mathscr{C}$, with probability of error bounded by 2^{-n}, whatever P is (as long as it is in $\mathscr{C}$), and for any n.

From Theorem 4.2 we know that for every $P \in \mathcal{I}_{\mathscr{C}}$ and every $x \in \mathcal{R}_P$, Q_x is m-random, hence ergodic and x is Q_x-random. We also prove a quantitative version of this fact. We recall that if $\mathcal{A}$ is an effective compact class of measures, $t_{\mathcal{A}} := \inf_{P \in \mathcal{A}} t_P$ is a universal $\mathcal{A}$-test, i.e. (i) it is lower semi-computable, (ii) $\int t_{\mathcal{A}}\,\mathrm{d}P \leq 1$ for every $P \in \mathcal{A}$ and (iii) $t_{\mathcal{A}}$ multiplicatively dominates every function satisfying (i) and (ii) (see [Gác08] for more details about such class tests). We will consider the class tests $t_{\mathscr{C}}$ and $t_{\mathcal{I}_{\mathscr{C}}}$.

Theorem 4.3. *Let $\mathscr{C}$ be an effective compact class of stationary ergodic probability measures. One has:*

1. $t_{\mathscr{C}}(x) \stackrel{*}{=} t_{\mathcal{I}_{\mathscr{C}}}(x)$
2. *The function $x \mapsto Q_x$ is $\mathcal{I}_{\mathscr{C}}$-layerwise computable and $\mathscr{C}$-layerwise computable.*

Proof.

1. Of course, $t_{\mathcal{I}_{\mathscr{C}}} \stackrel{*}{<} t_{\mathscr{C}}$ as $\mathscr{C} \subseteq \mathcal{I}_{\mathscr{C}}$. Conversely, the $P \in \mathcal{I}_{\mathscr{C}}$: $\int t_{\mathscr{C}}\,\mathrm{d}P = \int (\int t_{\mathscr{C}}\,\mathrm{d}Q)\,\mathrm{d}m(Q) \leq 1$ as m is supported on measures in $Q \in \mathscr{C}$, and $\int t_{\mathscr{C}}\,\mathrm{d}Q \leq 1$ for such measures. As a result, $t_{\mathscr{C}}$ is a $\mathcal{I}_{\mathscr{C}}$-test, so $t_{\mathscr{C}} \stackrel{*}{<} t_{\mathcal{I}_{\mathscr{C}}}$.

2. The proof is similar to the proof of Theorem 4.1. As $t_{\mathscr{C}}(x) = \inf_{Q \in \mathscr{C}} t_Q(x)$, if $x \in \mathcal{R}_{\mathscr{C}}^n$ then $Q_x \in \mathcal{R}_m^{n+c}$. Again, $\pi_2 : \mathcal{R}_m^{n+c} \times \mathcal{R}_{\mathscr{C}}^n \to \mathcal{R}_{\mathscr{C}}^n$ is computable and bijective so its inverse is computable and maps x to (Q_x, x). Hence $\pi_1 \circ \pi_2^{-1}$ is computable on $\mathcal{R}_{\mathscr{C}}^n$, uniformly in n, i.e. it is $\mathscr{C}$-layerwise computable. As $t_{\mathscr{C}} \stackrel{*}{<} t_{\mathcal{I}_{\mathscr{C}}}$, it is also $\mathcal{I}_{\mathscr{C}}$-layerwise computable.

$\square$

Observe that for generic sequences x, $t_{\mathscr{C}}(x) \stackrel{*}{=} t_{Q_x}(x)$. Indeed, $t_{\mathscr{C}}(x) = \inf_{P \in \mathscr{C}} t_P(x) = t_{Q_x}(x)$ as $t_P(x) = +\infty$ for every $P \in \mathscr{C} \setminus \{Q_x\}$.

4.6 A weaker answer: finitely decomposable measures

In the two preceding results, we need the effectivity of the ergodic decomposition. In particular situations, we still get a (weaker) result without this assumption.

Proposition 4.3. *Let P be a stationary measure such that m_P is supported on a closed set $\mathscr{C}$ of stationary ergodic measures. For every P-random x, Q_x is ergodic.*

To prove it we use the following lemma.

Lemma 4.3. *Let X, Y be computable metric spaces. Let $f_n : X \to Y$ be uniformly computable functions that converge P-a.e. to a function f. Let $A \subseteq Y$ be a closed set such that $f(x) \in A$ for P-a.e. x. For every P-random x, $\lim f_n(x) \in A$.*

Proof. It is already known if f is constant P-a.e. Let x_0 be a P-random point such that $\lim f_n(x_0) \notin A$. Let $B(y, r)$ be a ball with computable center and radius, containing $\lim f_n(x_0)$ and disjoint from A. Let $g_n(x) = \max(0, r - d(f_n(x), y))$. For P-a.e. x, $g_n(x)$ converge to 0, but $\lim g_n(x_0) = r - d(\lim f_n(x_0), y) > 0$. □

Proof of Proposition 4.3. For every n, define $Q_n : X \to \mathcal{P}(X)$ by $Q_n(x) = \frac{1}{n}(\delta_x + \ldots + \delta_{T^{n-1}x})$. A sequence x is generic if and only if $Q_n(x)$ is weakly convergent, and in that case Q_x is the limit of $Q_n(x)$. The functions Q_n are uniformly computable. As $Q_x \in \mathscr{C}$ for P-almost every x, $Q_x \in \mathscr{C}$ for every P-random x by Lemma 4.3. □

For instance, if P has a finite decomposition, i.e. if $P = \sum_{i=1}^{n} \alpha_i P_i$ where $\alpha_i \in [0, 1]$, $\sum_i \alpha_i = 1$ and all P_i are ergodic, then regardless of the computability of P, α_i, P_i, for every P-random x, $Q_x \in \{P_1, \ldots, P_n\}$ as the latter set is closed. In this particular case, Q_x is always m-random.

We do not know whether every finitely decomposable measure is effectively decomposable. For instance, are there distinct non-computable ergodic measures P_1, P_2 such that $P := \frac{1}{2}(P_1 + P_2)$ is computable? Such a measure P would be a finitely, non-effectively decomposable measure.

If a finitely, but non-effectively, decomposable measure P exists, we do not know whether for every P-random x, is Q_x-random, Q_x is m_P-random. We only know that Q_x is ergodic.

5 Inferring ergodic measures

In [GHR10] the computability of stationary measures in computable dynamical systems is investigated. It is proved that computable topological systems, that always admit invariant measures, do not necessarily admit computable ones; it is also proved that some interesting classes of dynamical systems always admit computable ergodic measures. Here we are interested in the computability of ergodic measures, relative to its typical trajectories. There are generally uncountably many ergodic measures, so most of them are not computable; nevertheless in many cases all of them can be computed using the observation of any typical trajectory of the system as oracle.

Definition 5.1. A stationary ergodic probability measure P is ***strongly inferable*** if the constant function $x \mapsto P$ is P-layerwise computable.

A stationary ergodic probability measure P is ***weakly inferable*** if for every P-random sequence x, P is computable relative to x.

In the first notion, we ask for some uniformity in the computability of P relative to its random elements. While in many systems there are many ergodic measures that are not computable, for the simple reason that there are uncountably many ergodic measures, in many systems all of them are inferable.

Example 2. The simplest system is the identity $\mathrm{id} : X \to X$. The ergodic stationary measures are the Dirac measures δ_x. Of course, δ_x is computable from x, its only random point, so every ergodic measure is strongly inferable.

Example 3. Consider $X = S \times [0,1]$ where $S = [0,1] \mod 1$ is the unit circle. Let $T(x,y) = (x+y \mod 1, y)$. On $S \times \{y\}$, T is a rotation by angle y. There are two types of ergodic measures: $\lambda \times \delta_y$ where $y \notin \mathbb{Q}$ and λ is the uniform (Lebesgue) measure on S; $\frac{1}{q}(\delta_x + \ldots + \delta_{x+(q-1)y}) \times \delta_y$ where $x \in S$, $y = p/q \in \mathbb{Q}$ with p, q mutually prime. Every ergodic measure is uniformly computable from its random points (even from all the points in its support), so every ergodic measure is strongly inferable.

For the shift transformation on the Cantor space, we do not know whether every ergodic measure is weakly inferable. As a consequence of Theorem 4.3, we obtain:

Corollary 5.1. *If P belongs to an effective compact class of ergodic measures, then P is strongly inferable.*

Proof. Let $\mathscr{C}$ be an effective compact class of ergodic measures that contains P. The P-layers are contained in the $\mathscr{C}$-layers. $x \mapsto Q_x$, which is $\mathscr{C}$-layerwise computable is then also P-layerwise computable. □

As a result, all Bernoulli measures, all Markovian ergodic measures are strongly inferable. Using Baire category (see [Par61]) it can be proved that there exist ergodic shift-invariant measures that avoid every effective compact class of ergodic measures.

6 Open questions

We summarize the questions left open in the general case.

- Is every stationary ergodic measure strongly inferable? weakly inferable?
- Is there a non-effectively decomposable stationary measure that is finitely decomposable? Given two different stationary ergodic measures P_1, P_2: are they computable from the mixture $P := \frac{1}{2}(P_1 + P_2)$? In particular, is it possible that P be computable but not P_1, P_2?
- Given a stationary (non-ergodic) measure, and x a P-random sequence. Is Q_x ergodic? Is x always Q_x-random? Is Q_x always m_P-randoms? The question is not even answered when P is computable, or finitely decomposable.

All these questions can be asked for any computable system.

References

[BDMS10] Laurent Bienvenu, Adam Day, Ilya Mezhirov, and Alexander Shen. Ergodic-type characterizations of algorithmic randomness. In *Computability in Europe (CIE 2010)*, volume 6158 of *Lecture Notes in Computer Science*, pages 49–58. Springer, 2010. 1, 7, 8

[Gác05] Peter Gács. Uniform test of algorithmic randomness over a general space. *Theoretical Computer Science*, 341:91–137, 2005. 3

[Gác08] Peter Gács. Lecture notes on descriptional complexity and randomness. Technical report, Boston University, 2008. http://www.cs.bu.edu/faculty/gacs/papers/ait-notes.pdf. 3, 9

[GHR10] Stefano Galatolo, Mathieu Hoyrup, and Cristóbal Rojas. Dynamics and abstract computability: computing invariant measures. *Discrete and Continuous Dynamical Systems - A (accepted)*, 2010. 11

[HR09a] Mathieu Hoyrup and Cristobal Rojas. Applications of effective probability theory to Martin-Löf randomness. In *ICALP 2009*, volume 5555 of *Lecture Notes in Computer Science*, pages 549–561. Springer, 2009. 1, 4, 8

[HR09b] Mathieu Hoyrup and Cristobal Rojas. Computability of probability measures and Martin-Löf randomness over metric spaces. *Inf. Comput.*, 207(7):830–847, 2009. 3

[HW03] Peter Hertling and Klaus Weihrauch. Random elements in effective topological spaces with measure. *Information and Computation*, 181(1):32–56, 2003. 3

[Lev73] Leonid A. Levin. On the notion of a random sequence. *Soviet Mathematics Doklady*, 14:1413–1416, 1973. 3

[ML66] Per Martin-Löf. The definition of random sequences. *Information and Control*, 9(6):602–619, 1966. 1, 3

[Nan08] Satyadev Nandakumar. An effective ergodic theorem and some applications. In *STOC '08: Proceedings of the 40th annual ACM symposium on Theory of computing*, pages 39–44, New York, NY, USA, 2008. ACM. 1

[Par61] K.R. Parthasarathy. On the category of ergodic measures. *Ill. J. Math.*, 5:648–656, 1961. 11

[V'y97] Vladimir V. V'yugin. Effective convergence in probability and an ergodic theorem for individual random sequences. *SIAM Theory of Probability and Its Applications*, 42(1):39–50, 1997. 5, 6, 8

[V'y98] Vladimir V. V'yugin. Ergodic theorems for individual random sequences. *Theoretical Computer Science*, 207(4):343–361, 1998. 1

[Wei00] Klaus Weihrauch. *Computable Analysis*. Springer, Berlin, 2000. 2