The dimension of ergodic random sequences

Mathieu Hoyrup[1]

1 LORIA, INRIA Nancy - Grand Est
LORIA (Carte, bât. B)
615 rue du jardin botanique
BP 239
54506 Vandœuvre-lès-Nancy
France
mathieu.hoyrup@loria.fr

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 open problem. 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. At the same time, we present extensions of some previous results.

As pointed out by a referee the main result can also be derived from results by Hochman [8], using rather different considerations.

1998 ACM Subject Classification E.4 Coding and Information Theory, F.4.1 Mathematical Logic, G.3 Probability and Statistics

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

Digital Object Identifier 10.4230/LIPIcs.STACS.2012.567

1 Introduction

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

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

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 gales (see [12, 15, 1] for the original results and [13] for a survey).

Let $p \in [0,1]$ be a computable real number and μ_p the Bernoulli measure over Cantor space given by $\mu_p[w] = p^{|w|_1}(1-p)^{|w|_0}$. It is well-known[1] that if an infinite binary sequence

[1] see [11] for a generalization of this result

29th Symposium on Theoretical Aspects of Computer Science (STACS'12).
Editors: Christoph Dürr, Thomas Wilke; pp. 567–576
Leibniz International Proceedings in Informatics
LIPICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany
SYMPOSIUM ON THEORETICAL ASPECTS OF COMPUTER SCIENCE

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

$$h(\mu_p) = -p\log(p) - (1-p)\log(1-p). \quad (1)$$

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[2]

$$K(x{\restriction_n}) = -\log \mu_p[x{\restriction_n}] + O(\log(n))$$

for μ_p-random sequences by Levin-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 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 [22], based on non-classical, constructive proofs of the theorems. He proved, in particular:

▶ **Theorem 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 x,*

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

The entropy of an ergodic measure is defined as

$$h(\mu) = \lim_{n\to\infty} -\frac{1}{n}\sum_{|w|=n} \mu[w]\log\mu[w]. \quad (2)$$

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

▶ **Theorem 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 x,*

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

[2] K is the prefix-free version of Kolmogorov complexity

The question whether $\liminf \frac{K(x\upharpoonright_n)}{n}$ coincides with $h(\mu)$ for every Martin-Löf μ-random was left open by V'yugin. An alternative proof of Theorem 2 approximating ergodic measures by Markovian measures was later developed by Nakamura [16], but also left the question open. In this paper we provide a positive answer to this question. As was pointed out by a referee, Hochman's constructive proof of the Shannon-McMillan-Breiman theorem [8] gives another proof of the result.

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 $n \in \mathbb{N}$, $x\upharpoonright_n$ 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-free 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 the cylinders $\mu[w]$, for $w \in \{0,1\}^*$. μ is computable if all $\mu[w]$ are computable real numbers, 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,

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

Martin-Löf's original definition of a random sequence (in [14]) was expressed in terms of tests rather than complexity, but the one given above, due to Levin and Chaitin [4] independentely, was proved to be equivalent by Levin [9] and Schnorr [20].

Let us briefly present a notion of randomness test that we will use in the sequel. A μ-***test*** is a lower semi-computable function $f : \{0,1\}^{\mathbb{N}} \to [0,+\infty]$ such that $\int f \, \mathrm{d}\mu \leq 1$. The definition of a Martin-Löf random sequence can be rephrased as follows: $x \in \mathsf{ML}_\mu$ if and only if the quantity

$$d_\mu(x) = \sup_n \{-\log \mu[x{\restriction}_n] - K(x{\restriction}_n)\},$$

called the ***randomness deficiency*** of x, is finite. It was proved in [9] that $t_\mu := 2^{d_\mu}$ is a μ-test and by Gács [7] that it is *optimal* in the sense that for every μ-test f, there exists a constant c_f such that $f \leq c_f t_\mu$. As a result, $x \in \mathsf{ML}_\mu$ if and only if $f(x) < \infty$ for each μ-test f if and only if $t_\mu(x) < \infty$. More can be found on this subject in [10] as well as in the recent textbooks [17, 5].

Ergodic theory

We recall some basic notions of ergodic theory, more details can be found in [21, 19]. We denote by $T : \{0,1\}^{\mathbb{N}} \to \{0,1\}^{\mathbb{N}}$ the ***shift map*** defined by $T(x_0x_1x_2\dots) = x_1x_2x_3\dots$. A measure μ over $\{0,1\}^{\mathbb{N}}$ 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 $w \in \{0,1\}^*$. μ is ***ergodic*** if for all Borel sets A such that $T^{-1}A = A$ up to a null set, $\mu(A) = 0$ or 1. Equivalently, μ is ergodic if for all $u, v \in \{0,1\}^*$,

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

Every Bernoulli measure is shift-invariant and ergodic.

3 Effective ergodic theorems

The following theorem, taken from [2], extends a result of Kučera from the uniform measure to any ergodic shift-invariant measure:

▶ **Theorem 3** (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 [3] and [6] 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 4** (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 5** (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 denote the Birkhoff averages by $A_n^f(x) = \frac{1}{n}(f(x) + \ldots + f \circ T^{n-1}(x))$.

If f is lower semi-computable, then there is a sequence of uniformly computable non-negative functions $f_n \nearrow f$. Applying Theorem 1 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 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 for 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 $t - f$,

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

We then extend this result further:

▶ **Corollary 6** (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 $h_N = \min(t, \sup_{n \geq N} f_n)$ where t is a μ-test dominating f. 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 4 to g_N and h_N gives the result. More precisely, for every $x \in \mathsf{ML}_\mu$ and every N,

$$\liminf_n A_n^f(x) \geq \liminf_n A_n^{g_N}(x) = \int g_N \, \mathrm{d}\mu$$

$$\limsup_n A_n^f(x) \leq \limsup_n A_n^{h_N}(x) = \int h_N \, \mathrm{d}\mu,$$

so

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

3.1 Further results

We briefly discuss the extent to which the assumptions in the preceding results are needed.

3.1.1 Δ_2^0 functions

In Corollary 6 that one cannot get rid of the assumption that f is dominated by a μ-test, as a limit of uniformly computable functions may not be finite on all Martin-Löf random points, even if it is integrable. Let us give an example of such an f. Let μ be the uniform measure over $[0,1]$ and α be a Δ^0_2 random real (a Δ^0_2 real is a limit of a sequence of uniformly computable reals). Define f by $f(x)=\frac{1}{\sqrt{|x-\alpha|}}$ for $x\neq\alpha$ and $f(\alpha)=+\infty$. f is a limit of uniformly computable functions. Indeed, let a_n be a computable sequence of reals converging to α: for all x, $f(x)=\lim_{n\to\infty}\frac{1}{2^{-n}+\sqrt{|x-a_n|}}$.

3.1.2 Upper semi-computable functions

In Theorem 5 we do not know whether the assumption that the upper semi-computable function f is dominated by a test is really needed. Without this assumption, we still get a partial result.

▶ **Proposition 7.** *Let $f:\{0,1\}^{\mathbb{N}}\to[0,+\infty]$ be upper semi-computable. For every $x\in\mathsf{ML}_\mu$,*

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

Proof. The argument to prove inequality $\geq$ is essentially the one used in Theorem 5 to prove that $\limsup A^f_n(x)\leq\int f\,\mathrm{d}\mu$ when f is lower semi-computable. Let $q<\int f\,\mathrm{d}\mu$ be a rational number. By the classical ergodic theorem there exists $K\in\mathbb{N}$ such that the effective closed set $C_{q,K}:=\{x:\forall k\geq K, A^f_k(x)\geq q\}$ has positive measure. By Theorem 3, if $x\in\mathsf{ML}_\mu$ then there exists n such that $T^n(x)\in C_{q,K}$, which implies $\liminf A^f_k(x)=\liminf A^f_k(T^n(x))\geq q$. Taking q closer and closer to $\int f\,\mathrm{d}\mu$ we get $\liminf A^f_k(x)\geq\int f\,\mathrm{d}\mu$.

We now prove the other inequality. If $\int f\,\mathrm{d}\mu=+\infty$ we are done, otherwise let $q>\int f\,\mathrm{d}\mu$ be a rational number. By the classical ergodic theorem, for each $K\in\mathbb{N}$, $\mu(C_{q,K})=0$ so $C_{q,K}$ contains no μ-random point. As a result, $\liminf A^f_k(x)\leq\int f\,\mathrm{d}\mu$ for each $x\in\mathsf{ML}_\mu$. ◀

Observe that it is possible to build an integrable upper semi-computable function f that is not dominated by a μ-test.

3.1.3 Π^0_2 sets

By Corollary 6, if D is a Δ^0_2-set, i.e. if both D and its complement are effective intersections of effective open sets, then the visit frequency of the trajectory of a Martin-Löf random sequence into D is always $\mu(D)$. What can be said about sets of higher complexity in the effective Borel hierarchy?

Let Ω be a left-c.e. random sequence (a Chaitin's Ω). Let $D_n=T^n[\Omega,\Omega+2^{-2n-2}]$. D_n is Σ^0_2 (and even Δ^0_2), uniformly in n: $D_n=T^n(\bigcup_i[\Omega,1]\cap[0,q_i+2^{-2n-2}])$ where $q_i\nearrow\Omega$ is a computable sequence of dyadic numbers. One easily checks that $\mu(D_n)\leq 2^{-n-2}$. Let $D=\bigcup_n D_n$: D is Σ^0_2, $\mu(D)<1$ but all the iterates of Ω belong to D.

As a result, the complement of D is a Π^0_2-set of positive measure, but no iterate of Ω belongs to this set. We can conclude that the complement of D does not contain any effective closed set of positive measure. As shown by the following result, the converse is also true: if a Π^0_2-set of positive measure contains no effective closed set of positive measure then there is a random sequence whose iterates avoid this set.

▶ **Proposition 8.** *Let $D=\bigcap_n U_n$ where U_n are (not necessarily uniformly) effective open sets. Assume $\mu(D)>0$. The following are equivalent:*

1. *every random point eventually falls into D,*
2. *D contains an effective closed set of positive measure,*
3. *$\liminf A^{\mathbf{1}_D}_n(x)>0$ for every random point x.*

Proof. $1 \Rightarrow 2$. Let K_0 be the complement of some level in a universal ML test.

Assume D contains no effective closed set of positive measure. We construct a decreasing sequence K_n of effective closed sets of positive measure. As $T^n(K_n)$ has positive measure, it is not a subset of D so there is i_{n+1} such that $T^n(K_n) \setminus U_{i_{n+1}} \neq \emptyset$. Let $K_{n+1} = K_n \setminus T^{-n}U_{i_{n+1}}$: $K_{n+1} \neq \emptyset$ so $\mu(K_{n+1}) > 0$ as it is an effective closed set that contains random sequences. In the limit, $\bigcap_n K_n$ is non-empty and if $x \in \bigcap_n K_n$ then for every n, $T^n(x) \notin U_{i_{n+1}}$, so $T^n(x) \notin D$.

$2 \Rightarrow 3$. Direct: if $C \subseteq D$ is an effective closed set of positive measure then $\liminf A_n^{\mathbf{1}_D} \geq \liminf A_n^{\mathbf{1}_C} = \mu(C) > 0$ by Theorem 3

$3 \Rightarrow 1$. Obvious. ◀

3.2 Recurrence time

Let (X, T) be a dynamical system. Given a set A and a point $x \in A$, the recurrence time of x is defined as the minimal $k \geq 1$ such that $T^k(x)$ belongs to A. Ergodic theory provides several results about the asymptotic behavior of the recurrence time of a point, which have applications in coding and information theory. We focus on the particular case of the shift map on the Cantor space and when A is a cylinder.

For $x \in \{0,1\}^{\mathbb{N}}$, let $R_n(x) := \min\{k \geq n : x_0 \ldots x_{n-1} = x_k \ldots x_{x+n-1}\} = \min\{k \in n : x\restriction_n = T^k(x)\restriction_n\}$. Ornstein and Weiss [18] proved that for a shift-invariant ergodic probability measure μ, $\log R_n(x)/n$ converge to the entropy $h(\mu)$ almost-surely, extending the convergence in probability earlier proved by Wyner and Ziv [23]. Nakamura [16] proved a weak version of that result for Martin-Löf random points, showing that $\limsup \log R_n(x)/n = h(\mu)$ for every Martin-Löf μ-random sequence x.

Here we show that the full result can be simply derived from Theorem 3.

▶ **Theorem 9.** *Let μ be a shift-invariant ergodic measure. For every $x \in \mathsf{ML}_\mu$,*

$$\lim_{n\to\infty} \frac{\log R_n(x)}{n} = h(\mu).$$

Proof. Let $f_n(x) = \log R_n(x)/n$ and $q < h(\mu)$ be rational. Note that f_n is computable on ML_μ, uniformly in n. From Ornstein and Weiss result, the set $\{x : \exists N \forall n \geq N, f_n(x) \geq q\}$ has measure one, hence there exists N such that the set $C_N := \{x : \forall n \geq N, f_n(x) \geq q\}$ has positive measure. Let $x \in \mathsf{ML}_\mu$: as C_N is effectively closed, by Theorem 3 there exists k such that $T^k(x) \in C_N$, which implies $\liminf f_n(T^k(x)) \geq q$. One easily sees that $R_n(x) \geq R_{n-1}(T(x))$, which implies $\liminf f_n(x) \geq \liminf f_n(T^k(x)) \geq q$. As this inequality holds for each $q < h(\mu)$, we get $\liminf f_n(x) = h(\mu)$ for every $x \in \mathsf{ML}_\mu$. Together with Nakamura's result, it proves the theorem. ◀

4 The effective Shannon-McMillan-Breiman theorem

We now present our main result.

▶ **Theorem 10** (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 [21, 19]. 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 10.

An easy calculation shows that

$$-\log \mu[x\upharpoonright_n] = \sum_{k=0}^{n-1} f_{n-1-k} \circ T^k(x) \tag{3}$$

where

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

▶ **Lemma 11.** *$f_k(x)$ converge for each $x \in \mathsf{ML}_\mu$.*

Proof. Define the computable martingale

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

By the effective Doob's convergence theorem (see Theorem 7.1.3 on page 270 in [5]), 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

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

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

We will use the following lemma (Corollary 2.2 on page 261 in [19], Lemma 4.26 on page 26 in [21]).

▶ **Lemma 12.** *$f^* := \sup_k f_k \in L^1$.*

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

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

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 12 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 6, from which the result follows directly. ◀

▶ **Proposition 14.** *For each $x \in \mathsf{ML}_\mu$,*

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

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

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

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 $\int \tilde{g}_N \, d\mu$ by the Effective Ergodic Theorem 5. 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 $\int (f^* + f) \, d\mu$ by Corollary 6 so their limits cancel each other.

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

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

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

Acknowledgments

The author would like to thank the anonymous referees for their remarks and for pointing out Hochman's results.

References

1 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.

2 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.

3 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. To appear in *Information and Computation*. ArXiv 1007.5249, 2010.

4 Gregory J. Chaitin. A theory of program size formally identical to information theory. *J. ACM*, 22(3):329–340, 1975.

5 Rod Downey and Denis Hirschfeldt. *Algorithmic Randomness and Complexity*. Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2010.

6 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.

7 Péter Gács. Exact expressions for some randomness tests. *Z. Math. Log. Grdl. M.*, 26:385–394, 1980.

8 Michael Hochman. Upcrossing inequalities for stationary sequences and applications. *The Annals of Probability*, 37(6):2135–2149, 2009.

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

10 Ming Li and Paul M. B. Vitanyi. *An Introduction to Kolmogorov Complexity and Its Applications*. Springer-Verlag, Berlin, 1993.

11 Jack Lutz. Gales and the constructive dimension of individual sequences. In Ugo Montanari, José Rolim, and Emo Welzl, editors, *Automata, Languages and Programming*, volume 1853 of *Lecture Notes in Computer Science*, pages 902–913. Springer Berlin / Heidelberg, 2000.

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

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

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

15 Elvira Mayordomo. A Kolmogorov complexity characterization of constructive Hausdorff dimension. *Inf. Process. Lett.*, 84(1):1–3, 2002.

16 Masahiro Nakamura. Ergodic theorems for algorithmically random sequences. *Proceedings of the Symposium on Information Theory and Its Applications*, 28:71–74, 2005.

17 A. Nies. *Computability and randomness*. Oxford logic guides. Oxford University Press, 2009.

18 Donald S. Ornstein and Benjamin Weiss. Entropy and data compression schemes. *IEEE Transactions on Information Theory*, 39:78–83, 1993.

19 Karl Petersen. *Ergodic Theory*. Cambridge Univ. Press, 1983.

20 Claus-Peter Schnorr. Process complexity and effective random tests. *J. Comput. Syst. Sci.*, 7(4):376–388, 1973.

21 Meir Smorodinsky. *Ergodic Theory, Entropy*, volume 214 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin Heibelberg New York, 1971.

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

23 Aaron D. Wyner and Jacob Ziv. Some asymptotic properties of the entropy of a stationary ergodic data source with applications to data compression. *IEEE Transactions on Information Theory*, 35:1250–1258, 1989.