HAL open science

The dimension of ergodic random sequences

Mathieu Hoyrup

To cite this version:

Mathieu Hoyrup. The dimension of ergodic random sequences. [Research Report] 2011. inria-00606457v1

HAL Id: inria-00606457

https://inria.hal.science/inria-00606457v1

Submitted on 6 Jul 2011 (v1), last revised 21 Jul 2011 (v4)

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 6, 2011

Abstract

Let μ 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. μ then the strong effective dimension $\mathrm{Dim}(x)$ of x equals the entropy of μ. Whether its effective dimension $\dim(x)$ also equals the entropy was left as an 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.

1 Introduction

The *effective dimension* and *strong effective dimension* of an infinite binary sequence x are defined as

$$\begin{aligned} \dim(x) &= \liminf_n \frac{K(x\restriction_n)}{n} \\ \mathrm{Dim}(x) &= \limsup_n \frac{K(x\restriction_n)}{n}, \end{aligned}$$

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 μ_p the Bernoulli measure given by $\mu_p[w] = p^{|w|_1}(1-p)^{|w|_0}$. It is well-known that when x is Martin-Löf random w.r.t. μ_p then $\dim(x) = \mathrm{Dim}(x) = h(\mu_p)$, where $h(\mu_p)$ is the entropy of μ_p defined by

$$h(\mu_p) = -p\log(p) - (1-p)\log(1-p). \tag{1}$$

keywords: Shannon-McMillan-Breiman theorem; Martin-Löf random sequence; effective Hausdorff dimension.

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

$$K(x\restriction_n) = -\log \mu_p[x\restriction_n] + O(1)$$

for μ_p-random sequences by Levein-Schnorr theorem, and on the other hand

$$-\frac{1}{n}\log \mu_p[x\restriction_n] = -\frac{|x\restriction_n|_1}{n}\log(p) - \frac{|x\restriction_n|_0}{n}\log(1-p)$$

which converge to $h(\mu_p)$ for μ_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 *asymptotically, in the average* (see Section **??** 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:

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

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

The entropy of an ergodic measures is defined as

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

given the same quantity as (1) when μ is Bernoulli.

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

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

K is the prefix version of Kolmogorov complexity

The question whether $\liminf \frac{K\restriction_n}{n}$ coincides with $h(\mu)$ was left open by V'yugin. An alternative proof of the second theorem, approximating ergodic measures by Markovian measures, was later developed by Nakamura [Nak05], but also left the question open.

A classical proof of Shannon-McMillan-Breiman 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.

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 $n \in \mathbb{N}$, $x\restriction_n$ is the prefix of x of length n, and is also denoted $x_0x_1\dots x_{n-1}$. 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 it 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 μ over $\{0,1\}^{\mathbb{N}}$ is determined by its value on cylinders $\mu[w]$, for $w \in \{0,1\}^*$. μ is computable if $\mu[w]$ is a computable real number, uniformly in w. Given a computable probability measure μ, a sequence $x \in \{0,1\}^{\mathbb{N}}$ is ***Martin-Löf μ-random***, denoted $x \in \mathsf{ML}_\mu$, if there is c such that for all n,

$$K(x\restriction_n) \geq -\log \mu[x\restriction_n] - c.$$

The function

$$t_\mu(x) = \sup\{-\log \mu[x\restriction_n] - K(x\restriction_n)\}$$

is lower semi-computable and $\int t_\mu \, \mathrm{d}\mu \leq 1$. Moreover, t_μ is maximal in the sense that for every integrable lower semi-computable function $f : \{0,1\}^{\mathbb{N}} \to [0,+\infty]$, there exists c_f

such that $f \leq c_f t$. We call such an f a ***μ-test***. It tests Martin-Löf randomness in the sense that $x \in \mathsf{ML}_\mu$ iff $f(x) < \infty$ for each μ-test f iff $t_\mu(x) < \infty$. We refer the reader to [Nie09].

Ergodic theory. We recall some basic notions of ergodic theory, more details can be found in [Smo71, Pet83]. We denote by $T : \{0,1\}^\mathbb{N} \to \{0,1\}^\mathbb{N}$ the ***shift map*** defined by $T(x_0x_1\ldots) = x_1x_2\ldots$. A measure μ over $\{0,1\}^\mathbb{N}$ is ***shit-invariant*** if for all Borel sets A, $\mu(T^{-1}A) = \mu(A)$, equivalently if $\mu[0w] + \mu[1w] = \mu[w]$ for all $w \in \{0,1\}^*$. μ 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, μ is ergodic if for all $u, v \in \{0,1\}^*$,

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

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 μ be a computable ergodic shift-invariant measure and $C \subseteq \{0,1\}^\mathbb{N}$ an effective closed set such that $\mu(C) > 0$. Every Martin-Löf μ-random sequence has a tail in C, i.e. for every $x \in \mathsf{ML}_\mu$ 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 μ be a computable ergodic shift-invariant measure and $C \subseteq \{0,1\}^\mathbb{N}$ an effective closed set such that $\mu(C) > 0$. For every Martin-Löf μ-random sequence x,*

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

We first generalize the result from sets to functions:

Theorem 3.3 (Effective Birkhoff ergodic theorem III)**.** *Let μ be a computable ergodic shift-invariant measure. Assume $f : \{0,1\}^\mathbb{N} \to [0, +\infty]$ is:*

- *either lower semi-computable,*
- *or upper semi-computable and bounded by a μ-test.*

For each $x \in \mathsf{ML}_\mu$,

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

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

If f is lower semi-computable, then there is a sequence of uniformly computable non-negative functions $f_n \nearrow f$. Applying Theorem 1.1 to f_n and $x \in \mathsf{ML}_\mu$ gives $\liminf_k A_k^f(x) \geq \liminf_k A_k^{f_n}(x) = \int f_n \,\mathrm{d}\mu$. By the monotone convergence theorem, $\int f_n \,\mathrm{d}\mu \nearrow \int f \,\mathrm{d}\mu$, so $\liminf_k A_k^f(x) \geq \int f \,\mathrm{d}\mu$. If $\int f \,\mathrm{d}\mu = \infty$ we are done. Otherwise, let $q > \int f \,\mathrm{d}\mu$ be a rational number. The set $C_K := \{x : \forall k \geq K, A_k^f(x) \leq q\}$ 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 $x \in \mathsf{ML}_\mu$ then there is n such that $T^n(x) \in C_K$. As a result, $\limsup A_k^f(x) = \limsup A_k^f(T^n(x)) \leq q$. As this is true of every $q > \int f \,\mathrm{d}\mu$, we get the result.

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

$$A_n^f(x) = A_n^t(x) - (A_n^{t-f}(x)) \to \int t \,\mathrm{d}\mu - \int (t - f) \,\mathrm{d}\mu = \int f \,\mathrm{d}\mu.$$

□

We then extend this result further:

Corollary 3.1 (Effective Birkhoff ergodic theorem IV)**.** *Let $f : \{0,1\}^{\mathbb{N}} \to [0, +\infty]$ be Δ_2^0 on ML_μ, i.e. there is a sequence f_n of uniformly computable functions such that $f(x) = \lim_n f_n(x)$ for each $x \in \mathsf{ML}_\mu$. Assume that f is dominated by a μ-test. For every $x \in \mathsf{ML}_\mu$,*

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

Proof. Let $g_N = \inf_{n \geq N} f_n$ and $h_N = \min(t, \sup_{n \geq N} f_n)$. On ML_μ, $g_N \nearrow f$ and $h_N \searrow f$. 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 $x \in \mathsf{ML}_\mu$ and every N,

$$\begin{aligned} \liminf_n A_n^f(x) &\geq \liminf_n A_n^{g_N}(x) = \int g_N \,\mathrm{d}\mu \\ \limsup_n A_n^f(x) &\leq \limsup_n A_n^{h_N}(x) = \int h_N \,\mathrm{d}\mu, \end{aligned}$$

so

$$\int f \,\mathrm{d}\mu = \sup_N \int g_N \,\mathrm{d}\mu \leq \liminf_n A_n^f(x) \leq \limsup_n A_n^f(x) \leq \inf_N \int h_N \,\mathrm{d}\mu = \int f \,\mathrm{d}\mu.$$

□

4 The effective Shannon-McMillan-Breiman theorem

We now present our main result.

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

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

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

$$-\log \mu[x\restriction_n] = \sum_{k=0}^{n-1} f_{n-1-k} \circ T^k(x) \tag{2}$$

where

$$\begin{aligned} f_k(x) &:= -\log \mu[x_0|x_1 \dots x_k] = -\log \frac{\mu[x_0 \dots x_k]}{\mu[x_1 \dots x_k]} \quad \text{for } k \geq 1, \\ f_0(x) &:= -\log \mu[x_0]. \end{aligned}$$

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

Proof. Define the computable martingale

$$\begin{aligned} d(\epsilon) &= 2 \\ d(x_0) &= \frac{1}{\mu[x_0]} \\ d(x_0 \dots x_k) &= \frac{\mu[x_1 \dots x_k]}{\mu[x_0 \dots x_k]} \quad \text{for } k \geq 1. \end{aligned}$$

By the effective Doob's convergence theorem, for each $x \in \mathsf{ML}_\mu$, $d(x_0 \dots x_k)$ converges, and so does $f_k(x) = \log d(x_0 \dots x_k)$. □

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

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

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

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* $x \in \mathsf{ML}_\mu$,

$$\lim_n \frac{1}{n} \sum_{k=0}^{n-1} f \circ T^k(x) = \int f \, \mathrm{d}\mu = h_\mu(P). \tag{3}$$

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

f^* is lower semi-computable and by Lemma 4.2 it is a μ-test. By construction, f is Δ^0_2 on ML_μ and it is dominated by f^* so it satisfies the conditions of Corollary 3.1, from which the result follows directly. □

Proposition 4.2. *For each* $x \in \mathsf{ML}_\mu$,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} f_{n-1-k} \circ T^k(x) - f \circ T^k(x) = 0. \tag{4}$$

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 $x \in \mathsf{ML}_\mu$,

$$|f_k(x) - f(x)| = \lim_j |f_k(x) - f_j(x)| = \limsup_j |f_k(x) - f_j(x)| \leq \sup_{j \geq N} |f_k(x) - f_j(x)|,$$

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_μ,

$$\begin{aligned}
\left| \frac{1}{n} \sum_{k=0}^{n-1} f_{n-1-k} \circ T^k - f \circ T^k \right| &\leq \frac{1}{n} \sum_{k=0}^{n-1} |f_{n-1-k} \circ T^k - f \circ T^k| \\
&= \frac{1}{n} \sum_{k=0}^{n-1-N} |f_{n-1-k} \circ T^k - f \circ T^k| + \frac{1}{n} \sum_{k=n-N}^{n-1} |f_{n-1-k} \circ T^k - f \circ T^k| \\
&\leq \frac{1}{n} \sum_{k=0}^{n-1-N} g_N \circ T^k + \frac{1}{n} \sum_{k=n-N}^{n-1} (f^* + f) \circ T^k \\
&\leq \frac{1}{n} \sum_{k=0}^{n-1-N} \tilde{g}_N \circ T^k + \frac{1}{n} \sum_{k=0}^{n-1} (f^* + f) \circ T^k - \frac{1}{n} \sum_{k=0}^{n-N-1} (f^* + f) \circ T^k.
\end{aligned}$$

Fix N and let $n \to \infty$. As $\tilde{g}_N \in L^1$ is lower semi-computable, the first term converges to $\int \tilde{g}_N \, \mathrm{d}\mu$ by the Effective Ergodic Theorem 3.3. As $f^* + f$ is Δ^0_2 on ML_μ and is dominated by the μ-test $2f^*$, the second and the third terms converge to $\int (f^* + f) \, \mathrm{d}\mu$ by Corollary 3.1 so their limits cancel each other.

As $\int \tilde{g}_N \, \mathrm{d}\mu \to 0$, we have proved equality (4). □

Putting together equalities (2), (3) and (4) give, for $x \in \mathsf{ML}_\mu$,

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

References

[AHLM07] Krishna B. Athreya, John M. Hitchcock, Jack H. Lutz, and Elvira Mayordomo. Effective strong dimension in algorithmic information and computational complexity. *SIAM J. Comput.*, 37(3):671–705, 2007. 1

[BDH+10] 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. Submitted. ArXiv 1007.5249, 2010. 4

[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. 4

[FGMN10] Johanna N.Y. Franklin, Noam Greenberg, Joseph S. Miller, and Keng Meng Ng. Martin-Löf random points satisfy Birkhoff's ergodic theorem for effectively closed sets. To appear in the *Proceedings of the American Mathematical Society*, 2010. 4

[Lut00] Jack H. Lutz. Dimension in complexity classes. In *IEEE Conference on Computational Complexity*, pages 158–169, 2000. 1

[Lut05] Jack H. Lutz. Effective fractal dimensions. *Mathematical Logic Quarterly*, 51(1):62–72, 2005. 1

[May02] Elvira Mayordomo. A kolmogorov complexity characterization of constructive hausdorff dimension. *Inf. Process. Lett.*, 84(1):1–3, 2002. 1

[Nak05] Masahiro Nakamura. Ergodic theorems for algorithmically random sequences. *Proceedings of the Symposium on Information Theory and Its Applications*, 2005. 3

[Nie09] A. Nies. *Computability and randomness*. Oxford logic guides. Oxford University Press, 2009. 4

[Pet83] Karl Petersen. *Ergodic Theory*. Cambridge Univ. Press, 1983. 4, 6

[Smo71] Meir Smorodinsky. *Ergodic Theory, Entropy*, volume 214 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin Heibelberg New York, 1971. 4, 6

[V'y98] Vladimir V. V'yugin. Ergodic theorems for individual random sequences. *Theoretical Computer Science*, 207(4):343–361, 1998. 2