

Algorithmic randomness and layerwise computability

Mathieu Hoyrup

► **To cite this version:**

Mathieu Hoyrup. Algorithmic randomness and layerwise computability. Johanna N. Y. Franklin; Christopher P. Porter. Algorithmic Randomness – Progress and Prospects, Cambridge University Press, pp.17, 2020, 10.1017/9781108781718 . hal-02975222

HAL Id: hal-02975222

<https://hal.inria.fr/hal-02975222>

Submitted on 22 Oct 2020

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.

Algorithmic randomness and layerwise computability

Mathieu Hoyrup

Inria Nancy
615, rue du jardin botanique
54600 Villers-lès-Nancy, France
mathieu.hoyrup@inria.fr

Abstract

In this article we present the framework of layerwise computability. We explain the origin of this notion, its main features and properties, and we illustrate it with several concrete examples: decomposition of measures, random closed sets, Brownian motion.

Contents

1	Introduction	1
1.1	Layerwise computability vs. computability	3
1.2	Layerwise computability vs. measure and probability	3
1.3	Advantages of the notion	4
2	Layerwise computability	4
2.1	Martin-Löf randomness	4
2.2	Randomness deficiency	5
2.3	Layerwise computability	6
2.4	No randomness <i>ex nihilo</i> and compactness	7
3	Recovering a distribution from a sample	8
3.1	Class of measures	8
3.2	Decomposition of measures	9
4	Representations of random objects	11
4.1	Different representations may induce the same notion of randomness	11
4.2	Random closed sets	12
4.3	Brownian motion	13

1 Introduction

Algorithmic randomness lies at the intersection of computability theory and probability theory. These two fields follow very different approaches. In computability theory, one is interested in the properties of *individual* sets or sequences, their computation power or their information content. In probability theory, individual objects are in a sense irrelevant: what happens at one point does not matter, only what happens at most points *taken as a whole* is important.

Algorithmic randomness was born from the friction between these two fields, providing a way to apprehend the whole via individual objects: a *single* algorithmically random sequence has the typical properties of *almost all* sequences.

The fundamental difference between the two fields comes with incompatibilities between their respective concepts, which often hinders the development of algorithmic randomness. The motivation for the introduction and development of layerwise computability was to reconcile the two approaches by providing a notion that fits with both of them.

Let us illustrate one of the main incompatibilities between computability theory and probability theory. In order to carry out the study of algorithmically random sequences, one is naturally lead to manipulate concepts from computability theory such as a Turing functional, which is a computable function from sequences to sequences, and concepts from measure and probability theory such as a measurable function or an L^1 function. There are two important differences between the two classes of functions:

- Computable functions are continuous while measurable functions are usually discontinuous,
- Computable functions are well-defined at each point while L^1 functions are ill-defined at single points as they are actually equivalence classes of functions, and the value of a measurable function at a single point is irrelevant as long as that point has measure 0.

In the study of algorithmically random sequences one often has to consider the particular values of functions at single random sequences, but which class of functions is relevant? The class of computable functions is too small (only continuous functions), the class of measurable functions is not adapted (they are not well-defined at single points). A new class of functions is needed. Layerwise computable functions are a compromise between the two classes: they can be discontinuous, and they have definite values at each random point. Observe that the usual effective versions of measurable function or Borel measurable function are not appropriate as, in the first case they are not well-defined at single points and in the second case they are not well-behaved w.r.t. random sequences (for instance they do not preserve randomness when they preserve measure).

Let us now come to the definition. A partial function $f : 2^\omega \rightarrow 2^\omega$ is layerwise computable if it is defined at each Martin-Löf random sequence and if the value $f(x)$ can be uniformly computed from x and any upper bound on the randomness deficiency of x . So on the set ML_n of random points whose randomness deficiency is bounded by n , f is a usual computable function. The main idea is that while the randomness deficiency of x can only be approximated from below given x , knowing an upper bound on it usually gives much information about x , making possible many algorithmic tasks that cannot be performed without this additional information.

There are good reasons to say that this is the appropriate class of functions that fill the gap between computability theory and measure theory. First layerwise computable functions are well-defined at each random sequence. Secondly, in the same way as the functions that are computable relative to some oracle are the continuous functions, the functions that are layerwise computable relative to some oracle are exactly the Borel measurable functions (this is Lusin's theorem). Therefore they are the right effective version of measurable functions that behave well w.r.t. algorithmic randomness. We will see through many examples that layerwise computability helps in the study of algorithmically random sequences.

In this article we present the notion of layerwise computable function, giving the basic properties that make this notion useful and illustrating it with several concrete examples.

Let us first show in more detail how layerwise computability interacts nicely with computability theory on the one hand and probability and measure theory on the other hand.

1.1 Layerwise computability vs. computability

In many respects, layerwise computable functions behave like computable functions, and in many results computability assumptions can be relaxed to layerwise computability.

- Randomness preservation: If f is a measure-preserving layerwise computable function, then the image by f of the set of random points is exactly the set of random points.
- Pushforward measure: If f is layerwise computable, then the image of the measure is computable.
- Composition: If f and g are layerwise computable *and g is measure-preserving*, then $f \circ g$ is layerwise computable (here we need an additional assumption on g , unneeded when f is a plain computable function).
- Reducibility: If f is layerwise computable, then $f(x)$ is truth-table reducible to x for each random x .
- Integration: the integral of a layerwise computable non-negative real-valued function is left-c.e., and it is computable when the function is bounded.
- Uniqueness: two layerwise computable functions that coincide almost everywhere coincide on every random point.
- Inverse: the inverse of a one-to-one layerwise computable function is layerwise computable (the analog result for computable functions only holds on effectively compact spaces, while this result holds on any complete computable metric space).

1.2 Layerwise computability vs. measure and probability

Layerwise computability interacts nicely with notions from measure and probability theory. Layerwise computable functions can be seen as an effective version of measurable functions, while the computable functions are an effective version of the continuous ones, a much smaller class.

- Effective measurability: There exist several equivalent definitions of effectively measurable set or function, based on approximations in measure by simple sets or functions. It happens then that a set or function is effectively measurable iff it coincides almost everywhere with a layerwise computable function (uniquely defined on the random points).
- Measurability: A function is continuous iff it is computable relative to some oracle. A function is measurable iff it is layerwise computable relative to some oracle (both the notion of randomness and the machine computing the function must be relativized). This is essentially Lusin's theorem.
- Almost everywhere convergence: If a sequence of layerwise computable functions converges effectively almost everywhere, then it converges uniformly on each ML_n . The limit is layerwise computable and the speed of convergence is layerwise computable.
- L^1 -computability: A real-valued function is a computable element of the L^1 -space iff it coincides almost everywhere with a layerwise computable function and its integral is computable.
- Probabilistic computation: A function f is layerwise computable iff it is computable by a probabilistic machine (given x and ϵ , the machine computes $f(x)$ for x in a set of measure $> 1 - \epsilon$).

1.3 Advantages of the notion

In studying the interaction of algorithmic randomness with probability theory, one is often led to mix techniques and objects from computability theory and measure theory. However, these two fields offer very different and often incompatible perspectives. One is therefore often forced to thoroughly rework arguments and results provided by probability and measure theory to make them fit in the computability theory framework. It is desirable to have the possibility of directly applying established results without reworking them. This situation is evidence that notions are lacking.

The framework of layerwise computability offers an economic way of proving new results with minimal effort. Often the results proved in particular situations with *ad hoc* arguments elaborated on proofs of classical theorems are just instantiations of more general, structural results.

Layerwise computability is not an artificial, technical notion as it naturally enters the scene in many situations arising from the friction between computability and measure:

- The inverse of a one-to-one computable function defined on a measure one set is often not computable but always layerwise computable.
- A lower semicomputable function with computable expectation is always layerwise computable.
- Measurable functions appearing in proofs of classical results are usually discontinuous hence not computable, but often layerwise computable.

In this article we try to illustrate these assertions by concrete examples. Our goal is to convince the reader that this notion is natural, useful, and that having this tool available can help saving a lot of time and energy in the development of new results. We will show how it appears naturally in concrete situations such as the study of random closed sets, Brownian motion and decomposition of measures.

A goal of the paper is also to give a flavor of the kind of arguments used to prove the results. They are usually rather short and based on topological considerations rather than calculations and *ad hoc* technical arguments. In particular, compactness arguments prove to be very powerful. We provide proof sketches and give precise references for the detailed proofs.

The article is organized as follows. In Section 2 we present the definitions and fundamental properties of layerwise computability. In Section 3 we present a situation where it appears naturally as the answer to the problem of recovering a probability distribution from the observation of a sample, which is not a computable task but in many situations a *layerwise computable* task. In Section 4 we present investigations about the different ways encountered in the literature of defining classes of random objects other than binary sequences, with a general result stating their equivalence together with illustrations of this result in the cases of random closed sets and of Brownian motion. In each case we will see that several representations of random objects that are computably very different are actually *layerwise computably* equivalent.

2 Layerwise computability

2.1 Martin-Löf randomness

Martin-Löf's notion of randomness, originally defined on the Cantor space 2^ω of infinite binary sequences [ML66], can be extended to other topological spaces in a straightforward way. In a topological space X with a countable basis numbered in some way B_0, B_1, \dots , an effective open

set is a union $\bigcup_{i \in W_e} B_i$ where W_e is a c.e. subset of \mathbb{N} . Given a computable Borel probability measure μ over X , a Martin-Löf μ -test is, exactly as on the Cantor space, a sequence of uniformly effectively open sets U_n such that $\mu(U_n) < 2^{-n}$. A sequence x is Martin-Löf μ -random if $x \notin \bigcap_n U_n$ for each test $(U_n)_{n \in \mathbb{N}}$. We denote by ML^μ the set of Martin-Löf μ -random sequences.

When the topological space is metrizable with a suitable computable structure based on the choice of a countable dense set called *simple* points, and once the right notion of computable measure has been defined, the existence of an optimal test can be shown: a test $(U_n^\mu)_{n \in \mathbb{N}}$ such that for every test $(V_n)_{n \in \mathbb{N}}$, there exists a constant c such that $V_{n+c} \subseteq U_n^\mu$ for all n . All the details about effective topological spaces and computable metric spaces can be found in [Wei00] and about Martin-Löf randomness on general spaces in [HW98, Gác05, HR09c].

2.2 Randomness deficiency

In 2^ω endowed with the uniform measure, some sequences are Martin-Löf random, some are not. But even among the random sequences, some of them can be qualified as *less random* than others. For instance, a random sequence starting with 1,000,000 zeroes is less random than the same sequence from which the 1,000,000 first bits have been removed.

Among the random sequences, the degree of randomness can be measured. The randomness deficiency of a sequence is a positive number which measures the regularity of the sequence: the larger the deficiency is, the more regular and therefore the less random the sequence is.

There are several ways of quantifying the randomness deficiency. They are not equivalent, but all of them are close to each other. Each notion of test for which an optimal test exists induces a notion of randomness deficiency. For our purposes, the choice of the particular measure of randomness deficiency is not important and nothing in the theory depends on this choice.

Via Martin-Löf tests. Given an optimal Martin-Löf μ -test U_n , a sequence x is random iff there exists n such that $x \notin U_n$. The minimal such n is a measure of randomness deficiency.

Via Kolmogorov complexity. A sequence $x \in 2^\omega$ is Martin-Löf random if and only if its prefixes are compressible only by a bounded amount of bits. This is Levin-Schnorr theorem [Lev73, Sch73]. The maximal number of bits gained by compressing the prefixes of x is a measure of the randomness deficiency.

More precisely, x is random iff there exists c such that for all n , $K(x \upharpoonright_n) \geq n - c$. More generally if μ is a computable probability measure over 2^ω , then x is Martin-Löf μ -random if there exists c such that for all n , $K(x \upharpoonright_n) \geq -\log \mu[x \upharpoonright_n] - c$. The minimal such c is a measure of randomness deficiency. Note that here K is either prefix complexity or monotone complexity. Each one induces a measure of randomness deficiency. They are not equal, but related and both acceptable in the sense below.

Via integral tests. Another type of randomness test is given by a left-c.e. nonnegative function $t : 2^\omega \rightarrow [0, +\infty]$ such that $\int t d\mu \leq 1$. There exists an optimal test, i.e. a test t_μ such that for every test t' there exists a constant c such that $t' \leq ct_\mu$. An element x is Martin-Löf μ -random iff $t_\mu(x)$ is finite. The value $t_\mu(x)$, or its logarithm, provides a measure of randomness deficiency.

Equivalence. All the usual measures of randomness deficiency are interchangeable for our purpose because they are related in the following way. Given two such notions of deficiency $d(x)$ and $d'(x)$, there exists a computable increasing function $\phi : [0, +\infty) \rightarrow [0, +\infty)$ such that $d' \leq \phi \circ d$ (and symmetrically). In particular, from any upper bound on $d(x)$ an upper bound

on $d'(x)$ can be computed, and vice-versa. We will say that a measure of randomness deficiency is *acceptable* if it is related in that way to the canonical ones presented above. We refer to [LV93] for other measures of randomness deficiency and their relationship. For investigations on the difference between universality and optimality, and its consequences in terms of layerwise computability, one can consult [Miy11, HS14].

2.3 Layerwise computability

We fix a computable probability measure μ over 2^ω and an acceptable measure of randomness deficiency, which is a left-c.e. function $d_\mu : 2^\omega \rightarrow [0, +\infty]$ such that $x \in 2^\omega$ is Martin-Löf μ -random if and only if $d_\mu(x) < +\infty$. Given a Martin-Löf μ -random sequence x , one can only compute approximations of $d_\mu(x)$ *from below*, but no upper bound on it can be computed. The fundamental idea behind the notion of layerwise computability is that having such an upper bound provides much additional information about x that can be used to perform algorithmic tasks that would be impossible without this information (this idea had been exploited earlier by Davie [Dav01, Dav04]).

Definition 2.1. A function $f : 2^\omega \rightarrow 2^\omega$ is μ -*layerwise computable* if there is a Turing functional that given $x \in \text{ML}^\mu$ and an upper bound on $d_\mu(x)$ as inputs, computes $f(x)$.

The name of the notion comes from the following alternative definition. Let

$$\text{ML}_n^\mu = \{x \in 2^\omega : d_\mu(x) \leq n\},$$

be the set of Martin-Löf μ -random sequences whose deficiency is bounded by n , so that the set of Martin-Löf μ -random sequences can be decomposed into “layers” $\text{ML}^\mu = \bigcup_n \text{ML}_n^\mu$. A function f is then μ -layerwise computable if it is computable on each layer ML_n^μ , uniformly in n .

As d_μ is left-c.e., ML_n^μ is an effective closed set (or Π_1^0 -set), i.e. the complement of an effective open set, uniformly in n .

Computability notions often admit a topological characterization. For instance, a function is computable if and only if the pre-image of each basic open set is an effective open set, uniformly. A similar characterization exists for layerwise computability.

Proposition 2.1. *A function $f : 2^\omega \rightarrow 2^\omega$ is μ -layerwise computable if and only if the pre-image of every basic open set is an effective open set on each ML_n^μ , uniformly.*

This means that there exist uniformly effectively open sets $U_{i,n}$ such that $f^{-1}(B_i) \cap \text{ML}_n^\mu = U_{i,n} \cap \text{ML}_n^\mu$. We say that $f^{-1}(B_i)$ is μ -*layerwise effective open*. This notion has a more intuitive characterization: A is μ -layerwise effective open if and only if there exists uniformly effectively closed sets A_i with uniformly computable measures such that $\bigcup_i A_i$ coincides with A on ML^μ .

We give some useful examples of layerwise computable functions:

- If A is an effective open set or an effective closed set and $\mu(A)$ is computable, then the characteristic function $\mathbf{1}_A : 2^\omega \rightarrow \{0, 1\}$ is μ -layerwise computable. We say that A is μ -layerwise decidable.
- More generally if $f : 2^\omega \rightarrow [0, +\infty]$ is left-c.e. and $\int f d\mu$ is computable, then f is μ -layerwise computable.
- Given a random sequence x over $\{-1, 1\}$ w.r.t. the uniform measure, the *random harmonic series* $\sum_n \frac{x_n}{n}$ converges to a finite number with probability 1. It is actually finite for each Martin-Löf random x , and the function mapping x to the sum is layerwise computable, though not computable even on set of measure one (because not continuous on a set of measure one). This example is due to Bienvenu (personal communication).

We have given the definition on the Cantor space but it extends directly to any computable metric space X . Instead of manipulating $x \in X$ directly as on the Cantor space, a Turing functional works on a name of x , i.e. a binary sequence that encodes a sequence of simple points converging fast to x .

The notions were introduced in [HR09a, HR09b] where more details can be found.

Extensions to other notions of randomness. The notion of a layerwise computable function is based on Martin-Löf’s notion of randomness which has the nice property of admitting an optimal randomness test which in turn provides a notion of randomness deficiency. However layerwise computability can also be defined for other notions of randomness for which there is no randomness deficiency available, such as Schnorr randomness. In that case, some (non-optimal) randomness test is fixed and provides a measure of randomness deficiency. Schnorr layerwise computability is often useful as it is stronger than Martin-Löf layerwise computability and closer to concepts from computable analysis [Miy13].

2.4 No randomness *ex nihilo* and compactness

A famous result first discovered by Shen (unpublished) states that measure-preserving computable functions not only preserve randomness, but also do not *create* randomness: every random element has a random pre-image. This result is sometimes called the *no randomness ex nihilo* principle. The precise result is as follows.

Theorem 2.1 (No randomness *ex nihilo* on Cantor space). *Let $f : 2^\omega \rightarrow 2^\omega$ be a computable function pushing a computable measure μ to a computable measure ν . A sequence y is Martin-Löf ν -random (if and) only if $y = f(x)$ for some Martin-Löf μ -random x .*

Proof. This result is simple to prove. Start from an optimal Martin-Löf μ -test U_n , let $P_n = 2^\omega \setminus U_n$. P_n are uniformly Π_1^0 sets and so are their images $f(P_n)$. As f pushes μ to ν , $\nu(f(P_n)) \geq \mu(P_n) \geq 1 - 2^{-n}$, so the sets $2^\omega \setminus f(P_n)$ form a Martin-Löf ν -test. As a result, if x is Martin-Löf ν -random, then $x \in f(P_n)$ for some n , so x is the image of a Martin-Löf μ -random sequence. \square

This result is simple but extremely powerful. An analog of this result holds on other spaces, unfortunately for a long time it had been explicitly stated on the Cantor space only. Moreover the general result was not known for a long time on non-compact compact spaces, where the argument cannot be directly applied (on non-compact spaces, the image of a Π_1^0 set is not Π_1^0 in general). As a consequence, many results on random objects other than binary sequences have overcomplicated proofs which can be simplified by using a generalization, which we present now, of Theorem 2.1 to other spaces.

The notion of Π_1^0 set exists on any effective topological space: it is the complement of an effective union of basic open sets. However in order to extend the result from the Cantor space to other spaces, one has to observe that the relevant notion is not that of a Π_1^0 set but that of an *effectively compact* set. A compact set A is effectively compact if one can computably enumerate the finite sets of basic open sets whose union cover A . We know by definition that the sets ML_n^μ are Π_1^0 sets. On complete metric space, they are actually more than that, as stated by the following result (Theorem 2 in [HR09b]).

Theorem 2.2 (Effective compactness). *Let (X, d) be a complete computable metric space and μ a computable probability measure. The sets ML_n^μ are effectively compact, uniformly in n .*

Proof sketch. It is well-known that if A is an effective open set, then $\mu(A) = 1$ implies $\text{ML}^\mu \subseteq A$. A careful analysis of the proof shows that a “layerwise” version also holds: given an index of

an effective open set A , one can compute $c \in \mathbb{N}$ such that for all n , if $\mu(A) > 1 - 2^{-n-c}$, then $\text{ML}_n^\mu \subseteq A$.

Now given an index of an effective open set U , we want to show that the condition $\text{ML}_n^\mu \subseteq U$ is c.e. Applying the previous observation to $A = U \cup U_n^\mu$ gives a constant c . One has

$$\begin{aligned} \mu(A) > 1 - 2^{-n-c} &\implies \text{ML}_n^\mu \subseteq A \\ &\implies A = X \\ &\implies \mu(A) = 1 > 1 - 2^{-n-c}, \end{aligned}$$

so $\text{ML}_n^\mu \subseteq U \iff \text{ML}_n^\mu \subseteq A \iff \mu(A) > 1 - 2^{-n-c}$ which is a c.e. condition as the measure of an effective open set is left-c.e. \square

While the image of a Π_1^0 set by a computable function is not Π_1^0 in general, the image of an effectively compact set is always effectively compact. Moreover in a computable metric space, the complement of an effective compact set is an effective open set. As a result, the argument in Theorem 2.1 directly extends to any complete computable metric space. Moreover the function can be assumed to be layerwise computable only. We then have the following result (Proposition 5 in [HR09b]).

Theorem 2.3. *Let (X, d) and (Y, d') be computable metric spaces. Assume that (X, d) is complete. Let μ be a computable probability measure over X and $f : X \rightarrow Y$ a μ -layerwise computable function.*

- *The push-forward measure $\nu = \mu \circ f^{-1}$ is computable.*
- *An element $y \in Y$ is Martin-Löf ν -random if and only if $y = f(x)$ for some Martin-Löf μ -random x .*
- *If f is moreover one-to-one, then its inverse is ν -layerwise computable.*

The third point shows that layerwise computability is stable under taking the inverse, which is not true of plain computability when the space is not compact.

This result is often useful to prove that two ways of defining a notion of randomness are equivalent, by finding a computable measure-preserving function between two classes of objects. Section 4 is devoted to this application.

3 Recovering a distribution from a sample

We now present results showing how layerwise computability appears naturally.

3.1 Class of measures

We consider the following situation. One is given a coin and has to compute its bias by repeatedly tossing it. The underlying probability measure is a Bernoulli measure. Basic probability theory gives estimates on the convergence of the frequencies of heads and tails, from which one can derive a probabilistic algorithm to compute the bias up to any precision with arbitrarily large probability. More precisely there is an algorithm that takes a random sequence x and some $\epsilon > 0$ as inputs and computes the frequency of 1's in x for all x in a set of measure at least $1 - \epsilon$. From this one can derive that the function mapping a random sequence to the corresponding Bernoulli measure is layerwise computable: for $\epsilon = 2^{-n}$ then the algorithm will compute a correct output for every x whose deficiency is at most n .

The proof is not difficult but rather technical, and very specific to the case of Bernoulli measures. The calculations of the estimates are useful to obtain an efficient algorithm, that runs in polynomial time. However if one is only interested in computability rather than complexity there is an alternative simple argument to prove the layerwise computability of the mapping, which is applicable to other classes of measures, more general than the class of Bernoulli measures. Bienvenu and Monin [BM12] even identified the classes of measures for which such an algorithm exists.

Observe that if such an algorithm exists for a class \mathcal{C} of measures, then two different measures $\mu, \nu \in \mathcal{C}$ must be *effectively* orthogonal in the sense that they do not share random sequences: indeed, for each $\mu \in \mathcal{C}$ the algorithm computes μ given any μ -random sequence, so a sequence cannot be at the same time μ -random and ν -random for distinct $\mu, \nu \in \mathcal{C}$.

We then assume that \mathcal{C} contains pairwise effectively orthogonal measures. We say that x is \mathcal{C} -random if x is μ -random for some $\mu \in \mathcal{C}$. This notion of randomness comes with a notion of deficiency: the \mathcal{C} -deficiency of x is defined as $d_{\mathcal{C}}(x) = \inf_{\mu \in \mathcal{C}} d_{\mu}(x)$. Observe that if x is μ -random, then $d_{\mathcal{C}}(x) = d_{\mu}(x)$ and $d_{\nu}(x) = +\infty$ for all $\nu \in \mathcal{C} \setminus \{\mu\}$. More on randomness for classes for measures can be found in [BGH⁺11].

Theorem 3.1. *Let \mathcal{C} be a class of probability measures over 2^{ω} . The following statements are equivalent:*

- \mathcal{C} is contained in an effectively compact class of pairwise effectively orthogonal measures,
- The function that for every $\mu \in \mathcal{C}$ maps each μ -random x to μ is layerwise computable: μ can be uniformly computed from x and any upper bound on $d_{\mathcal{C}}(x) = d_{\mu}(x)$.

Proof sketch. Assume that \mathcal{C} is an effectively compact class of pairwise effectively orthogonal measures. One can show that the sets $U_n = \bigcap_{\mu \in \mathcal{C}} U_n^{\mu}$ form a randomness test w.r.t. every measure in \mathcal{C} . As the measures in \mathcal{C} are effectively orthogonal, if $x \in 2^{\omega} \setminus U_n$, then $\{\mu : x \in 2^{\omega} \setminus U_n^{\mu}\}$ is a singleton, containing the measure making x random. That set is Π_1^0 in x and n , uniformly, so its element is computable in x and n , uniformly.

Conversely, if there is a layerwise computable function ϕ mapping random sequences to their corresponding measures, then one can show that the class of measures $\mathcal{D} = \{\mu : \forall n, \mu(\{x : \phi(x, n) = \mu\}) \geq 1 - 2^{-n}\}$ is an effectively compact class of effectively orthogonal measures containing \mathcal{C} (see [BM12] for a detailed proof). \square

The class of Bernoulli measures is an example of effectively compact class of effectively orthogonal measures. For Bernoulli measures, it is intuitive that the measure can be computed in some way by reading the sequence and progressively approximating the frequency of 1's in it. However, in order to prove Theorem 3.1 one does not need to enter into such details and the short argument is very abstract.

3.2 Decomposition of measures

We now consider a close but different situation. This time we are given a sequence that is random w.r.t. a particular measure, and that measure can be decomposed as a combination of measures in a certain class. Several theorems state the existence and uniqueness of such a decomposition, a famous example being De Finetti's theorem, which we present now.

De Finetti's theorem. A measure μ over 2^{ω} can be equivalently presented as a sequence X_i of $\{0, 1\}$ -valued random variables. The measure $\mu([w])$ of a cylinder $w = w_0 \dots w_n$ is then given by the probability that $X_0 = w_0, X_1 = w_1, \dots$ and $X_n = w_n$. The sequence X_i is said to

be *exchangeable* if for every finite sequence $a_0, \dots, a_n \in \{0, 1\}$ of bits and every finite sequence $i_0, \dots, i_n \in \mathbb{N}$ of pairwise distinct positions, the probability that $X_{i_0} = a_0, X_{i_1} = a_{i_1}, \dots$ and $X_{i_n} = a_n$ does not depend on the positions i_0, \dots, i_n . For instance every Bernoulli measure corresponds to an exchangeable sequence of random variables, but the notion is more general. For instance if μ_p and μ_q are two Bernoulli measures, then the average $\frac{\mu_p + \mu_q}{2}$ is not a Bernoulli measure but does correspond to an exchangeable sequence of random variables. More generally, given a probability measure ν over $[0, 1]$, the combination μ of Bernoulli measures given by $\mu([w]) = \int \mu_p[w] d\nu(p)$ corresponds to an exchangeable sequence of random variables.

De Finetti's theorem states that every exchangeable sequence of random variables can be (uniquely) decomposed as a combination of Bernoulli measures, for some measure ν over $[0, 1]$. Intuitively, the process of generating a binary sequence distributed according to μ can be decomposed as follows: first generate a random number $p \in [0, 1]$ distributed according to ν , then generate a binary sequence using a coin with bias p .

The question is then: having observed the outcome, i.e. the infinite binary sequence, can one recover the bias p ? The answer is positive and actually holds in much more general situations.

Theorem 3.2. *Let \mathcal{C} be a class of pairwise effectively orthogonal measures over 2^ω , and ν a computable probability measure supported on \mathcal{C} . Let $\mu([w]) = \int \eta([w]) d\nu(\eta)$. The following holds:*

- *A sequence x is μ -random if and only if it is η -random for some ν -random measure η .*
- *The function mapping μ -random x to the ν -random measure η such that x is η -random is μ -layerwise computable.*

Proof sketch. Here we use the integral tests mentioned in Section 2.2 as a measure of randomness deficiency. Let $f(x) = \inf_{\eta} t_{\nu}(\eta) t_{\eta}(x)$. The decomposition of μ directly gives $\int f d\mu \leq 1$, so f is a μ -test and $f \leq ct_{\mu}$ for some constant c .

As a result, if x is μ -random, then x is η -random for some ν -random measure η . Moreover there is a positive constant c' such that if n is an upper bound on the deficiency $t_{\mu}(x)$, then $c'n$ is an upper bound on the deficiency $t_{\eta}(x)$.

So given x and $n \geq t_{\mu}(x)$, the corresponding measure η is the unique measure such that $t_{\eta}(x) \leq c'n$, so $\{\eta\}$ is Π_1^0 in x and n , uniformly, hence η is computable in x and n , uniformly. \square

If one takes the class of Bernoulli measures for \mathcal{C} , then one gets the case of De Finetti's theorem. Whatever ν is, the frequency of 1's in the sequence is μ -layerwise computable. In that case, if one starts with a computable measure μ , a simple compactness argument shows that the measure ν over the class of Bernoulli measures (or equivalently over $[0, 1]$) is computable. This was generalized by Freer and Roy [FR09] to other spaces than the Cantor space.

Ergodic decomposition. Another interesting case of decomposition of measures is when μ is a shift-invariant probability measure, which means that the probability that a sequence starts with w is the same as the probability that the shifted sequence (obtained by removing the first bit of the sequence) starts with w . Formally it means that $\mu([w]) = \mu([0w]) + \mu([1w])$ for all w .

The ergodic decomposition theorem states that every shift-invariant measure can be uniquely decomposed as a combination of other shift-invariant measures, that in turn cannot be decomposed further. Those shift-invariant measures that have no non-trivial decomposition into other shift-invariant measures are called the *ergodic* measures. Two different ergodic measures are always effectively orthogonal, so Theorem 3.2 can be applied to the class \mathcal{C} of ergodic shift-invariant measures. This time, the computability of μ does not imply the computability of ν , which must be assumed [Hoy13].

4 Representations of random objects

4.1 Different representations may induce the same notion of randomness

We have seen that the notion of a Martin-Löf random element can be defined on many spaces endowed with a probability measure, by simply reproducing the original notion developed on the Cantor space.

Another common way to define an algorithmic notion of randomness for a class of objects is to encode them into infinite binary sequences and then export the notion of Martin-Löf randomness from the Cantor space to the space under consideration. The choice of the particular representation is an important issue as each particular representation induces its own notion of randomness.

However it often happens that different representations that are far from being computably equivalent actually induce the same notion of randomness. We give here a result explaining this phenomenon in a large number of situations, and whose proof is elementary. This result also shows that in these cases, representations that are far from being computably equivalent are actually *layerwise* computably equivalent. In particular they are computably equivalent at each random element.

We assume that the strongest representation of the objects is derived from a *complete* computable metric on the space of objects. An object is then represented by successive approximations in the underlying metric.

Formally, we consider a complete computable metric space (X, d) , i.e. a separable complete metric space with a distinguished dense sequence $(s_i)_{i \in \mathbb{N}}$ such that the real numbers $d(s_i, s_j)$ are uniformly computable. The metric d induces a first representation, the Cauchy representation: a point x is represented by any sequence of natural numbers i_0, i_1, \dots such that $d(s_{i_n}, x) < 2^{-n}$. We then consider several cases for the weaker representation.

Metric representations. Here we assume that the weaker representation is derived from a weaker metric. Let then d' be another metric on X that is effectively weaker than d : as a function from $X \times X$ to \mathbb{R} it is computable. This is equivalent to saying that (X, d') is another computable metric space (with the same countable dense set) such that the identity from (X, d) to (X, d') is computable.

Representing a point $x \in X$ by approximations in the metric d provides at least as much information about x as giving approximations of x in the metric d' , and possibly more.

Now we assume that (X, d) is endowed with a computable Borel probability measure μ (which is automatically computable on the space (X, d')). As d' is weaker than d , there are less Martin-Löf tests in (X, d') than in (X, d) so there might be strictly more Martin-Löf random points in (X, d') than in (X, d) . It is not the case.

Theorem 4.1. *Let (X, d) be a complete computable metric space, μ a computable probability measure and d' a metric that is effectively weaker than d .*

- *The Martin-Löf random points of (X, d, μ) are the same as the Martin-Löf random points of (X, d', μ) .*
- *Moreover, mapping $x \in (X, d')$ to $x \in (X, d)$ is μ -layerwise computable.*

Proof. The identity f from (X, d) to (X, d') is computable, one-to-one and preserves μ . As (X, d) is complete we can apply Theorem 2.3. \square

Other representations. We now consider the more general case. We still have the *Cauchy representation* $\delta_C : 2^\omega \rightarrow X$ derived from the computable metric d . We then assume another representation, i.e. a surjective partial function $\delta : 2^\omega \rightarrow X$, and we say that x is δ -Martin-Löf random if it has a δ -name that is Martin-Löf random w.r.t. the uniform measure λ (λ is the canonical measure over 2^ω , but any other measure can be chosen instead).

Observe that we have a notion of randomness derived from δ which makes no reference to any measure over X . Actually it is a notion of randomness for the measure μ defined as the push-forward of λ under δ : $\mu(A) = \lambda(\delta^{-1}(A))$ for all Borel sets A (here we have to assume that δ is Borel).

We can now compare δ -randomness with Martin-Löf randomness in the space (X, d, μ) . For this we have to assume that μ is computable.

Observe that if δ as a function from $2^\mathbb{N}$ to X is computable, then obviously the two randomness notions coincide. So we are interested in the situation when δ is not computable. We still have to assume some effectivity. Often, δ is “semicomputable” in some sense. We do not give a general definition, but give a result that illustrates the idea when the space X is the Cantor space. We then give an example in the next section.

We say that $\delta : 2^\mathbb{N} \rightarrow 2^\mathbb{N}$ is c.e. if there is a Turing machine that on oracle $x \in 2^\mathbb{N}$ enumerates $\delta(x)$, as a subset of \mathbb{N} .

Theorem 4.2. *Let μ a computable probability measure over $2^\mathbb{N}$ and $\delta : 2^\mathbb{N} \rightarrow 2^\mathbb{N}$ a c.e. map that pushes the uniform measure λ to μ .*

- *The Martin-Löf random points of $(2^\mathbb{N}, \mu)$ are exactly the δ -random points (i.e. the images of Martin-Löf random points by (δ, λ) by δ).*
- *Moreover, δ is λ -layerwise computable.*

Proof. Given n , the set $U_n = \delta^{-1}(\{x \in 2^\mathbb{N} : x_n = 1\})$ is an effective open set. Moreover, $\lambda(U_n) = \mu(\{x \in 2^\mathbb{N} : x_n = 1\})$ is computable, so U_n is λ -layerwise computable. As this is uniform in n , δ is λ -layerwise computable so Theorem 2.3 applies. \square

Observe that for an arbitrary x , $\delta(x)$ is not computable relative to x , but only c.e. relative to x . However, assuming that the push-forward measure of δ is computable implies that when x is Martin-Löf random, $\delta(x)$ is computable relative to x .

In the situations described by Theorem 4.1 and 4.2, several representations of the same objects that give more or less information about them induce the same notion of randomness, and the representations are actually computably equivalent on the random elements, in a layerwise fashion.

We now illustrate the preceding results to two classes of objects: closed subsets of 2^ω and continuous real-valued functions over $[0, 1]$.

4.2 Random closed sets

A way of obtaining a notion of Martin-Löf random closed subset of 2^ω is to fix a representation of closed sets by infinite sequences over a finite alphabet and to label a closed set as random if it has a name that is Martin-Löf random w.r.t. the uniform measure (or possibly another fixed Bernoulli measure). Again the notion of random closed set depends on the representation and on the measure on the space of names.

Mainly two representations have been proposed in the literature, in [BBC⁺07] and [Kjo09].

The hit-or-miss representation. In [BBC⁺07] a non-empty closed set $C \subseteq 2^\omega$ is represented by the tree *without dead ends* consisting of all the finite strings having an extension in C . This tree is encoded as an infinite sequence over the alphabet $\mathfrak{3} = \{0, 1, 2\}$ and 3^ω is endowed with the uniform measure. Hence a name of a closed set C tells which cylinders hit (intersect) C and which ones miss (are disjoint from) C , so the induced topology is called the *hit-or-miss* topology. This representation is equivalent to the one derived from the Hausdorff metric (with for instance the clopen sets as a dense countable subset), which is complete.

The induced measure is as follows: given that C intersects $[w]$, the probability that C intersects only one of $[w0]$ and $[w1]$ is $1/3$, the probability that C intersects both is also $1/3$.

The miss representation. In [Kjo09] a closed set C is represented by any set S of finite strings such that C is the set of infinite sequences having all their prefixes in S . Such a set S is encoded as an infinite sequence over $2 = \{0, 1\}$ and 2^ω is endowed with a Bernoulli measure with parameter $p \in [0, 1]$. The topology induced by this representation is called the *miss* topology: a representation of C tells which cylinders miss C , but does not tell which ones hit C . This representation can also be seen as the generation of a tree by a Galton-Watson process and this time the tree does have dead ends.

The induced measure on the trees is as follows: given that w belongs to the tree, $w0$ and $w1$ belong to the tree with probability p , independently. The induced measure on the closed set is then determined by the following conditions: the probability that C is non-empty is 0 if $p \leq 1/2$ and $(2p - 1)/p^2$ if $p > 1/2$, and given that C intersects $[w]$, the probability that C intersects only one of $[w0]$ and $[w1]$ is $1 - p$ and the probability that C intersects both is $2p - 1$.

Comparison. The two representations are topologically and computably very different, as the first one gives much more information about the closed set than the second one. However for $p = 2/3$ they induce the same measure (restricted to the non-empty closed sets) and the same notion of randomness. This equivalence has been proved in [DK09, Axo10], but the framework of layerwise computability provides the much simpler argument of Theorem 4.2.

Theorem 4.3. *The two notions of random closed set coincide for non-empty closed sets. Moreover, a hit-or-miss representation of a non-empty random closed set can be computed from a miss representation in a layerwise computable way.*

Proof. The miss representation is “semicomputable”: given S , one can enumerate the complement of the closed set represented by S . As the induced measure μ_1 is computable and coincides with the measure μ_2 restricted to the non-empty closed sets, the miss representation is actually layerwise computable and induces the same notion of randomness. \square

In particular, given a random miss representation, i.e. a tree with dead ends, it is layerwise decidable whether a node in the tree eventually dies.

4.3 Brownian motion

There are many different ways of constructing, defining, and representing Brownian motion. Each one of them induces a representation of paths and the different representations may or may not be computably reducible to each other.

In any case, one has to choose a way of representing objects and a measure over the set of objects.

Wiener measure and uniform distance. One approach to define Brownian motion is to declare that a path is a continuous function, i.e. an element of the set $\mathcal{C}[0, 1]$ of continuous real-valued functions over $[0, 1]$, and to endow $\mathcal{C}[0, 1]$ with a probability measure called the Wiener measure.

The Wiener measure over $\mathcal{C}[0, 1]$ is defined as the unique measure having normal conditional expectations with certain parameters: the value of $f(t)$ given $f(t_0), \dots, f(t_n)$ is normally distributed with parameters depending on t_0, \dots, t_n, t .

The usual way of representing a continuous function $f \in \mathcal{C}[0, 1]$ is to give the needed information in order to evaluate f at any real number. This is done by giving the values of f on the dyadic rationals together with a modulus of uniform continuity for f (i.e. a function giving, for each $\epsilon > 0$, some $\delta > 0$ such that $|x - y| < \delta$ implies $|f(x) - f(y)| < \epsilon$). This is also equivalent to endowing $\mathcal{C}[0, 1]$ with the uniform distance

$$d_\infty(f, g) = \max_{x \in [0, 1]} |f(x) - g(x)|$$

which makes it a complete computable metric space (taking for instance the rational piecewise affine functions as countable dense subset).

One then has to prove that W is a computable measure, this is done in [Fou00a]. The notion of a Martin-Löf W -random function then comes for free, using the definition of Martin-Löf test presented in Section 2.1.

Wiener measure and values on the rationals. Another way to represent a function $f \in \mathcal{C}[0, 1]$ is to give the values of f on the dyadic rationals, but no information about its modulus of continuity. Formally it can be done by identifying $f \in \mathcal{C}[0, 1]$ with the sequence $f(d_n)$ of values of f on the dyadic rationals, where $(d_n)_{n \in \mathbb{N}}$ is an effective enumeration of the set \mathbb{D} of dyadic rationals in $[0, 1]$. In this way, $\mathcal{C}[0, 1]$ is homeomorphic to a subset of $\mathbb{R}^{\mathbb{N}}$ and inherits the metric over $\mathbb{R}^{\mathbb{N}}$:

$$d_{\mathbb{D}}(f, g) = \sup_n 2^{-n} \frac{|f(d_n) - g(d_n)|}{1 + |f(d_n) - g(d_n)|}$$

The metric $d_{\mathbb{D}}$ is strictly weaker than d_∞ . Approximating $f \in \mathcal{C}[0, 1]$ in the metric $d_{\mathbb{D}}$ gives no information about the modulus of continuity of f , hence cannot be used in general to give approximations of f in the uniform metric. In other words, the identity from $(\mathcal{C}[0, 1], d_{\mathbb{D}})$ to $(\mathcal{C}[0, 1], d_\infty)$ is far from being computable (while its inverse is). However as a direct application of Theorem 4.1, for *random* functions the two types of approximations give the same information, in a layerwise computable way.

Corollary 4.1. *The Martin-Löf W -random elements of $(\mathcal{C}[0, 1], d_\infty)$ are exactly the Martin-Löf W -random elements of $(\mathcal{C}[0, 1], d_{\mathbb{D}})$. Moreover, the function mapping $f \in (\mathcal{C}[0, 1], d_{\mathbb{D}})$ to $f \in (\mathcal{C}[0, 1], d_\infty)$ is W -layerwise computable.*

The equivalence is proved in [Fou00b] without using layerwise computability, and the layerwise computability of the mapping is proved in [DF13]. Note that the result is much more general as it remains true when replacing W with any other (computable) measure over $\mathcal{C}[0, 1]$ and $d_{\mathbb{D}}$ with any computable metric over $\mathcal{C}[0, 1]$.

Franklin-Wiener series. Another way of defining Brownian motion is to consider it as the limit of discrete-time random walks.

A path is obtained by summing up a linear combination with random coefficients of triangular functions located at dyadic rationals. A path is then determined by a sequence of random, normally distributed, coefficients.

In this way we get another possible representation of paths by giving the sequence of coefficients, encoded in an element of 2^ω , and declaring a path random if it has a random encoding. This is the approach taken by Fouché in [Fou00b].

Classical theorems state that almost surely the series converge to a continuous function and that the induced measure over $\mathcal{C}[0, 1]$ is precisely the Wiener measure.

It is not difficult to see that giving the coefficients of the series is equivalent to giving the values of the limit function at the dyadic rationals: the value of f at some dyadic rational d can be computed from a finite number of coefficients using some formula, and reciprocally each coefficient can be computed from the values of f on a finite set of dyadic rationals using another formula.

As a result the Franklin-Wiener representation is computably equivalent to the representation derived from the metric $d_{\mathbb{D}}$. From this we conclude that again we obtain the same notion of randomness.

We mention the original definition of algorithmically random path due to Asarin and Prokovskii [AP86] using Kolmogorov complexity, which was proved to be equivalent to the definitions presented here by Fouché [Fou00a].

References

- [AP86] Eugene Asarin and Alexei Pokrovskii. Use of the Kolmogorov complexity in analyzing control system dynamics. *Avtomatika i Telemekhanika*, 1, 1986. (Russian, English translation in Automation and Remote Control, 47, Part 1, No 1).
- [Axo10] Logan M. Axon. *Algorithmically random closed sets and probability*. PhD thesis, University of Notre Dame, 2010.
- [BBC⁺07] George Barmpalias, Paul Brodhead, Douglas Cenzer, Seyyed Dashti, and Rebecca Weber. Algorithmic randomness of closed sets. *J. Log. and Comput.*, 17(6):1041–1062, 2007.
- [BGH⁺11] Laurent Bienvenu, Peter Gács, Mathieu Hoyrup, Cristóbal Rojas, and Alexander Shen. Algorithmic tests and randomness with respect to a class of measures. *Proceedings of the Steklov Institute of Mathematics*, 274(1):34–89, November 2011.
- [BM12] Laurent Bienvenu and Benoît Monin. Von neumann’s biased coin revisited. In *Proceedings of the 27th Annual IEEE Symposium on Logic in Computer Science, LICS 2012, Dubrovnik, Croatia, June 25-28, 2012*, pages 145–154. IEEE Computer Society, 2012.
- [Dav01] George Davie. The Borel-Cantelli lemmas, probability laws and Kolmogorov complexity. *Annals of Probability*, 29(4):1426–1434, 2001.
- [Dav04] George Davie. Characterising the Martin-Löf random sequences using computably enumerable sets of measure one. *Information Processing Letters*, 92(3):157–160, 2004.
- [DF13] George Davie and Willem L. Fouché. On the computability of a construction of brownian motion. *Mathematical Structures in Computer Science*, 23(6):1257–1265, 2013.
- [DK09] David Diamondstone and Bjørn Kjos-Hanssen. Members of random closed sets. In Klaus Ambos-Spies, Benedikt Löwe, and Wolfgang Merkle, editors, *Mathematical*

Theory and Computational Practice, 5th Conference on Computability in Europe, CiE 2009, Heidelberg, Germany, July 19-24, 2009. Proceedings, volume 5635 of *Lecture Notes in Computer Science*, pages 144–153. Springer, 2009.

- [Fou00a] Willem L. Fouché. Arithmetical representations of Brownian motion I. *J. Symb. Log.*, 65(1):421–442, 2000.
- [Fou00b] Willem L. Fouché. The descriptive complexity of Brownian motion. *Advances in Mathematics*, 155:317–343, 2000.
- [FR09] Cameron E. Freer and Daniel M. Roy. Computable exchangeable sequences have computable de Finetti measures. In Klaus Ambos-Spies, Benedikt Löwe, and Wolfgang Merkle, editors, *CiE*, volume 5635 of *Lecture Notes in Computer Science*, pages 218–231. Springer, 2009.
- [Gác05] Peter Gács. Uniform test of algorithmic randomness over a general space. *Theoretical Computer Science*, 341:91–137, 2005.
- [Hoy13] Mathieu Hoyrup. Computability of the ergodic decomposition. *Annals of Pure and Applied Logic*, 164(5):542–549, 2013.
- [HR09a] Mathieu Hoyrup and Cristobal Rojas. An application of Martin-Löf randomness to effective probability theory. In Klaus Ambos-Spies, Benedikt Löwe, and Wolfgang Merkle, editors, *CiE*, volume 5635 of *Lecture Notes in Computer Science*, pages 260–269. Springer, 2009.
- [HR09b] Mathieu Hoyrup and Cristobal Rojas. Applications of effective probability theory to Martin-Löf randomness. In Susanne Albers, Alberto Marchetti-Spaccamela, Yossi Matias, Sotiris E. Nikolettseas, and Wolfgang Thomas, editors, *ICALP 2009*, volume 5555 of *Lecture Notes in Computer Science*, pages 549–561. Springer, 2009.
- [HR09c] Mathieu Hoyrup and Cristóbal Rojas. Computability of probability measures and Martin-Löf randomness over metric spaces. *Information and Computation*, 207(7):830–847, 2009.
- [HS14] Rupert Hölzl and Paul Shafer. Universality, optimality, and randomness deficiency. Submitted. Available at <http://arxiv.org/abs/1409.8589>, 2014.
- [HW98] Peter Hertling and Klaus Weihrauch. Randomness spaces. In Kim G. Larsen, Sven Skyum, and Glynn Winskel, editors, *Automata, Languages and Programming*, volume 1443 of *Lecture Notes in Computer Science*, pages 796–807, Berlin, 1998. Springer. 25th International Colloquium, ICALP’98, Aalborg, Denmark, July 1998.
- [Kjo09] Bjørn Kjos-Hanssen. Infinite subsets of random sets of integers. *Mathematical Research Letters*, 16:103–110, 2009.
- [Lev73] Leonid A Levin. On the notion of a random sequence. *Soviet Math. Dokl.*, 14(5):1413–1416, 1973.
- [LV93] Ming Li and Paul M. B. Vitanyi. *An Introduction to Kolmogorov Complexity and Its Applications*. Springer-Verlag, Berlin, 1993.
- [Miy11] Kenshi Miyabe. The difference between optimality and universality. *Logic Journal of IGPL*, 2011.

- [Miy13] Kenshi Miyabe. L^1 -computability, layerwise computability and solovay reducibility. *Computability*, 2(1):15–29, 2013.
- [ML66] Per Martin-Löf. The definition of random sequences. *Information and Control*, 9(6):602–619, 1966.
- [Sch73] Claus-Peter Schnorr. Process complexity and effective random tests. *J. Comput. Syst. Sci.*, 7(4):376–388, 1973.
- [Wei00] Klaus Weihrauch. *Computable Analysis*. Springer, Berlin, 2000.