The dimension of ergodic random sequences
Mathieu Hoyrup

To cite this version:

HAL Id: inria-00606457
https://inria.hal.science/inria-00606457v4
Submitted on 21 Jul 2011

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
The dimension of ergodic random sequences

Mathieu Hoyrup

July 21, 2011

Abstract

Let \( \mu \) be a computable ergodic shift-invariant measure over \( \{0, 1\}^\mathbb{N} \). Providing a constructive proof of Shannon-McMillan-Breiman theorem, V’yugin proved that if \( x \in \{0, 1\}^\mathbb{N} \) is Martin-Löf random w.r.t. \( \mu \) then the strong effective dimension \( \text{Dim}(x) \) of \( x \) equals the entropy of \( \mu \). Whether its effective dimension \( \dim(x) \) also equals the entropy was left as a problem question. In this paper we settle this problem, providing a positive answer. A key step in the proof consists in extending recent results on Birkhoff’s ergodic theorem for Martin-Löf random sequences.

Keywords: Shannon-McMillan-Breiman theorem; Martin-Löf random sequence; effective Hausdorff dimension; compression rate; entropy.

1 Introduction

The effective dimension and strong effective dimension of an infinite binary sequence \( x \) are defined as

\[
\dim(x) = \liminf_n \frac{K(x|_n)}{n}, \\
\text{Dim}(x) = \limsup_n \frac{K(x|_n)}{n},
\]

where \( K(w) \) is the Kolmogorov complexity of \( w \).

They can be characterized as effective versions of Hausdorff and packing dimensions respectively, or by divergence of s-gales (see [Lut00, May02, AHLM07] for the original results and [Lut05] for a survey).

Let \( p \in [0, 1] \) be a computable real number and \( \mu_p \) the Bernoulli measure over Cantor space given by \( \mu_p[w] = p^{|w|} (1-p)^{|w|} \). It is well-known that if an infinite binary sequence \( x \) is Martin-Löf random w.r.t. \( \mu_p \) then \( \dim(x) = \text{Dim}(x) = h(\mu_p) \), where \( h(\mu_p) \) is the entropy of \( \mu_p \) defined by

\[
h(\mu_p) = -p \log(p) - (1-p) \log(1-p).
\] (1)
This result is not difficult to prove and reduces to the strong law of large numbers for Martin-Löf random sequences, as on the one hand\(^1\)

\[
K(x|_{[n+1,n+2|}) = - \log \mu_p[x|_{[n+1,n+2|)} + O(1)
\]

for \(\mu_p\)-random sequences by Levin-Schnorr theorem, and on the other hand

\[
- \frac{1}{n} \log \mu_p[x|_{[n]} = - \frac{|x|_{[n]}1}{n} \log(p) - \frac{|x|_{[n]}0}{n} \log(1-p)
\]

which converge to \(h(\mu_p)\) for \(\mu_p\)-random sequences, by the Strong Law of Large Numbers for Martin-Löf random sequences.

This result highlights the relationship between Shannon’s information theory, Kolmogorov algorithmic information theory and effective randomness.

Ergodic theory provides a natural extension of information theory in which many results can be transferred, with more involved proofs, from the case of independent identically distributed random variables to the ergodic case, where independence is only required \textit{asymptotically, in the average} (see Section 2 for a precise definition).

First, the strong law of large numbers extends to Birkhoff’s ergodic theorem. Second, the coincidence between local information and entropy extends through the Shannon-McMillan-Breiman theorem. Whether Martin-Löf randomness fits with these theorems has been an open problem for a while. The first results were proved by V’yugin [V’y98], based on non-classical, constructive proofs of the theorems. He proved, in particular:

\textbf{Theorem 1.1 (Effective Birkhoff ergodic theorem I).} Let \(\mu\) be a computable shift-invariant ergodic measure over \(\{0,1\}^\mathbb{N}\) and \(f \in L^1(\mu)\) be computable. For every Martin-Löf \(\mu\)-random sequence \(x\),

\[
\lim_{n \to \infty} \frac{1}{n} \sum_{k=0}^{n-1} f \circ T^k(x) = \int f \, d\mu.
\]

The entropy of an ergodic measure is defined as

\[
h(\mu) = \lim_{n \to \infty} - \frac{1}{n} \sum_{|w|=n} \mu[w] \log \mu[w].
\]

Observe that (1) and (2) are consistent as they give the same quantity when \(\mu\) is a Bernoulli measure.

\textbf{Theorem 1.2 (Effective Shannon-McMillan-Breiman theorem I).} Let \(\mu\) be a computable shift-invariant ergodic measure over \(\{0,1\}^\mathbb{N}\). For every Martin-Löf \(\mu\)-random sequence \(x\),

\[
\limsup_{n \to \infty} \frac{K(x|_{[n|})}{n} = \limsup_{n \to \infty} - \frac{1}{n} \log \mu[x|_{[n|}] = h(\mu).
\]

\(^1\)\(K\) is the prefix version of Kolmogorov complexity.
The question whether $\lim \inf \frac{K(w)}{n}$ coincides with $h(\mu)$ for every Martin-Löf $\mu$-random was left open by V'yugin. An alternative proof of Theorem 1.2 approximating ergodic measures by Markovian measures was later developed by Nakamura [Nak05], but also left the question open. In this paper we provide a positive answer to this question.

A classical proof of the Shannon-McMillan-Breiman theorem uses Birkhoff’s ergodic theorem, applied to some particular functions. The problem in making it effective is that these functions are not computable in general. Recent works have been achieved to push the effective ergodic theorem to the largest possible class of functions. Here we extend it enough to get the full effective Shannon-McMillan-Breiman theorem.

In Section 2 we recall basic notions of computability, randomness and ergodic theory. In Section 3 we develop effective versions of Birkhoff’s ergodic theorem. In Section 4 we present our main result.

2 Background and notations

We work on the Cantor space $\{0,1\}^\mathbb{N}$ of infinite binary sequences. A finite word $w \in \{0,1\}^*$ determines the cylinder $[w] \subseteq \{0,1\}^\mathbb{N}$ of infinite sequences starting with $w$. If $x \in \{0,1\}^\mathbb{N}$ and $n \in \mathbb{N}$, $x|_n$ is the prefix of $x$ of length $n$, and is also denoted $x_0x_1\ldots x_{n-1}$. The cylinders form a base of the product topology.

Effective topology. An open set $U \subseteq \{0,1\}^\mathbb{N}$ is effective if it is a recursively enumerable union of cylinders. A closed set is effective if its complement is an effective open set. A function $f : \{0,1\}^\mathbb{N} \to \mathbb{R}$ is computable if there is a Turing machine that on oracle $x$ and input $n$ computes a rational number $q$ such that $|q - f(x)| < 2^{-n}$. Equivalently, $f$ is computable if for every rational numbers $a < b$, $f^{-1}(a, b)$ is effectively open, uniformly in $a, b$. A function $f : \{0,1\}^\mathbb{N} \to [0, +\infty]$ is lower (resp. upper) semi-computable if there is a Turing machine that on oracle $x$ and input $n$ computes a rational number $q_n$ such that $f(x) = \sup_n q_n$ (resp. $f(x) = \inf_n q_n$). Equivalently, $f$ is lower (resp. upper) semi-computable if for every rational number $a$, $f^{-1}(a, +\infty]$ (resp. $f^{-1}[0, a)$) is effectively open, uniformly in $a$.

Kolmogorov complexity and Martin-Löf randomness. For $w \in \{0,1\}^*$, $K(w)$ is the prefix version of Kolmogorov complexity, defined by Levin and Chaitin independently. It is defined as the length of a shortest input of a universal Turing machine with prefix-free domain computing $w$ on that input.

A probability measure $\mu$ over $\{0,1\}^\mathbb{N}$ is determined by its value on cylinders $\mu[w]$, for $w \in \{0,1\}^*$. $\mu$ is computable if $\mu[w]$ is a computable real number, uniformly in $w$. Given a computable probability measure $\mu$, a sequence $x \in \{0,1\}^\mathbb{N}$ is Martin-Löf $\mu$-random, denoted $x \in \text{ML}_\mu$, if there is $c$ such that for all $n$,

$$K(x|_n) \geq -\log \mu[x|_n] - c.$$
Martin-Löf’s original definition [ML66] was expressed in terms of tests rather than complexity, but the one given here, due to Levin and Chaitin [Cha75] independently, was proved to be equivalent by Levin [Lev73] and Schnorr [Sch73].

The function
t_{\mu}(x) = \sup_{n} \{- \log \mu([x]_{n}) - K(x|_{n})\}

is lower semi-computable and \(\int 2^{t_{\mu}} d\mu \leq 1\). Moreover, it was proved in [Gác80] that \(2^{t_{\mu}}\) is maximal in the sense that for every integrable lower semi-computable function \(f : \{0,1\}^N \rightarrow [0, +\infty]\), there exists \(c_{f}\) such that \(f \leq c_{f}2^{t_{\mu}}\). We call such an \(f\) a \(\mu\)-test. It tests Martin-Löf randomness in the sense that \(x\) is \(\mu\)-random if \(f(x) < \infty\) for each \(\mu\)-test \(f\) iff \(t_{\mu}(x) < \infty\). More can be found on this subject in [Nie09, DH10].

**Ergodic theory.** We recall some basic notions of ergodic theory, more details can be found in [Smo71, Pet83]. We denote by \(T : \{0,1\}^N \rightarrow \{0,1\}^N\) the **shift map** defined by \(T(x_{0}x_{1} \ldots) = x_{1}x_{2} \ldots\). A measure \(\mu\) over \(\{0,1\}^N\) is **shift-invariant** if for all Borel sets \(A\), \(\mu(T^{-1}A) = \mu(A)\), equivalently if \(\mu([0w]) + \mu([1w]) = \mu([w])\) for all \(w \in \{0,1\}^*\). \(\mu\) is **ergodic** if for all Borel sets \(A\) such that \(T^{-1}A = A\) up to a null sets, \(\mu(A) = 0\) or \(1\). Equivalently, \(\mu\) is ergodic if for all \(u, v \in \{0,1\}^*\),

\[
\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=0}^{n-1} \mu([u] \cap T^{-k}[v]) = \mu[u] \cdot \mu[v].
\]

## 3 Effective ergodic theorems

The following theorem, taken from [BDMS10], extends a result of Kučera from the uniform measure to any ergodic shift-invariant measure:

**Theorem 3.1** (Effective Poincaré recurrence theorem). Let \(\mu\) be a computable ergodic shift-invariant measure and \(C \subseteq \{0,1\}^N\) an effective closed set such that \(\mu(C) > 0\). Every Martin-Löf \(\mu\)-random sequence has a tail in \(C\), i.e. for every \(x \in \text{ML}_{\mu}\) there exists \(k\) such that \(T^k(x) \in C\).

In [BDH+10] and [FGMN10] independently this result was used to prove that not only the orbit of \(x\) eventually falls into \(C\), but it does so with frequency \(\mu(C)\).

**Theorem 3.2** (Effective Birkhoff ergodic theorem II). Let \(\mu\) be a computable ergodic shift-invariant measure and \(C \subseteq \{0,1\}^N\) an effective closed set such that \(\mu(C) > 0\). For every Martin-Löf \(\mu\)-random sequence \(x\),

\[
\lim_{n \rightarrow \infty} \frac{1}{n} |\{k < n : T^k(x) \in C\}| = \mu(C).
\]

We first generalize the result from sets to functions:

**Theorem 3.3** (Effective Birkhoff ergodic theorem III). Let \(\mu\) be a computable ergodic shift-invariant measure. Assume \(f : \{0,1\}^N \rightarrow [0, +\infty]\) is:
• either lower semi-computable,

• or upper semi-computable and bounded by a \( \mu \)-test.

For each \( x \in \text{ML}_\mu \),

\[
\lim_{n \to \infty} \sum_{k=0}^{n-1} f \circ T^k(x) = \int f \, d\mu.
\]

**Proof.** Let us introduce the notation \( A^f_n(x) = \frac{1}{n}(f(x) + \ldots + f \circ T^{n-1}(x)) \).

If \( f \) is lower semi-computable, then there is a sequence of uniformly computable non-negative functions \( f_n \uparrow f \). Applying Theorem 1.1 to \( f_n \) and \( x \in \text{ML}_\mu \) gives \( \liminf_k A^f_k(x) \geq \liminf_k A^f_n(x) = \int f_n \, d\mu \). By the monotone convergence theorem, \( \int f_n \, d\mu \uparrow \int f \, d\mu \), so \( \liminf_k A^f_k(x) \geq \int f \, d\mu \). If \( \int f \, d\mu = \infty \) we are done. Otherwise, let \( q > \int f \, d\mu \) be a rational number. The set \( C_K := \{ x : \forall k \geq K, A^f_k(x) \leq q \} \) is effectively closed and by the classical ergodic theorem, there exists \( K \) such that \( \mu(C_K) > 0 \). Theorem 3.1 tells us that if \( x \in \text{ML}_\mu \) then there is \( n \) such that \( T^n(x) \in C_K \). As a result, \( \limsup A^f_k(x) = \limsup A^f_k(T^n(x)) \leq q \).

As this is true of every \( q > \int f \, d\mu \), we get the result.

Now, if \( f \) is upper semi-computable and \( f \leq t \) where \( t \) is a \( \mu \)-test, then for \( x \in \text{ML}_\mu \), applying the preceding result to \( t \) and \( t - f \),

\[
A^f_n(x) = A^t_n(x) - (A^{t-f}_n(x)) \to \int t \, d\mu - \int (t-f) \, d\mu = \int f \, d\mu.
\]

We then extend this result further:

**Corollary 3.1** (Effective Birkhoff ergodic theorem IV). Let \( f : \{0, 1\}^\mathbb{N} \to [0, +\infty] \) be \( \Delta^0_2 \) on \( \text{ML}_\mu \), i.e. there is a sequence \( f_n \) of uniformly computable functions such that \( f(x) = \lim_n f_n(x) \) for each \( x \in \text{ML}_\mu \). Assume that \( f \) is dominated by a \( \mu \)-test. For every \( x \in \text{ML}_\mu \),

\[
\lim_{n \to \infty} \sum_{k=0}^{n-1} f \circ T^k(x) = \int f \, d\mu.
\]

**Proof.** Let \( g_N = \inf_{n \geq N} f_n \) and \( h_N = \min(t, \sup_{n \geq N} f_n) \). On \( \text{ML}_\mu \), \( g_N \uparrow f \) and \( h_N \downarrow f \). By the monotone and dominated convergence theorem, the convergences hold in \( L^1(\mu) \).

Applying Theorem 3.2 to \( g_N \) and \( h_N \) gives the result. More precisely, for every \( x \in \text{ML}_\mu \) and every \( N \),

\[
\liminf_n A^f_n(x) \geq \liminf_n A^{g_N}_n(x) = \int g_N \, d\mu
\]

\[
\limsup_n A^f_n(x) \leq \limsup_n A^{h_N}_n(x) = \int h_N \, d\mu,
\]

so

\[
\int f \, d\mu = \sup_N \int g_N \, d\mu \leq \liminf_n A^f_n(x) \leq \limsup_n A^f_n(x) \leq \inf_N \int h_N \, d\mu = \int f \, d\mu.
\]
4 The effective Shannon-McMillan-Breiman theorem

We now present our main result.

**Theorem 4.1** (Effective Shannon-McMillan-Breiman theorem II). Let $\mu$ be a computable shift-invariant probability measure. For each $x \in \text{ML}_\mu$,

$$
\lim_{n \to \infty} \frac{K(x \upharpoonright n)}{n} = \lim_{n \to \infty} -\frac{1}{n} \log \mu[x \upharpoonright n] = h(\mu).
$$

A proof of the classical result, stating the result for a.e. $x$, can be found in [Smo71, Pet83]. It makes use of martingale convergence theorems and ergodic theorems. The main difficulty in adapting the proof is to make sure that the effective versions of the ergodic theorem can be applied. The rest of this section is devoted to the proof of Theorem 4.1.

An easy calculation shows that

$$
-\log \mu[x \upharpoonright n] = \sum_{k=0}^{n-1} f_{n-1-k} \circ T^k(x).
$$

where

$$
f_k(x) := -\log \mu[x_0 \ldots x_k] = \log \frac{\mu[x_0 \ldots x_k]}{\mu[x_1 \ldots x_k]} \quad \text{for } k \geq 1,
$$

$$f_0(x) := -\log \mu[x_0].$$

**Lemma 4.1.** $f_k(x)$ converge for each $x \in \text{ML}_\mu$.

**Proof.** Define the computable martingale

$$
d(\epsilon) = 2
$$

$$
d(x_0) = \frac{1}{\mu[x_0]}
$$

$$
d(x_0 \ldots x_k) = \frac{\mu[x_1 \ldots x_k]}{\mu[x_0 \ldots x_k]} \quad \text{for } k \geq 1.
$$

By the effective Doob’s convergence theorem (see Theorem 7.1.3 on page 270 in [DH10]), for each $x \in \text{ML}_\mu$, $d(x_0 \ldots x_k)$ converges, and so does $f_k(x) = \log d(x_0 \ldots x_k)$. \hfill \Box

Let $f(x)$ be the limit. We write

$$
-\frac{1}{n} \log \mu[x \upharpoonright n] = \frac{1}{n} \sum_{k=0}^{n-1} (f_{n-1-k} \circ T^k(x) - f \circ T^k(x)) + \frac{1}{n} \sum_{k=0}^{n-1} f \circ T^k(x)
$$

and prove that the first term tends to 0 while the second term converges to $\int f \, d\mu = h(\mu)$.

We will use the following lemma (Corollary 2.2 on page 261 in [Pet83], Lemma 4.26 on page 26 in [Smo71]).
Lemma 4.2. \( f^* := \sup_k f_k \in L^1 \).

As \( f_k \to f \) a.e. and the convergence is dominated by \( f^* \in L^1, f_k \to f \) in \( L^1 \).

Proposition 4.1. For each \( x \in \ML_\mu \),

\[
\lim_{n} \frac{1}{n} \sum_{k=0}^{n-1} f \circ T^k(x) = \int f \, d\mu = h_\mu(P). \tag{4}
\]

Proof. That \( \int f \, d\mu = h(\mu) \) is a classical result and follows from \( h(\mu) = \lim_k \int f_k \, d\mu \) and the \( L^1 \)-convergence of \( f_k \) to \( f \).

\( f^* \) is lower semi-computable and by Lemma 4.2 it is a \( \mu \)-test. By construction, \( f \) is \( \Delta_2^0 \) on \( \ML_\mu \) and it is dominated by \( f^* \) so it satisfies the conditions of Corollary 3.1, from which the result follows directly. \( \square \)

Proposition 4.2. For each \( x \in \ML_\mu \),

\[
\lim_{n \to \infty} \frac{1}{n} \sum_{k=0}^{n-1} f_{n-1-k} \circ T^k(x) - f \circ T^k(x) = 0. \tag{5}
\]

Proof. Let

\[
g_N = \sup_{k \geq N} |f_k - f| \quad \text{and} \quad \tilde{g}_N = \sup_{k,j \geq N} |f_k - f_j|.
\]

For \( x \in \ML_\mu \),

\[
|f_k(x) - f(x)| = \lim_j |f_k(x) - f_j(x)|
\]

\[
= \lim_j \sup_j |f_k(x) - f_j(x)|
\]

\[
\leq \sup_{j \geq N} |f_k(x) - f_j(x)|,
\]

so \( g_N(x) \leq \tilde{g}_N(x) \). As \( f_k \to f \) a.e., \( \tilde{g}_N \to 0 \) a.e. As \( \tilde{g}_N \leq 2f^* \in L^1 \), \( \tilde{g}_N \to 0 \) in \( L^1 \) by the dominated convergence theorem. On \( \ML_\mu \),

\[
\left| \frac{1}{n} \sum_{k=0}^{n-1} f_{n-1-k} \circ T^k - f \circ T^k \right|
\]

\[
\leq \frac{1}{n} \sum_{k=0}^{n-1} |f_{n-1-k} \circ T^k - f \circ T^k|
\]

\[
= \frac{1}{n} \sum_{k=0}^{n-1} |f_{n-1-k} \circ T^k - f \circ T^k| + \frac{1}{n} \sum_{k=n-N}^{n-1} |f_{n-1-k} \circ T^k - f \circ T^k|
\]

\[
\leq \frac{1}{n} \sum_{k=0}^{n-1} g_N \circ T^k + \frac{1}{n} \sum_{k=n-N}^{n-1} (f^* + f) \circ T^k
\]

\[
\leq \frac{1}{n} \sum_{k=0}^{n-1} \tilde{g}_N \circ T^k + \frac{1}{n} \sum_{k=0}^{n-1} (f^* + f) \circ T^k - \frac{1}{n} \sum_{k=0}^{n-N-1} (f^* + f) \circ T^k.
\]
Fix $N$ and let $n \to \infty$. As $\tilde{g}_N \in L^1$ is lower semi-computable, the first term converges to $\int \tilde{g}_N \, d\mu$ by the Effective Ergodic Theorem 3.3. As $f^* + f$ is $\Delta^0_2$ on $\text{ML}_\mu$, and is dominated by the $\mu$-test $2f^*$, the second and the third terms converge to $\int (f^* + f) \, d\mu$ by Corollary 3.1 so their limits cancel each other.

As $\int \tilde{g}_N \, d\mu \to 0$, we have proved equality (5). \qed

Putting equalities (3), (4) and (5) together gives, for $x \in \text{ML}_\mu$,

$$\lim_{n} - \frac{1}{n} \log \mu[x \mid n] = \lim_{n} \frac{1}{n} \sum_{k=0}^{n-1} f_{n-1-k} \circ T^k(x) = h(\mu).$$

References


