Open access • Book Chapter • DOI:10.1007/978-3-642-02927-1_46

Applications of Effective Probability Theory to Martin-Löf Randomness

Mathieu Hoyrup, Cristobal Rojas

Published on: 06 Jul 2009 - International Colloquium on Automata, Languages and Programming

Topics: Randomness, Computable function, Computability, Probability theory and Computable analysis

Applications of Effective Probability Theory to Martin-Löf Randomness

Mathieu Hoyrup, Cristobal Rojas

Applications of Effective Probability Theory to Martin-Löf Randomness

Mathieu Hoyrup and Cristóbal Rojas

Abstract

We pursue the study of the framework of *layerwise computability* introduced in [HR09a] and give three applications. (i) We prove a general version of Birkhoff's ergodic theorem for random points, where the transformation and the observable are supposed to be *effectively measurable* instead of *computable*. This result significantly improves [V'y97, Nan08]. (ii) We provide a general framework for deriving sharper theorems for random points, sensitive to the speed of convergence. This offers a systematic approach to obtain results in the spirit of Davie [Dav01]. (iii) Proving an effective version of a theorem of Ulam, we positively answer a question raised in [Fou08]: can random Brownian paths reach any random number? All this shows that *layerwise computability* is a powerful framework to study Martin-Löf randomness, with a wide range of applications.

Contents

1 Introduction

Algorithmic randomness emerged as an early achievement of Kolmogorov's program to base probability theory on the theory of computing. Yet a framework allowing the combination of these two theories is still lacking: for instance, computable analysis is mainly concerned with effective versions of topological notions, and not probabilistic/mesure-theoretic ones. For this reason, the study of algorithmic randomness has not reached its expected range of application: general probability theory. Let us recall the main contributions of algorithmic randomness to probability theory developed so far.

Theorems for random points. The main novelty brought by algorithmic randomness is that probabilistic laws can be strengthened in principle, holding at every random point and not only with probability one. Classical examples can be found in [ML66, Vov87, V'y97] for instance. When proving this kind of result the key hypothesis is the *computability* of the random variables involved. However, it is well-known that computability notions are the effective versions of topological ones (the computable functions are precisely the effectively continuous ones, the semi-decidable sets are precisely the effectively open sets, and so on). Hence the computability assumption on random variables is (i) inappropriate in principle, as probability theory is grounded on measure theory and not on topology; (ii) a priori too strong, as in the classical setting only properties as measurability, integrability are required. This leads to the following:

Problem 1. *Theorems for random points should hold for "effectively measurable" objects and not only computable ones.*

This problem has already been independently investigated in [Nan08, GHR10] where ergodic theorems for random points are proved for different types of "almost everywhere computable" functions. These works are, however, still far from catching the effective version of measurable functions. For instance in Birkhoff's ergodic theorem, nothing can be said about the mean sojourn time of algorithmically random points in fractal sets having effective constructions, as the *Smith-Volterra-Cantor* (or *fat Cantor*) set[1].

[1]The Smith-Volterra-Cantor set $A \subseteq [0,1]$ is homeomorphic to the Cantor set and has Lebesgue measure

Information given by the randomness degree. A further contribution of algorithmic randomness to probability theory consists in making use of the "randomness degree" of a random point x to get additional information about the way x satisfies a given probabilistic law. For instance in [Dav01], the speed of convergence in the Strong Law of Large Numbers is computed from the "compressibility coefficient" of each random sequence. This kind of result gives a much sharper insight into probabilistic phenomena and, we believe, new tools are needed in order to make this approach systematic and applicable on abstract spaces:

Problem 2. *Having a general framework to get sharper theorems for random points, using the information given by the randomness degree.*

Layerwise computability. In [HR09a], working in the context of computable probability spaces (to which Martin-Löf randomness has been recently extended, see [Gác05, HR09b]), effective versions of measure-theoretic notions were examined and another contribution of algorithmic randomness to probability theory was developed: the setting of a new framework for computability adapted to the probabilistic context. This was achieved by making a fundamental use of the existence of a universal Martin-Löf test to endow the space with what we call the *Martin-Löf layering*. In this new framework, which we call ***layerwise computability***, the *layerwise* versions of virtually all computability notions can be naturally defined. The contributions of this setting can be summarized in the following principle, supported by the main results in [HR09a]:

Correspondence Principle (CP)**.** *Under effectivity assumptions, measure-theoretic notions correspond exactly to* layerwise *versions of topological ones.*

Intuitively, this gives evidence that the *layering structure* grasps a large part of the probabilistic phenomena: each probabilistic notion, that by nature intimately depends on the underlying measure μ, can be expressed without referring to μ but only to its imprint on the space, captured by the *layering*. In this paper, elaborating on [HR09a] and developing layerwise computability further, we give solutions to Problems 1 and 2. The CP is at the core of these solutions, that we briefly present now.

Solution to Problem 1. We prove general versions of theorems for random points and *effectively measurable* random variables, in particular Birkhoff's ergodic theorem. This is a significant improvement of [Nan08, GHR10] as it implies in particular a positive result for the Smith-Volterra-Cantor set. To prove these results we develop tools allowing to adapt the existent techniques (used in the *computable* context) to the *layerwise computable* context. Then, the results for *effectively measurable* objects follow from the CP. This strategy is very general and applicable in a wide range of situations.

$\frac{1}{2}$.

Solution to Problem 2. As a further illustration of the CP we prove that under effectivity assumptions, *almost everywhere* convergence corresponds to the *layerwise* version of *uniform* convergence. This result gives evidence that the *layering* encodes information from which sharper results can be stated, providing a systematic approach to obtain results in the spirit of [Dav01]. In particular, we use it to compute the speed of convergence of random points (in a given *layering*) in both the Strong Law of Large Number and the Ergodic Theorem, in their general versions. The explicit connection between our framework and the results in [Dav01] is also given.

As a preliminary step, we use Martin-Löf randomness and an isomorphism theorem to derive an effective version of Ulam's theorem about tightness of probability measures. This gives yet another tool to solve problems related to algorithmic randomness, as it enables us to give a positive answer to a question raised in [Fou08] for algorithmically random Brownian motion (see Sect. 5.3).

In Sect. 2 we recall the background on computable probability spaces and Martin-Löf randomness and prove the effective version of an Ulam's result. In Sect. 3 we set the framework for *layerwise computability* and state the results relating it to effective measurability. In Sect. 4 we study the convergence of random variables from the effective point of view. We finish in Section 5 by applying all this machinery to obtain the general results announced above, giving solutions to Problems 1 and 2.

2 Preliminaries

2.1 Computable probability spaces

We work on the well-studied computable metric spaces (see [Wei00]).

Definition 2.1.1. A ***computable metric space*** is a triple $(X, d, \mathcal{S})$ where:

1. (X, d) is a separable metric space,
2. $\mathcal{S} = \{s_i : i \in \mathbb{N}\}$ is a countable dense subset of X with a fixed numbering,
3. $d(s_i, s_j)$ are uniformly computable real numbers.

$\mathcal{S}$ is called the set of ***ideal points***. If $x \in X$ and $r > 0$, the metric ball $B(x, r)$ is defined as $\{y \in X : d(x, y) < r\}$. The set $\mathcal{B} := \{B(s, q) : s \in \mathcal{S}, q \in \mathbb{Q}, q > 0\}$ of ***ideal balls*** has a canonical numbering $\mathcal{B} = \{B_i : i \in \mathbb{N}\}$. An ***effectively open set*** is an open set U such that there is a r.e. set $E \subseteq \mathbb{N}$ with $U = \bigcup_{i \in E} B_i$. A compact set K is ***effectively compact*** if the set $\{\langle i_1, \ldots, i_n\rangle : K \subseteq B_{i_1} \cup \ldots \cup B_{i_n}\} \subseteq \mathbb{N}$ is r.e. Let $K \subset X$. A set V is ***effectively open in*** K if there is an effective open set U such that $V \cap K = U \cap K$. A set V is ***decidable in*** K if V and $X \setminus V$ are effectively open in K. A function $f : X \to Y$ is ***computable on*** K if the preimages of effectively open sets are effectively open in K, in a uniform way. A real function $f : X \to [-\infty, +\infty]$ is ***lower semi-computable*** if the sets $f^{-1}(q_i, +\infty)$ are uniformly effectively open, it is ***upper semi-computable*** if $-f$ is lower semi-computable.

Any object that has some effectivity can be naturally encoded into a (possible more than one) integer, called its *Gödel number*.

Remark 2.1.1. Let K be effectively compact. It is not difficult to see that the complement $X \setminus K$ is an effective open set, uniformly in K, and that if U is an effective open set, then $K \setminus U$ is effectively compact, uniformly in U, K.

Several approaches to the computability of Borel probability measures have been proposed and happen to give the same notion, which can then be considered as a robust one.

Definition 2.1.2 (from [Eda96, Sch07, HR09b])**.** Let $(X, d, \mathcal{S})$ be a computable metric space. A Borel probability measure μ on X is ***computable*** if $\mu(B_{i_1} \cup \ldots \cup B_{i_n})$ are lower semi-computable, uniformly in $i_1, \ldots, i_n$.

Definition 2.1.3 (from [HR09b])**.** A ***computable probability space*** is a pair (X, μ) where X is a computable metric space and μ is a computable Borel probability measure on X.

2.1.1 Algorithmic randomness.

Martin-Löf randomness was first defined in [ML66] on the space of infinite symbolic sequences. Its generalization to abstract spaces has been investigated in [ZL70, HW03, Gác05, HR09b]. We follow the approaches [Gác05, HR09b] developed on any computable probability space (X, μ).

Definition 2.1.4. A ***Martin-Löf test (ML-test)*** V is a sequence of uniformly effective open sets V_n such that $\mu(V_n) < 2^{-n}$. A point x ***passes*** a ML-test V if $x \notin \bigcap_n V_n$. A point is ***Martin-Löf random (ML-random)*** if it passes all ML-tests. The set of ML-random points is denoted by ML_μ.

Theorem 2.1.1 (adapted from [ML66])**.** *Every computable probability space (X, μ) admits a universal Martin-Löf test, i.e. a ML-test U such that for all $x \in X$, x is ML-random $\iff$ x passes the test U. Moreover, for each ML-test V there is a constant c (computable from any Gödel number of V) such that $V_{n+c} \subseteq U_n$ for all n.*

From now and beyond, we fix a particular universal ML-test U. One can assume w.l.o.g. that $U_{n+1} \subseteq U_n$.

When the underlying space is complete, even if is is unbounded the finite character of probability measures makes the probabilistic phenomena concentrate in a small region. This is formally expressed by Ulam's theorem: on a complete separable metric every Borel probability measure is *tight*. We prove its effective version:

Theorem 2.1.2 (Effective Ulam's theorem)**.** *On a complete computable metric space, every computable Borel probability measure is effectively tight: the sets $K_n := X \setminus U_n$ are uniformly effective compact sets and $\mu(K_n) > 1 - 2^{-n}$.*

The proof of this result uses the following theorem taken from [HR09b] (Thm. 5.1.1 and Cor. 6.2.1).

Theorem 2.1.3. *For every (X, μ) computable probability space there is a computable measure ν on the Cantor space $\{0,1\}^{\mathbb{N}}$ such that ML_μ and ML_ν are computably homeomorphic, i.e. there is a computable bijection $F : ML_\nu \to ML_\mu$ with computable inverse. This homeomorphism pushes ν to μ.*

2.1.2 Effective measurability.

The following is an adaptation of [HR09a] to complete spaces. Let $[\mathfrak{S}]_\mu$ be the set of Borel subsets of X quotiented by the equivalence relation $A \sim_\mu B \iff d_\mu(A, B) := \mu(A \Delta B) = 0$. d_μ is a metric on $[\mathfrak{S}]_\mu$, which has a natural computable metric structure.

Definition 2.1.5. A Borel set A is ***μ-recursive*** if $[A]_\mu$ is a computable point of $[\mathfrak{S}]_\mu$. A measurable function $f : (X, \mu) \to Y$ is ***μ-recursive*** if there is a basis $\hat{\mathcal{B}} = \{\hat{B}_1, \hat{B}_2, \ldots\}$ of Y effectively equivalent to $\mathcal{B}$ such that $f^{-1}(\hat{B}_i)$ are uniformly μ-recursive.

Definition 2.1.6. A set A is ***effectively μ-measurable*** if there are uniformly effective compact sets C_n and open sets U_n such that $C_n \subseteq A \subseteq U_n$ and $\mu(U_n \setminus C_n) < 2^{-n}$. A function $f : X \to Y$ is ***effectively μ-measurable*** if there is a basis $\hat{\mathcal{B}} = \{\hat{B}_1, \hat{B}_2, \ldots\}$ of Y effectively equivalent to $\mathcal{B}$ such that $f^{-1}(\hat{B}_i)$ are uniformly effectively μ-measurable.

In the original definition the sets C_n are complements of effective open sets. When the space is complete, requiring C_n to be effectively compact gives the same notion, using effective tightness (Thm. 2.1.2) and Rmk. 2.1.1.

Theorem 2.1.4 (from [HR09a])**.** *1. A is μ-recursive $\iff$ A is equivalent to an effectively μ-measurable set.*

2. f is μ-recursive $\iff$ f is equivalent to an effectively μ-measurable function.

3 Layerwise computability

Now we enter in the main novelty of this article. With effective versions of measure-theoretic notions at our disposal, one can hope to solve Problem 1. However the notions developed so far are difficult to handle (and rather heavy, see Def. 2.1.6) or even not well-defined on algorithmically random points (see Def. 2.1.5). It was demonstrated in [HR09a] that algorithmic randomness and the universal test offer an alternative elegant way of handling effective measurability notions. Let (X, μ) be a *complete* computable probability space. It comes with a canonical universal ML-test U_n, with $U_{n+1} \subseteq U_n$. We proved that the sets $K_n := X \setminus U_n$ are uniformly effective compact sets (Thm. 2.1.2). Hence the set of ML-random points is *layered* by an increasing sequence of effective compact sets: $\mathrm{ML} = \bigcup_n K_n$.

Definition 3.0.7 (Martin-Löf Layering)**.** Let (X, μ) be a computable probability space. We call the sequence $(K_n)_{n\in\mathbb{N}}$ the ***Martin-Löf layering*** of the space.

Definition 3.0.8 (Layerwise computability notions)**.**

1. A set $A \subseteq X$ is ***layerwise semi-decidable*** if for all n, A is effectively open on K_n, uniformly in n, i.e. there are uniformly effective open sets U_n such that $A \cap K_n = U_n \cap K_n$,

2. A set $A \subseteq X$ is ***layerwise decidable*** if for all n, A is decidable on K_n, uniformly in n, i.e. both A and $X \setminus A$ are layerwise semi-decidable,

3. A function $f : (X, \mu) \to Y$ is ***layerwise computable*** if for all n, f is computable on K_n, uniformly in n, i.e. $f^{-1}(B_i)$ are uniformly layerwise semi-decidable.

More generally, every computability (or *effective topological*) notion has its layerwise counterpart[2]. For instance, the layerwise counterpart of effective uniform convergence (i.e, in the sup norm) of functions will be examined in section 4. In general, layerwise computability is not stable under composition, simply because layerwise computable functions may not preserve randomness. This can be overcome under measure presevation:

Proposition 3.0.1 (from [HR09a]). *Let (X, μ) be a computable probability space and $T : X \to X$ a layerwise computable function which preserves μ (i.e. $\mu \circ T^{-1} = \mu$). Then:*

1. *T preserves ML-randomness. Moreover there is a contant c such that $T(K_n) \subseteq K_{n+c}$ for all n.*

2. *If $f : X \to Y$ is layerwise computable then so is $f \circ T$, uniformly in f and T.*

3.1 Relation with effective measurability

If $T : X \to X$ is measurable and preserves μ, then $T^{-1} : [\mathfrak{S}]_\mu \to [\mathfrak{S}]_\mu$ is well-defined and continuous. In [HR09a], we prove the following equivalences:

Theorem 3.1.1. *Let $A \subseteq X$ be a set, $f : X \to Y$ a function and $T : X \to X$ a measure-preserving function.*

1. *A is effectively μ-measurable $\iff$ A is layerwise decidable.*

2. *$f : X \to Y$ is effectively μ-measurable $\iff$ f is layerwise computable.*

3. *T^{-1} is computable $\iff$ T is equivalent to a layerwise computable function.*

Therefore, under effectivity assumptions measure-theoretical notions are the layerwise versions of topological ones. Observe that the latter are expressed without referring to μ but only to the Martin-Löf layering. In other words, the layering catches the essential part of the probabilistic features. This is the first illustration of the Correspondence Principle (see Introduction).

[2]Layerwise computable functions should not be confused with compact-computable functions as defined in [Mor01]. Indeed, the compact sets involved in [Mor01] are not only effectively compact but also *effectively closed*, i.e. they contain a dense sequence of computable points (which is of course not the case for the K_n's, as random points are generally not computable).

3.2 Layerwise tests

We now state the theorem which will allow to solve Problem 1: making theorems on random points hold under effective measurability assumptions. The point is that some classical results underlying the practice of algorithmic randomness have – surprisingly – they layerwise counterpart.

Definition 3.2.1. A ***layerwise Martin-Löf test*** A is a sequence of uniformly layerwise semi-decidable sets A_n such that $\mu(A_n) < 2^{-n}$. A ***layerwise integrable test*** is a layerwise lower semi-computable function $t : X \to [0, +\infty]$ such that $\int t \, \mathrm{d}\mu < \infty$.

Theorem 3.2.1. *Let U be a layerwise semi-decidable set, A a layerwise ML-test and t a layerwise integrable test.*

1. *If $\mu(U) = 1$ then $ML_\mu \subseteq U$.*
2. *If x is ML-random, then $x \notin \bigcap_n A_n$. Moreover, there is a constant c (computable from a Gödel number of the sequence A) such that $A_{n+c} \cap K_n = \emptyset$ for all n.*
3. *If x is ML-random, then $t(x) < \infty$. Moreover, there is a constant c such that $t < 2^{n+c}$ on K_n.*

4 Convergence of random variables

In [V'y97] the following result for convergence of random variables on random points is stated: if computable functions converge almost everywhere in an effective way then they converge on ML-random points. Here we improve this in several ways:

- using layerwise tests, we weaken the hypothesis: the functions are now assumed to be effectively measurable only, which gives a solution to Problem 1,
- using the layering, we get information about the speed of convergence on random points, providing a solution to Problem 2,
- under effectivity assumptions, we get a characterization of a probabilistic notion (namely, almost everywhere convergence) as the layerwise version of a topological one (namely, uniform convergence), which further illustrate the Correspondence Principle, beyond Theorem 3.1.1.
- we give other results for random points under different types of assumptions on the convergence of the sequence.

Observe that what follows works on any computable probability space (algorithmic randomness was only developed on the Cantor space when [V'y97] was written). Let $f_i : X \to \mathbb{R}$ be a sequence of random variables and f another random variable (expected to be the limit of f_i). Let $D_n(\delta) := \{x : \exists i \geq n, |f_i - f| > \delta\}$. It is a standard observation that the sequence f_i converge almost everywhere to f if and only if the measure of the sets $D_n(\delta)$ tends to zero, for each δ. This motivates the following:

Definition 4.0.2. Functions f_n converge ***effectively almost everywhere (effectively a.e.)*** if $\mu(D_n(\delta))$ converge to 0, effectively from δ. In other words there is a computable function $n(\delta, \varepsilon)$ such that $\mu(D_{n(\delta,\varepsilon)}(\delta)) < \varepsilon$.

As already said, V'yugin [V'y97] proved that if f_n are uniformly *computable* functions that converge effectively a.e. then they converge at each ML-random point. Actually, the result also holds when the functions f_n are uniformly *effectively μ-measurable*. Indeed, for each δ the sets $D_{n(\delta,2^{-n})}(\delta)$ form a layerwise ML-test hence by Thm. 3.2.1 we directly get the convergence at every ML-random point. We can even go further: the layering gives information about the speed of convergence at *each random point*. By Thm. 3.2.1 again there is a constant c (computable from a Gödel number of the sequence f_n and the function $n(\delta, \varepsilon)$) such that, for $x \in K_n$ we have $x \notin D_N(\delta)$, with $N = n(\delta, 2^{-n-c})$. In other words, if we know that $x \in K_n$ we can compute a number N such that for all $i \geq N$ we have $|f_i(x) - f(x)| < \delta$. This motivates the following, which is the layerwise version of effective convergence for the uniform norm:

Definition 4.0.3. Functions f_i ***converge layerwise effectively uniformly*** to f if for each k, the restrictions of f_i to K_k converge to the restriction of f to K_k for the uniform norm, effectively from k. In other words, there is a computable function $n(\delta, k)$ such that $\|f_i - f\|_{K_k} := \sup_{K_k} |f_i - f| \leq \delta$ for all $i \geq n(\delta, k)$.

In the same way that uniform convergence implies pointwise convergence, such functions converge on each ML random point.

Proposition 4.0.1. *If f_i are uniformly layerwise computable functions that converge layerwise effectively uniformly to f then f is layerwise computable.*

As said above, effective a.e. convergence implies layerwise effective uniform convergence. Actually this is a characterization, which provides another illustration of the Correspondence Principle:

Theorem 4.0.2. *Let f_n be uniformly effectively μ-measurable functions. Then f_n converge effectively a.e. if and only if f_n converge layerwise effectively uniformly.*

At the same time, this result gives evidence that layerwise computability is a solution to Problems 1 and 2: this convergence for random points holds for *effectively μ-measurable* functions and not only *computable* ones, and the speed of convergence can be computed from the layer a random point belongs to.

Corollary 4.0.1. *If f_i are uniformly effectively μ-measurable functions that converge effectively a.e. to f, then f is effectively μ-measurable.*

The simplicity of this proof shall be very general, as soon as the Correspondence Principle holds: a result about effective measure-theoretical notions can be split into two parts (i) the layerwise version of the corresponding result for effective topological (i.e. computability) notions (as Prop. 4.0.1) and (ii) the characterizations of effective measure-theoretical notions as layerwise topological ones (as Thm. 3.1.1 and 4.0.2).

4.0.1 Monotone convergence.

The sequence f_i is ***monotonic*** if $f_{i+1} \leq f_i$ for all i. When the sequence is monotonic Thm. 4.0.2 can be proved with weaker effectivity assumptions:

Proposition 4.0.2. *Let f_i be a monotonic sequence of uniformly layerwise lower semi-computable. Then the following statements are equivalent:*

1. *f_i converge effectively a.e. to 0,*
2. *f_i converge layerwise effectively uniformly to 0.*

It is an interesting fact that for monotonic upper semi-computable sequences, the almost everywhere convergence to 0 is always effective:

Proposition 4.0.3. *Let f_i be a monotonic sequence of uniformly layerwise upper-computable functions which converge a.e. to 0. Then f_i converge* effectively *a.e. to 0.*

4.0.2 Non-effective convergence.

When the convergence a.e. is not effective we can still say something concerning random points.

Theorem 4.0.3. *Let f_n, f be uniformly layerwise computable functions, and c some (not necessarily computable) constant.*

- *If f_n converge a.e. to a constant c then $\liminf f_n(x) \leq c \leq \limsup f_n(x)$ for all $x \in ML$.*
- *If f_n converge a.e. to a layerwise computable function f, then $\liminf f_n(x) \leq f(x) \leq \limsup f_n(x)$ for all $x \in ML$.*

5 Applications

5.1 Ergodic theorems for effectively measurable functions

We now apply the developed tools to solve Problem 1 for ergodic theorems, namely Poincaré recurrence theorem and Birkhoff's ergodic theorem. Let us first recall the first version of Birkhoff's theorem for random points, proved by V'yugin [V'y97].

Theorem 5.1.1 (Ergodic theorem for random points, V'yugin)**.** *Let μ be a computable probability measure on $X = \{0,1\}^{\mathbb{N}}$, $T : X \to X$ a computable measure-preserving map and $f \in L^1(X, \mu)$ a computable function (called observable). Then:*

(i) For every ML-random sequence x, the limit $\overline{f}(x) := \lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} f \circ T^i(x)$ exists.

(ii) If the system is moreover ergodic, then $\overline{f}(x) = \int f \, \mathrm{d}\mu$ for every ML-random x.

This theorem embodies several probabilistic laws. For instance, if we consider the system to be (σ, λ) where σ is the shift transformation[3] and λ is the uniform measure, and take as observable the indicator function $\vec{1}_{[1]}$ (which is computable since the cylinder [1] is a *decidable* set), we get the Law of Large numbers for random sequences. Let us point out some facts about the hypothesis of this theorem:

1. It is established for Cantor space $X = \{0,1\}^{\mathbb{N}}$,

2. the transformation T and the observable f are supposed to to be *computable*.

The first point is not a real restriction, as the proof for general spaces remains unchanged. However the second condition is an unnatural restriction as the classical Birkhoff's result, belonging to the measure-theoretic setting, is stated for measurable functions whereas *computability* corresponds to the effective version of *continuity*. Moreover, when passing to general (usual) spaces, this restriction becomes much more important since the theorem cannot be applied to indicators of sets anymore (in connected spaces only trivial sets are decidable). In [Nan08], the Ergodic theorem is extended to include functions having some discontinuities at computable points[4]. A further step is given in [GHR10] where the result is proved to hold for the indicator functions of every (not necessarily constructive) set of continuity[5]. Yet, nothing can be said about some natural sets having effective constructions, like the ***Smith-Volterra-Cantor set*** (or *fat Cantor set*) whose Lebesgue measure is $\frac{1}{2}$ but has empty interior, so it is not a continuity set.

In what follows we give a definite solution by proving the Ergodic Theorem for *effectively measurable* functions. In particular, indicator functions of sets like the fat Cantor set fall in this class. As an introductory result, let us do so with Poincaré recurrence theorem first, whose proof is simpler.

Theorem 5.1.2 (Poincaré recurrence theorem for random points)**.** *Let (X, μ) be a computable probability space, $T : X \to X$ an effectively μ-measurable ergodic measure-preserving map and A a layerwise semi-decidable set with positive measure. Then every ML-random point falls infinitely often in A by iteration of T.*

Theorem 5.1.3 (Ergodic theorem for random points)**.** *Let (X, μ) be a computable probability space, $T : X \to X$ an effectively μ-measurable measure-preserving map and $f \in L^1(X, \mu)$ be an effectively μ-measurable observable. Then:*

(i) For every ML-random point x, the limit $\overline{f}(x) := \lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} f \circ T^i(x)$ exists.

(ii) If the system is moreover ergodic, then $\overline{f}(x) = \int f \, \mathrm{d}\mu$ for every ML-random x.

Proof. (i) The tools developed in section 3.2 allow to prove the first point by following the original V'yugins's proof. Given f and T, he constructs an integrable function $\sigma(T, f, x)$

[3] Defined by $(\sigma(\omega))_i = \omega_{i+1}$

[4] These functions belong to the class of Braverman's "graph computable functions".

[5] A set A is a set of μ-continuity if its boundary has μ-measure zero.

which is finite iff $\frac{1}{n}\sum_{i=0}^{n-1} f\circ T^i(x)$ converge. When f and T are computable, σ is lower semi-computable hence it is an integrable test, which proves the convergence at each random point. Actually the definition of σ makes sense on any computable probability space. When f and T are only effectively measurable, they are *layerwise* computable by Thm. 3.1.1 hence all $f\circ T^n$ are uniformly layerwise computable (Prop. 3.0.1). In this case, σ is then *layerwise* lower semi-computable so it is a layerwise integrable test; as a result, it is finite on all ML-random points by Thm. 3.2.1, from which (i) follows. (ii) By Thm. 4.0.3, the limit is the expected limit $\int f \,\mathrm{d}\mu$ when the system is ergodic. □

5.2 Layerwise computable speed of convergence on random points

In this section we show how the framework developed so far provides a solution to Problem 2. Let us first recall some results established by Davie [Dav01].

5.2.1 Davie's results.

To state them some background is needed first. On the Cantor space, implicitely endowed with the uniform measure λ, the *compressibility coefficient* of a sequence ω is defined as $d_\lambda(\omega) = \sup_n\{n - H(\omega_{1:n})\}$ where $H(w)$ is the *Kolmogorov-Chaitin complexity* of the finite word w. A fundamental result from algorithmic randomness and information theory is that a sequence is ML-random w.r.t. λ if and only if $d_\lambda(\omega)$ is finite. Davie defines $K^c := \{\omega : d_\lambda(\omega) \leq c\}$ and proves:

Theorem 5.2.1 (Davie, 2001)**.** *If A_i are uniformly effective open sets such that $\sum_i \mu(A_i)$ is a finite computable real number, then there is a computable function $n(c)$ such that for all $\omega \in K^c$ and all $m > n(c)$, $\omega \notin A_m$.*

Theorem 5.2.2 (Davie, 2001)**.** *There is a computable function $n(c,\varepsilon)$ such that for all $\omega \in K^c$ and all $n > n(c,\varepsilon)$, $\left|\frac{S_n(\omega)}{n} - \frac{1}{2}\right| < \varepsilon$ where $S_n(\omega)$ is the number of ones in the prefix of ω of length n.*

The equivalence between the paradigm of effective measure theory (Martin-Löf's approach) and the paradigm of compressibility (Chaitin's approach) is a strong non-trivial result, partly based on the technical *coding theorem*. Davie's results follow this line as they relate the compressibility coefficient of a sequence to the way the sequence satisfies a probability law, and thus their proofs consist in a finer use of the coding theorem. In our framework, we stay on the side of effective measure theory. In this way, the relation between the layer a random point belongs to and the way it satisfy laws is much simpler to derive, as it is essentially already contained in the existing proofs. This provides a solution to Problem 2. At the same time, as layerwise computability provides a solution to Problem 1 too, our results hold for effectively measurable sets/functions. As an illustration, we first state here the refined version of classical results in algorithmic randomness due to Solovay. The proofs are straightforward combinations of the usual proofs together with Thm. 3.2.1. Note that the first one is the generalization of Thm. 5.2.1 due to Davie.

Proposition 5.2.1 (Borel-Cantelli 1)**.** *There is a computable function $n(c,p)$ such that if A_n are uniformly layerwise semi-decidable sets such that $\alpha := \sum_n \mu(A_n)$ is finite and computable, then there is a constant c, computable from a Gödel number of the sequence A_n and α, such that if $x \in K_p$ then $x \notin A_n$ for all $n \geq n(c,p)$.*

We can also get a weaker result when the sum is not computable.

Proposition 5.2.2 (Borel-Cantelli 2)**.** *Let A_i be uniformly layerwise semi-decidable sets such that $\sum_i \mu(A_i) < \infty$. There is c, computable from a description of the sequence A_i, such that avery x in K_n falls in the A_i's at most 2^{n+c} times.*

5.2.2 (Very) Strong Law of Large Numbers.

We can now easily prove:

Theorem 5.2.3. *Let $X_i : (X,\mu) \to \mathbb{R}$ be i.i.d. effectively μ-measurable random variables such that $\int |X_i|^4 \,\mathrm{d}\mu < +\infty$. Let $S_n := X_0 + \cdots + X_{n-1}$. Hence, there is a computable function $n(c,\varepsilon)$ such that if $x \in K_c$ then for all $n > n(c,\varepsilon)$, $\left|\frac{S_n(x)}{n} - \int X_0 \,\mathrm{d}\mu\right| < \varepsilon$.*

Proof. All what we need is the convergence in the SLLN to be *effective*. This easily follow from the classical estimate (see [Bil79] for instance) $\mu\{x : \exists i \geq n, |n^{-1}S_n(x) - \int X_0 d\mu| > \delta\} \leq \frac{C}{(n-1)\delta^4}$, where C is a constant independent of n and δ. The result now follows from Theorem 4.0.2.

□

5.2.3 Effective convergence in Birkhoff's theorem

The convergence of the Birkhoff averages is not effective in general. In [V'y97], on the Cantor space V'yugin builds a computable probability measure which is invariant under the shift transformation, and such that the convergence of the averages of $\vec{1}_{[1]}$ is not effective. This measure is an infinite combination of ergodic measures and it is still an open question if a computable ergodic measure could be built for which the convergence is not effective.

However, in [GHR09] it is shown that for a class of ergodic systems, the convergence in Birkhoff theorem is effective. Let us recall that a system is ergodic if and only if for any two integrable functions f and g, the quantity $\gamma_n(f,g) := \left|\frac{1}{n}\sum_{i<n} \int f \circ T^i . g \,\mathrm{d}\mu - \int f \,\mathrm{d}\mu \int g \,\mathrm{d}\mu\right|$ goes to 0. A system is said to be **$\ln^2$-*ergodic for* (f,g)**, if there is a constant $c_{f,g} > 0$ such that $\gamma_n(f,g) \leq \frac{c_{f,g}}{(\ln(n))^2}$ for all $n \geq 2$.

Theorem 5.2.4 (Ergodic theorem for random points)**.** *Let (X,μ) be a computable probability space, $T : X \to X$ be an effectively measurable measure-preserving map and $f \in L^1(X,\mu)$ be an effectively measurable function. If T is $\ln^2$-ergodic for (f,f), then there is a computable function $n(c,\varepsilon)$ such that if $x \in K_c$ then for all $n > n(c,\varepsilon)$, $\left|\frac{1}{n}\sum_{i<n} f(x) \circ T^i(x) - \int f \,\mathrm{d}\mu\right| < \varepsilon$.*

5.2.4 Relation between K_n and K^n

Let X be the Cantor space endowed with a computable Borel probability measure μ. The compressibility coefficient can be adapted to μ: it is known as the *randomness deficiency* $d_\mu(\omega) := \sup_n\{-\log \mu[\omega_{1:n}] - H(\omega_{1:n})\}$. This function is known to be the logarithm of a universal integrable μ-test, which means that for every integrable μ-test t there is a constant a such that $\log t \leq a + d_\mu$. On the other hand, every computable probability space admits a universal integrable test t_μ (see [Gác05, HR09b]). Generalizing Davie, let us define $K^c := \{x : t_\mu(x) \leq 2^c\}$. As $\mathrm{ML}_\mu = \bigcup_c K^c$, the sequence $(K^c)_{c\in\mathbb{N}}$ can be used as an alternative layering and underly alternative versions of Def. 3.0.8 and 4.0.3. Actually, this would lead to the same notions. Indeed, using classical results from algorithmic randomness and information theory (see [Gác79, LV93, Gác]), it can be proved that there is a constant c such that $K_n \subseteq K^{n+c}$ and $K^n \subseteq K_{n+2\log n+c}$ for all n. Hence K^n are also uniformly effective compact sets and all layerwise computability notions relative to K^n are equivalent to the notions relative to K_n.

5.3 An application to Brownian motion

The study of Brownian motion from the algorithmic randomness point of view is carried out in [Fou00, Fou08]. Algorithmically random paths, called *complex oscillations* as they are defined in terms of Kolmogorov-Chaitin complexity, are the Martin-Löf random points of the computable probability space $(\mathcal{C}([0,1]), W)$, where $\mathcal{C}([0,1])$ is the space of continuous functions $x : [0,1] \to \mathbb{R}$ with the uniform norm and W is the Wiener probability measure. In [Fou00] it is proved that if $t \in [0,1]$ is computable and x is a complex oscillation then $x(t)$ is not computable. At the end of [Fou08] the following question is raised: can it be lower semi-computable?

We say that $y \in \mathbb{R}$ is λ-ML-random if $y = n + z$ where $n \in \mathbb{Z}$ and $z \in [0,1]$ is ML-random w.r.t. the Lebesgue measure λ on $[0,1]$. As noticed in [Fou08], it is a corollary of [KHN07] that $x(t)$ is actually λ-ML-random. But then can it be a Chaitin's Ω (which are lower semi-computable λ-ML-random reals)? The compactness of the layers (Thm. 2.1.2) enables us to give a positive answer. Indeed, Prop. 3.0.1 can be reinforced using Thm. 2.1.2:

Proposition 5.3.1. *Let (X,μ) and (Y,ν) be computable probability spaces such that X is complete. Let $T : X \to Y$ be a layerwise computable function which maps μ to ν. Then $T(\mathrm{ML}_\mu) = \mathrm{ML}_\nu$, i.e. T preserves randomness but it is also* onto.

Now, given a computable $t \in [0,1]$, the function $T_t : \mathcal{C}([0,1]) \to \mathbb{R}$ mapping x to $x(t)$ is computable. It pushes the Wiener measure W to a gaussian measure G. As G has bounded density w.r.t. the uniform measure and vice versa, ML_G is exactly the set of λ-ML-random reals. Hence,

Corollary 5.3.1. *Let x be a complex oscillation. For each computable $t \in [0,1]$, $x(t)$ is λ-ML-random. Moreover, given any λ-ML-random y and any computable t, there exists a complex oscillation x such that $x(t) = y$.*

References

[Bil79] Patrick Billingsley. *Probability and Measure*. John Wiley, New York, 1979. 13

[Dav01] George Davie. The Borel-Cantelli lemmas, probability laws and kolmogorov complexity. *Annals of Probability*, 29(4):1426–1434, 2001. 1, 3, 4, 12

[Eda96] Abbas Edalat. The Scott topology induces the weak topology. In *LICS '96: Proceedings of the 11th Annual IEEE Symposium on Logic in Computer Science*, page 372, Washington, DC, USA, 1996. IEEE Computer Society. 5

[Fou00] Willem L. Fouché. Arithmetical representations of brownian motion i. *J. Symb. Log.*, 65(1):421–442, 2000. 14

[Fou08] Willem L. Fouché. Dynamics of a generic brownian motion: Recursive aspects. *Theor. Comput. Sci.*, 394(3):175–186, 2008. 1, 4, 14

[Gác] Peter Gács. Lecture notes on descriptional complexity and randomness. http://www.cs.bu.edu/faculty/gacs/papers/ait-notes.pdf. 14

[Gác79] Péter Gács. Exact expressions for some randomness tests. In *Proceedings of the 4th GI-Conference on Theoretical Computer Science*, pages 124–131, London, UK, 1979. Springer-Verlag. 14

[Gác05] Peter Gács. Uniform test of algorithmic randomness over a general space. *Theoretical Computer Science*, 341:91–137, 2005. 3, 5, 14

[GHR09] Stefano Galatolo, Mathieu Hoyrup, and Cristobal Rojas. A constructive Borel-Cantelli lemma. Constructing orbits with required statistical properties. *Theor. Comput. Sci.*, 410(21-23):2207–2222, 2009. 13

[GHR10] Stefano Galatolo, Mathieu Hoyrup, and Cristóbal Rojas. Effective symbolic dynamics, random points, statistical behavior, complexity and entropy. *Information and Computation*, 2010. To appear. 2, 3, 11

[HR09a] Mathieu Hoyrup and Cristobal Rojas. An application of Martin-Löf randomness to effective probability theory. In *CiE*, pages 260–269, 2009. 1, 3, 6, 7

[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, 5, 14

[HW03] Peter Hertling and Klaus Weihrauch. Random elements in effective topological spaces with measure. *Information and Computation*, 181(1):32–56, 2003. 5

[KHN07] Bjørn Kjos-Hanssen and Anil Nerode. The law of the iterated logarithm for algorithmically random brownian motion. In *LFCS*, pages 310–317, 2007. 14

[LV93] Ming Li and Paul M. B. Vitanyi. *An Introduction to Kolmogorov Complexity and Its Applications*. Springer-Verlag, Berlin, 1993. 14

[ML66] Per Martin-Löf. The definition of random sequences. *Information and Control*, 9(6):602–619, 1966. 2, 5

[Mor01] Takakazu Mori. Computabilities of fine-continuous functions. In *CCA '00: Selected Papers from the 4th International Workshop on Computability and Complexity in Analysis*, pages 200–221, London, UK, 2001. Springer-Verlag. 7

[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, 2, 3, 11

[Sch07] Matthias Schröder. Admissible representations of probability measures. *Electronic Notes in Theoretical Computer Science*, 167:61–78, 2007. 5

[Vov87] Vladimir G. Vovk. The law of the iterated logarithm for random kolmogorov, or chaotic, sequences. *Theory of Probability and Applications*, 32:413–425, 1987. 2

[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. 1, 2, 8, 9, 10, 13

[Wei00] Klaus Weihrauch. *Computable Analysis*. Springer, Berlin, 2000. 4

[ZL70] A.K. Zvonkin and L.A. Levin. The complexity of finite objects and the development of the concepts of information and randomness by means of the theory of algorithms. *Russian Mathematics Surveys*, 256:83–124, 1970. 5

A Proofs

Proof of Thm. 2.1.2

First, this is true on the Cantor space C with a computable Borel probability measure ν: C is known to be effectively compact, so $K_n^\nu := C \setminus U_n$ is effectively compact, uniformly in n (Rmk. 2.1.1). Now, we use Thm. 2.1.3, which provides a computable function $F : \mathrm{ML}_\nu \to \mathrm{ML}_\mu$. The sets $K'_n := F(K_n^\nu)$ are then uniformly effectively compact, with measures $\mu(K'_n) > 1 - 2^{-n}$. By the universality of U_n, there is a constant c such that $X \setminus K'_{n+c} \subseteq U_n$ for all n. In other words, $K_n = X \setminus U_n \subseteq K'_{n+c}$ so $K_n = K'_{n+c} \setminus U_n$ is effectively compact, uniformly in n.

Proof of Thm. 3.2.1

We use the following corollary of Kurtz' theorem: if K is an effective compact set which has measure zero, then it contains no ML-random point.

1. Let $n \in \mathbb{N}$. There is an effectively open set U_n such that $A \cap K_n = U_n \cap K_n$. As $\mu(A) = 1$, $\mu(K_n \setminus U_n) = \mu(K_n \setminus A) = 0$. But $K_n \setminus U_n$ is an effective compact set so it contains no random point. As $K_n \subseteq \mathrm{ML}_\mu$ it follows that $K_n \setminus U_n = \emptyset$, so $K_n \subseteq A$. And this is true for every n.

2. Let U_n^p be uniformly effective open sets such that $A_p \cap K_n = U_n^p \cap K_n$. Let $C_p = K_p \setminus A_p = K_p \setminus U_p^p$: this is an effective compact set and $\mu(C_p) > 1 - 2^{-p+1}$, so by using universality of K_p there is a constant c such that $K_p \subseteq C_{p+c}$ for all p. As a result, $K_p \cap A_{p+c} = K_p \setminus C_{p+c} = \emptyset$.

3. We can suppose w.l.o.g. that $\int t \, \mathrm{d}\mu < 1$. Let $A_n = \{x : t(x) > 2^n\}$. As usual, $\mu(A_n) < 2^{-n}$. Now, $A_n = t^{-1}(2^n, +\infty]$ is a layerwise semi-decidable set by definition of layerwise lower computability. So A_n is a layerwise ML-test, so there is c such that $A_{n+c} \cap K_n = \emptyset$. Hence, $t \leq 2^{n+c}$ on K_n.

Proof of Thm. 4.0.2

First, we need some preparation. Given a sequence of random variables f_i, we define the deviation sets:

$$A_{i,j}(\delta) := [|f_i - f_j| > \delta] \quad \text{and} \quad D_n(\delta) := \bigcup_{i,j \geq n} A_{i,j}(\delta) \ .$$

If f is another function (expected to be their limit in some sense), we also define:

$$A_i^f(\delta) := [|f_i - f| > \delta] \quad \text{and} \quad D_n^f(\delta) := \bigcup_{i \geq n} A_i^f(\delta) \ .$$

We remark that one could replace $D_n(\delta)$ by $D_n^f(\delta) := \{x : \exists i \geq n, |f_i - f| > \delta\}$ where f is the almost sure limit. Indeed, one easily has $D_n^f(\delta) \subseteq D_n(\delta) \subseteq D_n^f(\delta/2) \mod 0$. Hence,

- let $n(\delta, \varepsilon)$ be adapted to D: it is adapted to D^f.

- let $n(\delta,\varepsilon)$ be adapted to D^f: $n(\delta/2,\varepsilon)$ is then adapted to D.

We now prove the result. We first use Thm. 3.1.1: the functions f_n are uniformly layerwise computable. Suppose f_n converge effectively a.e. Let $U_k := \bigcup_{p\geq 1} D_{n(2^{-p},2^{-k-p})}(2^{-p})$. This is a layerwise semi-decidable set and, by definition of $n(\delta,\varepsilon)$, $\mu(U_k) \leq \sum_{p\geq 1} 2^{-k-p} \leq 2^{-k}$. By Thm. 3.2.1 there is c such that for all k, $K_k \cap U_{k+c} = \emptyset$. Let $m(p,k) = n(2^{-p}, 2^{-k-p-c})$: it is a computable function and if $x \in K_k$, $x \notin U_{k+c}$ which means that for all $p \geq 1$, for all $i,j \geq m(p,k)$, $|f_i(x) - f_j(x)| \leq 2^{-p}$. Hence, for all p,k,

$$\|f_i - f_j\|_{K_k} \leq 2^{-p} \quad \text{for all } i,j \geq m(p,k).$$

Conversely, suppose that there is a computable function $n(\delta,k)$ such that for all $i,j \geq n(\delta,k)$, $\|f_i - f_j\|_{K_k} < \delta$. Let $m(\delta,\varepsilon)$ be any computable function such that $m(\delta,\varepsilon) = n(\delta,k)$ with $2^{-k} < \varepsilon$. One has $K_k \cap D_{m(\delta,\varepsilon)}(\delta) = \emptyset$, so $\mu(D_{m(\delta,\varepsilon)}(\delta)) < 1 - \mu(K_k) < 2^{-k} < \varepsilon$.

Proof of Prop. 4.0.2

(1)⇒ **(2)** Suppose $f_i \to 0$ effectively a.e. There is a computable function $m(k)$ such that

$$\mu\left(\left\{x : f_{m(k)}(x) > 2^{-k}\right\}\right) < 2^{-k} \quad \text{for all } k.$$

Let U_k be this layerwise effective open set. There exists a constant c such that for all k, $K_k \cap U_{k+c} = \emptyset$. Hence $\|f_n\|_{K_k} \leq 2^{-k-c}$ for all $n \geq m(k+c)$.
(2) ⇒ **(1)** The same as in Theorem 4.0.2.

Proof of Thm. 4.0.3

Let $x \in X$: $\liminf f_n(x) \leq c \iff \forall q > c, k, \exists n > k, f_n(x) < q$. So for each $q > c$ and k, $\mu(\bigcup_{n>k}\{x : f_n(x) < q\}) = 1$. When q is rational, this is moreover a layerwise semi-decidable set, so it contains all ML-random points. As this is true for every rational $q > 0$ and every k, $\liminf f_n(x) \leq c$ for every ML-random point x. Replacing f by $-f$ and c by $-c$, one obtains $\limsup f_n(x) \geq c$. When f_n converge to a layerwise computable function f, the first result applied to $f_n - f$ gives the second one.

Proof of Prop. 5.3.1

As the image of an effective compact set by a computable function is an effective compact set, $T(K_n)$ are uniformly effectively compact. As ν is the push-forward of μ, $\nu(T(K_n)) = \mu(T^{-1}(T(K_n))) \geq \mu(K_n) > 1 - 2^{-n}$. Using the minimality of the layering K'_n of (Y,ν), there is a constant c such that $K'_n \subseteq T(K_{n+c})$ for all n. Hence $\mathrm{ML}_\nu \subseteq T(\mathrm{ML}_\mu)$.