CDMTCS
Research
Report
Series

Randomness and Ergodic Theory: An Algorithmic Point of View

Cristóbal Rojas González

École Polytechnique, Paris, France
Università di Pisa, Italy

CDMTCS-333
September 2008

Centre for Discrete Mathematics and
Theoretical Computer Science

ÉCOLE POLYTECHNIQUE

UNIVERSITÀ DI PISA

Doctorat en mathématiques–informatique

–

Aléatoire et théorie ergodique : un point de vue algorithmique

Cristóbal ROJAS GONZÁLEZ

–

Thèse dirigée par

Giuseppe LONGO / Stefano GALATOLO

Soutenue le 18 juin 2008

Rapporteurs

Eugene ASARIN
Cristian CALUDE
Serge TROUBETZKOY

JURY

Eugene ASARIN	rapporteur
Pierre-Louis CURIEN	président
Gilles DOWEK	examinateur
Stefano GALATOLO	directeur
Giuseppe LONGO	directeur
Karl PETERSEN	examinateur
Serge TROUBETZKOY	rapporteur

Remerciements (Agradecimientos)

Je commence par un grand merci à Giuseppe Longo, pour m'avoir éclairé de sa vision des sciences (au travers de ses réflexions sur l'aléatoire, la physique et les sciences de la nature en général) sans pour autant l'imposer, et pour m'avoir aidé et soutenu à trouver mes propres moyens mathématiques pour traiter techniquement ces questionnements. Je remercie aussi Stefano Galatolo pour avoir co-dirigé cette thèse, pour m'avoir guidé dans ma decouverte de la théorie ergodique et pour les très agréables moments partagés. Je remercie les membres du jury: Serge Troubetzkoy, Eugène Asarin et Cristian Calude pour avoir accepté d'être rapporteurs et Karl Petersen, Pierre-Louis Curien et Gilles Dowek pour avoir eu la gentillesse de faire partie de ce jury.

Ce travail de thèse a été développé en étroite collaboration avec Mathieu Hoyrup. Pour avoir voulu se lancer avec moi dans ce projet, pour tout ce que nous avons appris, cherché et résolu ensemble et que j'ai vecus avec autant de plaisir : merci l'ami.

Je tiens à remercier Peter Gács pour s'être interessé à notre travail et avoir accepté de nous rendre visite pour travailler avec nous.

Un profond merci est déstiné à ma femme Tamara qui a tout abandoné pour devenir ma compagne dans cette aventure. Pour m'avoir supporté et soutenu dans des moments difficiles ainsi que pour tout le bonheur et la folie avec lesquels elle a su remplir ma vie, je lui suis très reconnaissant.

Toda mi gratitud va para mi familia, para quienes la distancia no ha sido obstáculo a la hora de hacerme sentir el calor de hogar cada vez que lo necesité. Especialmente agradecido estoy de mis padres, Ruth y José Luis, por haberme ofrecido las mejores oportunidades, por haberme orientado con sabiduría y luego por haberme apoyado incondicionalmente en mis decisiones.

Merci à la famille Thibaud pour m'avoir ouvert toutes ses portes (ainsi qu'à mes amis), pour m'avoir aidé une infinité de fois, enfin pour m'avoir adopté comme un des leurs.

Je voudrais remercier toute la cohorte des chiliens, en particulier Fernando, Ricardo et Eduardo, pour *une quantité non énumerable de moments tellement inoubliables.*

Finalement je vous remercie vous, cher lecteur, et je vous souhaite courage pour la suite.

Thesis advisor Author

Giuseppe Longo / Stefano Galatolo **Cristóbal Rojas González**

Aléatoire et théorie ergodique: un point de vue algorithmique.

Abstract

The first of our two main goals is investigating the relation between the notions of *algorithmic randomness* and *dynamical typicalness*. The former express the "algorithmic unpredictability" of individual infinite sequences with respect to a given probability measure and is modeled with the tools of computability theory. The latter, usually modeled in the framework of ergodic theory/dynamical systems, express the "physical plausibility" of an initial condition with respect to its evolution under some dynamics (a physically plausible point should follow the "expected" or "typical" behaviour of the system). Roughly, the motivation is the following: absolutely complete knowledge of the state of a physical system is unattainable (for many reasons). Now, supposing we access the physical world by "algorithmic means" (measurements are always finite approximations possibly elaborated later in computers) then it makes sense to modelize a "physical" point as being "algorithmically unknownable". To study this, it is essential to develop algorithmic tools adapted to the usual context in which ergodic theory takes place and this is what the entire first part of this thesis is devoted to. Then we consider two notions of *algorithmic randomness* due to Martin-Löf and Schnorr and prove three main results, thus establishing the relationship with *typicality*. i) In any computable metric space with an arbitrary probability measure μ, there always exists a universal uniform Martin-Löf-randomness test; ii) If μ is computable, then the trajectory of a Martin-Löf random point under **any** *computable ergodic* dynamics always follow the typical statistical behaviour of the system; iii) A point is Schnorr random if and only if its trajectory follows the typical behaviour of **any** *computable mixing* dynamics. As a second goal we study from a highly theoretical point of view the problem of simulating randomness or typical statistical behaviours on a computer, with a particular interest in the ergodic behaviour of dynamical systems. We prove that for several classes of dynamical systems having a single "physically relevant" invariant measure, i) this measure is computable and ii) there exist computable points (the only ones we have access in computer simulations) following the typical behaviour of the system. As a direct application we obtain the existence of computable reals which are Borel-normal with respect to any base.

Articles

The following article is a short survey on the subject. It was written at the end of the first year of doctoral work and it states many of the questions that this thesis tries to answer.

> [Roj08]:**Computability and information in models of randomness and chaos**, *Mathematical Structures in Computer Science* (2008), Cambridge University Press. *vol* 18 *pp.* 291-307.

Written with Mathieu Hoyrup, the following article contains a large part of Chapters 2, 3 and 4.

> [HR07]:**Computability of probability measures and Martin-Löf randomness over metric spaces**, submitted to *Information and Computation*, preprint available at http://arxiv.org/abs/0709.0907

The following is a joint work with Mathieu Hoyrup and Stefano Galatolo. The main results of this article are similar to those of Chapter 5. Some of the techniques we present in this thesis are somewhat different.

> [GHR07b]:**An effective Borel-Cantelli lemma. Constructing orbits with required statistical properties** Submitted to the *Journal of the European Mathematical Society.* Preprint available at http://arxiv.org/abs/0711.1478

Written jointly with Mathieu Hoyrup and Stefano Galatolo, the following article contains a large part of Chapter 4.

> [GHR07a]:**Algorithmically random points in measure preserving systems, statistical behaviour, complexity and entropy**, submitted to *Information and Computation.* Preprint available at http://arxiv.org/abs/0801.0209

This joint work with Peter Gács and Mathieu Hoyrup develops parts of Chapter 2 and 4.

> [GHR08a]:**Randomness–A dynamical point of view**, *in preparation.*

The following work with Stefano Galatolo and Mathieu Hoyrup is currently in development.

> [GHR08b]:**Computable invariant measures**, *in preparation.*

Contents

Introduction

This thesis is mainly concerned with the understanding of the relation between the notions of *algorithmic randomness* and *dynamical typicalness*, as well as investigating the problem of simulating randomness or typical statistical behaviours on a computer.

Algorithmic randomness and typicalness The theory of algorithmic randomness makes use of computability tools to express what means for an object to be algorithmically "random" or "unpredictable" or "inaccessible", with respect to a probability distribution μ. This algorithmic modelization of randomness have been formalized in different ways and there exist several definitions, Martin-Löf's one being the most celebrated. In all cases, being random can be identified with a high degree of "non-computability", expressed for instance by means of Kolmogorov complexity which measures the minimal amount of information from which an object can be *computed*. Then algorithmically random objects are those being maximally complex.

Let us now explain what we mean by *dynamical typicalness*. In general, the premise for the use of probabilistic methods is the assumption that the result x of the physical process under investigation arises *randomly* with respect to some probability distribution μ which can be discovered, hypothesized, inferred, measured... in several ways. Under this premise (of the process *governed by the probability law* μ) those properties of x which are satisfied with probability one are taken as probabilistic laws and subjecting x to such properties is predicted. In the context of a physical process modeled by a dynamical system (X, T) (X represents the set of all possible states of the system and T describes the dynamics), μ corresponds to a distribution for which the probability of events does not change in time, reflecting the fact that the system is in an *equilibrium* situation. Then, if the starting state is x, the orbit $\{x, T(x), T^2(x), ...\}$ represents the system evolving in time. In this case, there exists a natural class of (asymptotic) properties allowing a definition of "random state" or "physically plausible state" with respect to the dynamic T and the equilibrium distribution μ. It is the class associated to the ergodic Birkhoff theorem which states that for each (integrable) function $f : X \to \mathbb{R}$ (which represents a quantified observation), the time average along the orbit $O(x) = \{x, T(x), T^2(x), ...\}$ converges to the spatial mean $\int f d\mu$ with probability one. Then, this typical statistical behaviour of the system is "expected" to be followed by typical initial conditions. The set of points evolving in this expected way under the action of the dynamics T are then called ***T-typical***. The T-trajectories starting from T-typical points are exactly those described by the statistics of μ. If this distribution is "physical" in some sense, then what one expects to see in "reality" or in a good computer simulation is one of these trajectories. Hence, a "physically plausible" point is expected to show this typical behaviour under the evolution of *any* dynamics. Of course, if we consider all possible dynamics, then there are no such points. We will then consider only those dynamics that can be "computed" by algorithmic means and define the set of ***typical*** points as those being T-typical with respect to **every** *computable* ergodic dynamics.

But why shall we consider algorithmically random points as modelizing physically plausible ones? It is known that absolutely complete knowledge of the state of a physical system is unattainable (for many reasons). Consequently, an idealized, perfectly known number, say for example $\frac{1}{2}$ (if phase-space is modeled by $[0, 1]$), does not really have a physical sense. As an elementary example, if we consider a system modeled by $T(x) = 2x (\text{mod } 1)$ over $[0, 1]$, we expect orbits to be dense in $[0, 1]$ (among others properties), but the orbit of $\frac{1}{2}$ (and actually of any dyadic number) is eventually constant and equal to 0 (this follows from the particular relation between dyadic numbers and the definition of T). These numbers posses too many regularities, and this may produce "exceptional" behaviours[1].

Now, supposing that our access to the physical world is "algorithmical" in the sense that measurements are always finite approximations possibly elaborated later in computers, then it make sense to modelize a "physical" point as being "algorithmically unknowable", that is, without any algorithmical regularities.

The role of algorithmic randomness in dynamical systems, especially in ergodic ones, has already been the subject of previous research ([V'y97, V'y98, KST94, CHJW01]) but,

[1] *Individually*, each of these too regular points (a periodic point in a chaotic system for instance) does not really represent a physical state. On the other hand, all together, they can describe typical properties very well: they are a powerful tool to study the system.

without a more general theory of randomness, all these works are restricted to symbolic spaces.

The first part of this thesis is devoted to the transfer of some tools of computability theory to more general spaces thus providing a robust framework more adapted to the variety of physico-mathematical models and allowing a more systematic study of certain questions, among which are those related to randomness and ergodic theory. Following this direction, in Chapter 3 we study Martin-Löf algorithmic randomness in general metric spaces and prove the existence of a universal uniform test, improving Gács's result (see [Gác05]). We have also considered a slightly weaker notion of algorithmic randomness, due to C. Schnorr ([Sch71]). We generalize this notion to computable probability spaces and give some properties. In Chapter 4, we show that every Martin-Löf-random point is typical for every computable dynamics and that the asymptotic Kolmogorov complexity of every random orbit equals the metric entropy of the system (these are generalizations of the results in [V'y98, Bru83]). Concerning Schnorr's version of algorithmic randomness, we have established the following characterization: a point x is Schnorr-random *if and only if* it is typical for every computable "polynomially mixing" dynamics.

Pseudo-randomness. Computer simulations (the trajectory of some initial condition drawn on the screen) has become a very important technique used to *infer* the equilibrium distribution μ. As we have seen, the trajectory of any random initial condition would allow to recover this distribution. The problem is that algorithmically random points are strongly non-computable, and consequently it is impossible to observe the trajectory of such a point in a computer simulation which is only able to show orbits starting from *computable* initial conditions. Worst, the set of computable points has probability 0! From the simulation point of view, the fact that a given property holds with probability one says nothing about its *observability* with a computer–a typical example is normality in the sens of Borel: a real number is absolutely normal[2] with probability one, but constructing such points is extremely complicate [BF02]. However, it is widely accepted that computable simulations show the right ergodic behaviour. The evidence is mostly heuristic. Most arguments are based on the various "shadowing" results (see e.g. [HK95] chapter 18). In this kind of

[2] That is, normal with respect to every base.

approach (different from our), it is possible to prove that in a suitable system, any "pseudo"-trajectory, as the ones which are obtained in simulations with some computation error is near to a real trajectory of the system. So we know that what we see in a simulation is near to some real trajectory, but we do not know if the trajectory is typical in some sense. The main limitation/ of this approach is however that shadowing results hold only in particular systems, having some uniform hyperbolicity, while many physically interesting systems are not like this.

In our approach we consider real trajectories instead of "pseudo" ones and we ask:

are there computable points which are typical for a given dynamics?

In Chapter 5, we use the algorithmic tools developed so far to show a general method allowing (under certain conditions) the construction of computable points satisfying a given probabilistic law. In particular, we show that if T is a "polynomially mixing" dynamics then there exist computable points which are typical for T. A direct application shows the existence of computable real numbers which are *absolutely normal*.

All these statements require that the dynamics and the invariant measure are computable. The first assumption can be easily checked on concrete systems if the dynamics is given by a map which is effectively defined. The second is more delicate: it is well known that given a map on a metric space, there can be a continuous (even infinite dimensional) space of probability measures which are invariant for the map, and many of them will be non computable. An important part of the theory of dynamical systems is devoted to the selection of measures which are particularly meaningful. From this point of view, an important class is given by SRB invariant measures, which are measures being in some sense the "physically meaningful ones" (for a survey on this topic see [You02]). In the final part of Chapter 5, we show that in several classes of dynamical systems where SRB measures are proved to exist, these measures are also computable from our formal point of view, hence providing several classes of nontrivial concrete examples where our results can be applied.

Part I

Computability and Mathematics

Chapter 1

Computability on continuous spaces

1.1 Basic Notation

As usual, $\mathbb{N}$, $\mathbb{Z}$, $\mathbb{Q}$ and $\mathbb{R}$ will denote the set of natural, integers, rational and reals numbers. Let Σ denote a finite or countably infinite alphabet. The set of all finite strings made up with elements of a set Σ, will be denoted by Σ^*. $\Sigma^{\mathbb{N}}$ will denote the set of infinite sequences of elements of Σ. The spaces $\{0,1\}^{\mathbb{N}}$ and $\mathbb{N}^{\mathbb{N}}$, will be called ***Cantor*** space and ***Baire*** space, respectively. Elements of $\Sigma^* \cup \Sigma^{\mathbb{N}}$ will be denoted by $\omega = \omega_1, \omega_2, ...$ (or $\sigma = \sigma_1, ..$ when $\Sigma = \mathbb{N}$). We will write $\omega_{1:k}$ to mean $\omega_1, .., \omega_k$. We shall write $x \sqsubseteq \omega$ if x is a prefix of ω, where $x \in \Sigma^*$ and $\omega \in \Sigma^* \cup \Sigma^{\mathbb{N}}$. Given a finite sequence $x \in \Sigma^*$ we denote by $[x] = \{\omega \in \Sigma^{\mathbb{N}} : x \sqsubseteq \omega\}$ the associated ***cylinder***. These sets generate the standard topology (the product of the discrete topology) which can be metrized: $d(\omega, \omega') = 2^{-\min\{i : \omega_i \neq \omega'_i\}}$.

1.2 Recursive functions

The theory of algorithms begins with the mathematical formulation of the intuitive notions of *mechanical* (or *effective* or *constructive*) procedure on symbolic objects. Several very different formalizations were independently proposed (by Turing, Church, Kleene, Post, Markov...) in the 30's. Each constructed *computation model* defines a class of integer functions which can be computed by some *algorithmic* (with respect to the model) procedure. Later, all this models have proved to be equivalent: they define the same class of functions, which we shall call ***partial recursive functions*** (for formal definitions see for example [HR87]). Besides, this fact supports a working hypothesis known as Church's Thesis, which states that every (intuitively formalizable) algorithm is a partial recursive function. As algorithms are allowed to run forever on some inputs, the functions they compute may be ***partial***, that is, not defined everywhere. A recursive function defined everywhere is called ***total***. The ***domain*** of a partial recursive function $\varphi : \mathbb{N} \to \mathbb{N}$, is the set $\text{dom}(\varphi) \subset \mathbb{N}$ of inputs on which the algorithm computing φ eventually halts. A set $E \subset \mathbb{N}$ is said to be ***semi-decidable*** if it is the domain of a partial recursive function. If $E = \text{dom}(\varphi)$, then the algorithm computing φ allows to "semi-decide" whether $n \in E$ in the sense that the algorithm halts on input n if and only if $n \in E$ (it runs forever otherwise). If E and $\mathbb{N} \setminus E$ are semi-decidable, then E is said to be ***decidable***. It is easy to see that a set E is decidable if and only if the ***indicator function*** of E, $1_E : \mathbb{N} \to \{0, 1\}$ defined by $1_E(n) = 1 \Leftrightarrow n \in E$, is a *total recursive function*. In this case there is an algorithm which halts on every input n, and answer "yes" if $n \in E$ or "no" if $n \notin E$. The ***range*** of a partial recursive function φ is the set $\text{range}(\varphi) = \{\varphi(n) : n \in \text{dom}(\varphi)\}$.

Strictly speaking, recursive functions only work on natural numbers, but this can be easily extended to the objects (thought of as "finite" objects) of any countable set once a bijection with integers has been chosen in a way that the "finite" objects can be *algorithmically* recovered, in some sense, from their numbers. For example, let $\{0, 1\}^*$ be the set of finite binary words and denote by $l(n) = s$ the n-th word in lexicographical order. This correspondence defines a bijection $l : \mathbb{N} \to \{0, 1\}^*$. Let $\langle\cdot\rangle : \{0, 1\}^* \to \mathbb{N}$ be its inverse. Note that we can recover, in an *effective* way, $l(n)$ from n and $n = \langle l(n)\rangle$ from $l(n)$. Recursive functions can then be defined over $\{0, 1\}^*$:

$$\varphi\colon\ \{0,1\}^* \to \{0,1\}^* \text{ is } \textbf{\textit{recursive}} \text{ if so is } \langle s\rangle \mapsto \langle\varphi(s)\rangle.$$

In exactly the same way, there are *effective* bijections between $\mathbb{N}$ and finite tuples of integers $(n_1, .., n_k)$. Let us denote these bijections by $\langle, , .., \rangle\colon \mathbb{N} \times \mathbb{N} \cdots \times \mathbb{N} \to \mathbb{N}$. Then

$$\varphi\colon \mathbb{N}^k \to \mathbb{N}^l \text{ is } \textbf{\textit{recursive}} \text{ if so is } \langle n_1, .., n_k\rangle \mapsto \langle\varphi(n_1, .., n_k)\rangle.$$

With this intuitive description it known that there exists an effective procedure to enumerate the class of all partial recursive functions, associating to each of them its ***Gödel number***. Hence there exists a universal recursive function $\varphi_u : \mathbb{N} \to \mathbb{N}$ satisfying for all $e, n \in \mathbb{N}$, $\varphi_u(\langle e, n\rangle) = \varphi_e(n)$ where e is the Gödel number of φ_e. In classical recursion theory, a set of natural numbers is called ***recursively enumerable*** (***r.e.*** for short) if it is the range of some partial recursive function. That is, if there exists an algorithm listing (or enumerating) the set. Given a partial recursive function φ_e one can always compute the gödel number e' of a *total* recursive $\varphi_{e'}$ such that $\text{range}(\varphi_e) = \text{range}(\varphi_{e'})$. Hence we can suppose that r.e. sets are the range of total recursive functions. We shall denote by E_e the r.e. set associated to φ_e, namely: $E_e = \text{range}(\varphi_e) = \{\varphi_u(\langle e, n\rangle) : n \in \mathbb{N}\}$. A basic result in recursion theory says that a set E is recursively enumerable if and only if it is semi-decidable. Moreover, this equivalence is *effective* in the sens that, if $E = E_e = \text{range}(\varphi_e)$ then one can compute the gödel number e' of a partial recursive function $\varphi_{e'}$ such that $\text{dom}(\varphi_{e'}) = E_e$ and, conversely, if $E = \text{dom}(\varphi_e)$ then one can compute the gödel number e' of a *total* recursive function $\varphi_{e'}$ such that $E = \text{range}(\varphi_{e'})$. We shall freely use this equivalence. We say that $\sigma \in \mathbb{N}^{\mathbb{N}}$ is a ***recursive sequence*** if there exists a total recursive φ such that $\sigma_n = \varphi(n)$ for all $n \in \mathbb{N}$.

Definition 1.2.0.1. A ***numbered set*** $\mathcal{O}$ is a countable set together with a surjection $\nu_{\mathcal{O}} : \mathbb{N} \to \mathcal{O}$ called the ***numbering***. We write o_n for $\nu(n)$ and call n the ***name*** of o_n. ⌟

Examples 1.2.0.1.

1. The set $\mathbb{Q}$ of rational numbers can be injectively numbered $\mathbb{Q} = \{q_0, q_1, \ldots\}$ in an *effective* way: the number i of a rational a/b can be computed from a and b, and vice versa. We fix such a numbering.

2. The set of partial recursive functions $\mathcal{R} = \{\varphi_e : e \in \mathbb{N}\}$ is a numbered set, the Gödel numbers being the names.

3. The collection $\{E_e = \text{range}(\varphi_e) : e \in \mathbb{N}\}$ of all r.e. subsets of $\mathbb{N}$ is a numbered set.

⌟

We will sometimes use the word *algorithm* instead of *recursive function* when the inputs or outputs are interpreted as finite objects. The operative power of an algorithm on the objects of such a numbered set obviously depends on what can be effectively recovered from their numbers.

1.2.1 Uniformity

All through this work, we will use recursive functions to define *computability* or *constructivity* notions on more general objects. Depending on the context, these objects will take particulars names (computable, recursively enumerable, constructively open, decidable, etc...) but the definition will always follow the scheme:

object x is **constructive** *if there exists a recursive φ: $D \to E$ satisfying* property *$P(\varphi, x)$*

for example,

a set $E \subset \mathbb{N}$ is recursively enumerable *if there exists a recursive φ: $\mathbb{N} \to \mathbb{N}$ satisfying $E = \mathrm{range}(\varphi)$.*

Each time, a *uniform version* will be implicitly defined:

a sequence $(x_i)_i$ is constructive **uniformly in *i*** *if there exists φ: $\mathbb{N} \times D \to E$ such that for all $i \in \mathbb{N}$, $\varphi_i := \varphi(i, \cdot)$ satisfy* property *$P(\varphi_i, x_i)$*

in our example,

a sequence $(E_i)_i$ is recursively enumerable uniformly in i *if there exists φ: $\mathbb{N} \times \mathbb{N} \to \mathbb{N}$ such that for all $i \in \mathbb{N}$, $\varphi_i := \varphi(i, \cdot)$ satisfy $E_i = \mathrm{range}(\varphi_i)$.*

Let us illustrate this in other context,

Definition 1.2.1.1. A real number $x \in \mathbb{R}$ is said to be ***computable*** if there exists a total recursive $\varphi : \mathbb{N} \to \mathbb{Q}$ satisfying $|x - \varphi(n)| < 2^{-n}$ for all $n \in \mathbb{N}$. ⌟

Hence by a sequence of reals $(x_i)_i$ computable ***uniformly in i*** we mean that there exists $\varphi : \mathbb{N} \times \mathbb{N} \to \mathbb{Q}$ such that for all $i \in \mathbb{N}$, $\varphi_i := \varphi(i, \cdot)$ satisfy $|x - \varphi_i(n)| < 2^{-n}$ for all $n \in \mathbb{N}$.

1.3 Computability on symbolic spaces

The first step is to extend the notion of recursive function to functions over the space of infinite sequences on some alphabet Σ (finite or countably infinite). Intuitively, a function $F : D \subset \Sigma^{\mathbb{N}} \to \Sigma^{\mathbb{N}}$ is computable on D if there exists an algorithm $\mathcal{A}_F$ which computes the image $F(\omega)$ of any $\omega \in D$ in the following sense: the user enters some precision n (interpreted as accuracy 2^{-n}) to the algorithm which, after asking finitely many times the user for finite prefix of ω, halts outputting $F(\omega)_{1:n}$ which corresponds to a finite approximation of $F(\omega)$ up to 2^{-n}. In other words, the algorithm computing F is able to output finite prefixes of

$F(\omega)$, having access to finite prefixes of ω. Thus, if in some input $x \in \Sigma^*$ the algorithm outputs $y \in \Sigma^*$ then for any extension of x, $\mathcal{A}_F$ must output an extension of y. One can see that what underlies the algorithm $\mathcal{A}_F$ is a monotonic (with respect to the prefix order) recursive function $\varphi : \Sigma^* \to \Sigma^*$. Formally we have:

Definition 1.3.0.2. [Computability on symbolic spaces] For each monotonic $\varphi : \Sigma^* \to \Sigma^*$ (that is $x' \sqsubseteq x \Rightarrow \varphi(x') \sqsubseteq \varphi(x)$) we denote by $\overline{\varphi} : \Sigma^{\mathbb{N}} \to \Sigma^* \cup \Sigma^{\mathbb{N}}$ the function defined by:

$$\overline{\varphi}(x) = \sup_{x' \sqsubseteq x} \varphi(x')$$

and call $\mathrm{dom}(\overline{\varphi})$ the set of those $x \in \Sigma^{\mathbb{N}}$ for which $\overline{\varphi}(x) \in \Sigma^{\mathbb{N}}$. A function $F : \Sigma^{\mathbb{N}} \to \Sigma^{\mathbb{N}}$ is said to be ***recursive*** on $D \subset \Sigma^{\mathbb{N}}$ (and we call D, the domain of computability) if there exists a recursive monotonic function $\varphi : \Sigma^* \to \Sigma^*$ such that $D \subset \mathrm{dom}(\overline{\varphi})$ and $F = \overline{\varphi}$ on D. ⌟

Example 1.3.0.1.

The shift transformation $\sigma : \Sigma^{\mathbb{N}} \to \Sigma^{\mathbb{N}}$ defined by $\sigma(\omega)_i = \omega_{i+1}$ is the most simple recursive transformation. ⌟

1.4 Representations and constructivity

Let us discuss somewhat about extending the notion of computability to more general spaces. Often, given some space with a well defined structure S (a topological space for example), it is rather natural to define an *effective* version and then to obtain computability notions (of elements or functions for instance) based on recursive functions over finite objects. This approach has the advantage of allowing a rather elegant framework, giving more importance to structure than to the very way in which computations are performed. However, when a given set X comes with a natural structure which is different from S, there is in general no direct way to define computability over it. In this case, there are two possibilities: (i) force the structure S to appear somehow over X or (ii) set up new computability notions with respect to the new structure. In the first approach, the computability notions induced over X might be not the interesting ones (see example 1.6.1.1, 2 and remark 1.6.1.1, 1) and the second may seem somewhat arbitrary in some situations.

Another way to proceed is to "represent" infinite objects via "codings" into infinite sequences of integers, which allow to directly transfer all computability notions to any represented set.

Definition 1.4.0.3. A ***representation*** on a set X is a surjective (partial) function $\rho : \mathbb{N}^{\mathbb{N}} \to X$. If $\rho(\sigma) = x$, we say that σ is a ***description*** of x. ⌟

This approach has the advantage of allowing a single general definition of computability. On the other hand, caring about the role of codes when doing computations needs a lot of notation and may quickly become confuse, specially when many represented sets are involved, each with its own code.

We shall use representations but, as all spaces we will consider, will come with its own structure, each of them will have a *canonical* representation induced by its structure and this will allow us to *translate* all the computability notions induced by the representation in terms of the structure of the space. Thus, the way in which objects are coded will be only implicitly present.

Let X and Y be sets with fixed representations ρ_X and ρ_Y respectively. Computability can then be extended to any represented set as follows:

Definition 1.4.0.4 (Constructivity notions)**.**

1. An element $x \in X$ is ***constructive*** if there is a recursive sequence σ such that $\rho_X(\sigma) = x$.

2. A function $f :\subseteq X \to Y$ is ***constructive*** on $D \subseteq X$ if there exists a recursive function $F : \mathbb{N}^{\mathbb{N}} \to \mathbb{N}^{\mathbb{N}}$ which ***realizes*** f on D, that is, the following diagram commutes on $\rho_X^{-1}(D)$:

$$\begin{tikzcd} \mathbb{N}^{\mathbb{N}} \arrow[r, "F"] \arrow[d, "\rho_X"'] & \mathbb{N}^{\mathbb{N}} \arrow[d, "\rho_Y"] \\ X \arrow[r, "f"] & Y \end{tikzcd} \qquad \text{(that is, } f \circ \rho_X = \rho_Y \circ F \text{ on } \rho_X^{-1}(D))$$

⌟

Let $f :\subseteq X \to Y$ be constructive on D. Then if $x \in D$, for every description σ of x, $F(\sigma)$ is a description of the same element y, in this case $y = f(x)$. If a function F satisfy this property for some x, we say that it is ***extensional*** on x and that y is ***x-constructive***. If x is constructive then x-constructivity and constructivity are equivalent. An x-constructive element may induce a function $f : X \to Y$ which is constructive only on $\{x\}$. Note that two sequences of natural numbers can be merged into a single one, so the product $X \times Y$ of two represented sets has a canonical representation. In particular, it makes sense to speak about (x, y)-constructive elements.

One of the advantages of defining constructivity with the help of representations, is that the constructivity of the composition of constructive functions follows directly and *once for all.*

Proposition 1.4.0.1. *Let X, Y and Z be represented spaces. If $g : Y \to Z$ and $f : X \to \mathrm{dom}(g) \subset Y$ are constructive then the composition $g \circ f : X \to Z$ is constructive.*

Proof. It follows from the fact that the composition of recursive functions is recursive. □

In the following, we shall use representations to induce the notion of *constructive* or *computable* function between more general spaces. In each case we will characterize this functions in terms of structure and we shall use these characterizations later in order to prove that a given function is computable (or constructive, or semi-computable, etc...). Hence, computability could have been defined using the corresponding characterizations and without the help of representations (of course in this case, the computability of the composition would need to be proved case by case).

1.5 Enumerative Lattices

The existence of a universal recursive function φ_u is a remarkable fact. In many situations, it allows to *uniformly enumerate* the collection of constructive elements of a given space. For example, in the space of subsets of integers, the constructive elements are the semi-decidable (or r.e.) sets (which are only semi-computable) and, as we have already seen, the universal function induces a recursive enumeration of all the r.e sets $(E_e)_e$ such that E_e is r.e *uniformly in e*. On the contrary, for decidable sets (which are computable in a stronger sense) there is no such enumeration. This is a common situation in computability theory. In order to grasp *semi-computability* and express it in a general way we introduce the notion of *enumerative lattice*.

1.5.1 Definitions

Let us recall some concepts:

Definition 1.5.1.1. A ***complete lattice*** is a partial order $(L, \leq)$ in which every subset has both a supremum and an infimum. ⌟

If $(L, \leq)$ is a complete lattice, we will denote by $\perp$ and $\top$ the *least* and the *greatest* elements, that is, the supremum and the infimum of the empty set, respectively.

Examples 1.5.1.1.

1. The power set of a given set, ordered by inclusion. Supremum is given by the union and the infimum by the intersection.

2. The real line with the standard ordering.

3. The topology of a given topological space, ordered by inclusion. Supremum is given by the union and the infimum by the *interior* of the intersection.

⌟

We say that function $f : L \to L'$ of complete lattices is ***monotone*** if it preserves the order. If it preserves suprema and infima (hence the order), then we call it a ***morphism***.

Definition 1.5.1.2. An ***enumerative lattice*** is a triple $(L, \leq, \mathcal{P})$ where $(L, \leq)$ is a complete lattice and $\mathcal{P} \subseteq L$ (called ***simple elements***) is a numbered set such that every element $x \in L$ is the supremum of some subset of $\mathcal{P}$. ⌟

Definition 1.5.1.3. Given an enumerative lattice $(L, \leq, \mathcal{P})$, the ***canonical representation*** is defined by $\rho_L(\sigma) = \sup_n p_{\sigma_n}$. ⌟

From here and beyond, when dealing with enumerative lattices the canonical representation will be implicitly used. Hence, canonical constructivity notions derives directly from definition 1.4.0.4.

1.5.2 Constructive elements

Let $(L, \leq, \mathcal{P})$ be an enumerative lattice. From definition 1.4.0.4, it follows that

Definition 1.5.2.1. [Constructive elements] An element $x \in L$ is ***constructive*** if there is a recursive function $\varphi\colon \mathbb{N} \to \mathbb{N}$ such that $x = \sup\{p_{\varphi(n)} : n \in \mathbb{N}\}$. ⌟

Examples 1.5.2.1.

1. $(\{\bot, \top\}, \leq, \{\top\})$ with $\bot < \top$ is the most simple example.

2. $(2^{\mathbb{N}}, \subseteq, \{\text{finite sets}\})$ is an enumerative lattice. The constructive elements are the r.e sets from classical recursion theory.

3. $(\overline{\mathbb{R}}, \leq, \mathbb{Q})$ with $\overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty, +\infty\}$ is an enumerative lattice: the constructive elements are the so-called ***lower semi-computable*** real numbers.

⌟

The following result is one of the main reasons to introduce enumerative lattices:

Proposition 1.5.2.1. *Let $(X, \leq, \mathcal{P})$ be an enumerative lattice. There is an enumeration $(x_i)_{i \in \mathbb{N}}$ of all the constructive elements of X such that x_i is constructive uniformly in i.*

Proof. The universal recursive function φ_u induce a recursive enumeration of all constructive elements of X. Indeed, if e is the gödel number of the recursive function making x constructive, then $x = \sup_{n \in \mathbb{N}} p_{\varphi_u(\langle e,n \rangle)}$. □

The following easy lemma may also be usefull.

Lemma 1.5.2.1. *The supremum of a uniform sequence of constructive elements is again constructive.*

Definition 1.5.2.2. Two sets of simple elements $\mathcal{P}$ and $\mathcal{P}'$ are said to be ***equivalent*** if any $p_i \in \mathcal{P}$ is constructive in $(X, \leq, \mathcal{P}')$ uniformly in i and vice-versa. ⌟

1.5.3 Constructive functions

The following characterization of (total) constructive functions between enumerative lattices will be useful in proving constructivity. Let $(L, \leq, \mathcal{P})$ and $(L', \leq, \mathcal{P}')$ be two enumerative lattices and let $(x_i)_i$ the collection of constructive elements of L.

Lemma 1.5.3.1. *A function of enumerative lattices $f : L \to L'$ is constructive iff the following conditions hold:*

1. *f is monotone and commutes with suprema of infinite increasing sequences.*

2. *$f(x_i) \in L'$ is constructive uniformly in i.*

Proof. Let f be constructive. Let F be the function realizing f. Let $x \leq y$ in L. Let σ, σ' be descriptions of x and y respectively. We note that for any n, $\sigma_{1:n}\sigma'$ is another description of y. Hence, since there exists n such that $F[\sigma_{1:n}] \subset [F(\sigma)_1]$, we conclude $f(x) \leq f(y)$ and condition 1 now follows from the fact that F realizes f. Let x_i be constructive (in i) and let σ be a recursive description of it. Then $F(\sigma)$ is a recursive description of $f(x_i)$ which is then constructive, uniformly in i. Condition 2 follows. For the converse, let $\sigma \in \mathbb{N}^{\mathbb{N}}$. We define $F(\sigma)$ as follows. For sake of clarity, denote $n(\sigma)$ the name of the constructive element $\sup_{i \leq n} p_{\sigma_i}$ (that is, $\sup_{i \leq n} p_{\sigma_i} = x_{n(\sigma)}$) and by $\varphi : \mathbb{N}^2 \to \mathbb{N}$ the function making $f(x_i)$ uniformly constructive (that is $f(x_i) = \sup_n p_{\varphi(i,n)}$) which exists by condition 2. Then we define $F(\sigma)$ as $(F(\sigma))_{\langle i,j \rangle} := \varphi(i(\sigma), j)$. Clearly, F is recursive and by condition 1 it realizes f. □

Remarks 1.5.3.1.

1. The proof of the above lemma even shows condition 2 can be replaced by: there exists $\varphi : \mathbb{N} \to \mathbb{N}$ such that $f(\sup_{j \leq k} p_{i_j}) = y_{\varphi\langle i_1,..,i_k \rangle}$ for all $\langle i_1, .., i_k \rangle \in \mathbb{N}$. That is, the image of suprema of finitely many simple elements is uniformly constructive.

2. In *domain theory*, conditions 1 and 2 are known as *Scott-continuity*. For more on the relations between the enumerative-lattice approach and domain theory we refer to [Hoy08].

⌟

For morphism, constructivity takes a much simpler form:

Proposition 1.5.3.1. *A morphism $f : L \to L'$ is constructive if and only if the images of simple points are uniformly constructive.*

Proof. It follows directly from the property $f(\sup\{p_{i_1}, .., p_{i_k}\}) = \sup\{f(p_{i_1}), .., f(p_{i_k})\}$. □

Definition 1.5.3.1. An ***isomorphism*** of enumerative lattices is a bijective constructive morphism, with constructive inverse. ⌟

1.6 Computable Metric Spaces

A *computable metric space* is a separable metric space with a "distinguished" countable set (whose elements are called ***ideal points***) satisfying a computability condition: the distances between ideal points can be computed by an algorithm with arbitrary (but finite) precision. With this structure, the computable elements, sets or functions are naturally defined. Following [Wei00] we define:

Definition 1.6.0.2. A ***computable metric space*** (CMS) is a triple $\mathcal{X} = (X, d, \mathcal{S})$, where

- (X, d) is a separable complete metric space.
- $\mathcal{S} = (s_i)_{i\in\mathbb{N}}$ is a numbered dense subset of X (called ***ideal points***).
- The real numbers $(d(s_i, s_j))_{i,j}$ are all computable, uniformly in i, j.

⌟

The third condition in the above definition determines the effectivity of the numbering of $\mathcal{S}$: from the names (i and j) of s_i and s_j, we can uniformly recover their mutual distance.

Examples 1.6.0.1.

1. Symbolic spaces $(\Sigma^{\mathbb{N}}, d, S)$ with Σ a finite alphabet. Let us fix some element of Σ denoting it by 0. The dense set S is the set of ultimately 0-stationary sequences.

2. $(\mathbb{R}^n, d_{\mathbb{R}^n}, \mathbb{Q}^n)$ with the Euclidean metric and the standard numbering of $\mathbb{Q}^n$.

3. If (X_1, d_1, S_1) and (X_2, d_2, S_2) are two computable metric spaces, then the product space $(X_1 \times X_2, d, S_1 \times S_1)$ is a computable metric space where the distance d is given by $d((x_1, x_2), (y_1, y_2)) = \max(d_1(x_1, y_1), d_2(x_2, y_2))$.

⌟

The numbered set of ideal points $(s_i)_i$ induces the numbered set of ***ideal balls*** $\mathcal{B} := \{B(s_i, q_j) : s_i \in S, q_j \in \mathbb{Q}_{>0}\}$. We denote by $B_{i,j}$ the ideal ball $B(s_i, q_j)$. From the name of an ideal ball, we can recover the names of its center and radius, and vice-versa. We identify each point $x \in X$ to the collection of all ideal balls to which x belongs. We call this collection a ***complete description*** of x.

Definition 1.6.0.3. In a computable metric space $(X, d, \mathcal{S})$ the ***canonical representation*** $\rho_X : \mathbb{N}^{\mathbb{N}} \to X$ is defined by $\rho_X(\sigma) = x$ if $(B_{\sigma_n})_n$ is a complete description of x. ⌟

Again, when working with computable metric spaces, we will implicitly use the canonical representation. Hence, constructivity notions derive directly from definition 1.4.0.4. Usually, when dealing with computable metric spaces, constructive functions or points are called ***computable***. We shall do likewise.

1.6.1 Computable sets

Any open set is the union of some collection of ideal balls. Then we directly have:

Proposition 1.6.1.1. *Let $\tau_X \subset 2^X$ be the topology of X. Then the triple $(\tau_X, \subset, \mathcal{B})$ is an enumerative lattice.*

Proof. Straightforward. □

Definition 1.6.1.1. A set $U \subset X$ is called ***constructively open*** if it is a constructive element of the enumerative lattice $(\tau_X, \subset, \mathcal{B})$. That is, if there exists a recursive $\varphi : \mathbb{N} \to \mathbb{N}$ such that $U = \bigcup_{i \in \mathbb{N}} B_{\varphi(i)}$. ⌟

If D is an arbitrary subset of X, a set U is ***constructively open in D*** is there exists a constructively open U' such that $U = U' \cap D$, equivalently, if U is a constructive element of the enumerative lattice $(\tau_X \cap D, \subset, \mathcal{B})$ where $\tau_X \cap D := \{O \cap D : O \in \tau_X\}$ is the *trace* topology on D.

Examples 1.6.1.1.

1. In $\mathbb{N}^\mathbb{N}$ (or $\{0,1\}^\mathbb{N}$), ideal balls are cylinders. Then a set $U \subset \mathbb{N}^\mathbb{N}$ is constructively open if there exists $\varphi : \mathbb{N} \to \mathbb{N}^*$ such that $U = \cup_n [\varphi(n)]$.

2. Given a constructively open set A and an ideal ball B_i, the relation $B_i \subset A$ is in general not semi-decidable. Consider for example $[0,1]$ with the euclidean distance and the rationals as ideal points. Define $E = \{\langle i, n\rangle \in \mathbb{N}$ such that $n \geq i+1$ and $\varphi_i(i)$ does not stop in less than n steps$\}$. This is a r.e subset of $\mathbb{N}$. Then the following set is constructively open:

$$A = \bigcup_{\langle i,n\rangle \in E} \left(2^{-(i+1)}(1+2^{-n}), 2^{-i}\right)$$

and it is easy to see that $(2^{-(i+1)}, 2^{-i}) \subseteq A$ if and only if $\varphi_i(i)$ does not stop.

⌟

Remarks 1.6.1.1.

1. Any enumerative lattice $(L, \leq, \mathcal{P})$ can be turned into a constructive topological space[1] by defining for each $p \in \mathcal{P}$ the subbasis element $V_p := \{x \in L : p < x\}$. In particular,

[1]That is, a topological space with a countable subbase ν effectively given, and such that every point is uniquely determined by the collection of elements of ν containing it. A point is then called computable if this collection is recursively enumerable.

this can be done for the enumerative lattice $(\tau_X, \subset, \mathcal{B})$. If an element A of τ_X is computable in the corresponding constructive topological space then it is constructively open. The example 2 above shows that the converse is false. Then the two notions are not equivalent.

2. The computability of the distance implies that the relations $d(s_i, s_j) < q_n$ and $d(s_i, s_j) > q_n$ are semi-decidable, uniformly in i, j, n. Equivalently, one can semi-decide whether $s_i \in B(s_j, q_n)$, and whether $s_i \notin \overline{B}(s_j, q_n) = \{x : d(s_j, x) \leq q_n\}$ (uniformly in i, j, n).

3. In the case of $\mathbb{N}^{\mathbb{N}}$ or $\{0,1\}^{\mathbb{N}}$ (see example 1 above) cylinders are even *decidable* since, being clopen sets, they have no boundary. In connected spaces, the only *decidable* sets are X and $\emptyset$.

⌟

Definition 1.6.1.2. If $B(s, q)$ and $B(s', q')$ are ideal balls, we write $B(s, q) < B(s', q')$ (it reads ***manifestly included***) to mean $d(s', s) + q < q'$. ⌟

A basic topological characterization says that a set U is open if and only if for any point $x \in U$ there exists $\epsilon > 0$ for which $B(x, \epsilon) \subset U$. Here is the effective version: for each ideal point $s \in \mathcal{S}$ one can semi-decide whether $s \in U$ and, if $s \in U$ then one can also find some rational q such that $B(s, q) \subset U$.

Proposition 1.6.1.2. *U is constructively open if and only if there exists a partial recursive φ: $\mathcal{S} \to \mathbb{Q}$ such that:*

(i) $\varphi(s)$ halts iff $s \in U$ and

(ii) $B(s, \varphi(s)) \subset U$ holds.

Proof. The if part follows from the equality $U = \bigcup_{s \in U} B(s, \varphi(s))$. The only if part follows from remark 2: for each ideal point s and ideal ball $B_{i,j}$, semi-decide whether $s \in B_{i,j}$ and compute some $0 < \epsilon \in Q$ satisfying $\epsilon < r_j - d(s, s_i)$ (or equivalently $B(s, \epsilon) < B_{i,j}$). □

Let (X, S_X, d_X) and (Y, S_Y, d_Y) be computable metric spaces. Denote by $\mathcal{U}^X$ and $\mathcal{U}^Y$ the corresponding collections of recursively open sets. By proposition 1.5.2.1, there exists an enumeration of $\mathcal{U}^X = (U_e^X)_{e \in \mathbb{N}}$ such that U_e^X is constructively open, uniformly in e. And the same holds for $\mathcal{U}^Y$.

The following is an important property.

Proposition 1.6.1.3. *The union $\cup_{i\leq k}U_{e_i}$ is constructively open uniformly in $\langle e_1,..,e_k\rangle$. The same holds for the intersection.*

Proof. We use proposition 1.6.1.2. Let U_i and U_j be constructively open. Then, given an ideal point s, we can semi-decide whether $s\in U_i$ *or* $s\in U_j$ and compute some ϵ such that $B(s,\epsilon)\subset U_i$ *or* $B(s,\epsilon)\subset U_j$. Hence we can semi-decide whether $s\in U_i\cup U_j$ and compute ϵ such that $B(s,\epsilon)\subset U_i\cup U_j$. In the same way, we can semi-decide whether $s\in U_i$ *and* $s\in U_j$, and compute ϵ and ϵ' such that $B(s,\epsilon)\subset U_i$ *and* $B(s,\epsilon')\subset U_j$. Hence we can semi-decide whether $s\in U_i\cap U_j$ and compute $\delta=\min\{\epsilon,\epsilon'\}$ such that $B(s,\delta)\subset U_i\cap U_j$. □

1.6.2 Computable functions

Let $F:\{0,1\}^{\mathbb{N}}\to\{0,1\}^{\mathbb{N}}$ be a recursive function and let $\varphi:\{0,1\}^*\to\{0,1\}^*$ be the monotonic recursive function allowing to compute F (see definition 1.3.0.2). We note that for any $w,x\in\{0,1\}^*$, one can semi-decide whether $w\sqsubseteq\varphi(x)$ and then the set $F^{-1}[w]=\{x:w\sqsubseteq\varphi(x)\}$ is constructively open, uniformly in w. In words: the preimage of cylinders are uniformly constructively open sets. This can be thought as the recursive version of continuity. Moreover, this *recursive continuity* is actually equivalent to recursivity:

Proposition 1.6.2.1. *Let $F:\mathbb{N}^{\mathbb{N}}\to\mathbb{N}^{\mathbb{N}}$ be such that for each $y\in\mathbb{N}^*$ the set $F^{-1}[y]$ is constructively open, uniformly in y. Then F is recursive.*

Proof. For $x\in\mathbb{N}^*$, we define $\varphi(x)$ by induction in $|x|$. If Λ is the empty word, let $\varphi(\Lambda)=\Lambda$. Suppose φ has been defined over $\mathbb{N}^n=\{x\in\mathbb{N}^*:|x|=n\}$ and let $\varphi(x)=y$ for some $x\in\mathbb{N}^*$. We have to define $\varphi(xi)$ for some $i\in\mathbb{N}$ to be an extension of y. The idea is to decide whether the additional bit i (in xi) suffices to give an additional bit j[2] in which case case we will put $\varphi(xi)=yj$ and $\varphi(xi)=y$ otherwise[3]. □

As for symbolic spaces, the computability of a function between computable metric spaces is equivalent to a sort of *recursive continuity*:

[2]That is whether there exists j such that $[xi]\subset F^{-1}[yj]$ which, as $F^{-1}[yj]$ is constructively open, can be semi-decided.

[3]That is, when there exists k and k' such that $k\neq k'$ and $[xi]\cap F^{-1}[yk]\neq\emptyset$ and $[xi]\cap F^{-1}[yk']\neq\emptyset$, which can be semi-decided too.

Proposition 1.6.2.2 (Computable Functions)**.** *A function $T : X \to Y$ is* **computable** *iff $\hat{T}$: $\tau_Y \to \tau_X$ defined by $\hat{T}(U) := T^{-1}(U)$, is a constructive function of enumerative lattices.*

Proof. By lemma 1.5.3.1, (since $\hat{T}$ is a morphism) $\hat{T}$ is a constructive function of enumerative lattices if and only if the preimage $T^{-1}(U_e)$ of a recursively open set $U_e \subset Y$ is constructively open, uniformly in e (recursive continuity). The result now follows from the following claim:

Claim 1.6.2.1. *Let X be a computable metric space. A set $U \subset X$ is constructively open iff $\rho_X^{-1}(U)$ is constructively open in $\rho_X^{-1}(X) \subset \mathbb{N}^{\mathbb{N}}$.*

Proof. Let $U \subset X$ be constructively open. To show that so is $\rho_X^{-1}(U)$ it is enough to show that $\rho_X^{-1}(B_i)$ is constructively open uniformly in i. Let B_i be an ideal ball. If $\rho_X(\sigma) \in B_i$, then there exists some n such that $B_{\sigma_n} < B_i$ (which can be semi-decided) and then, $[\sigma_{1:n}] \subset \rho_X^{-1}(B_i)$. Thus, it suffice to enumerate all finite words $s \in \mathbb{N}^*$ and to keep those for which the test $B_{s_{|s|}} < B_i$ stops. Let us call $S \subset \mathbb{N}^*$ the set so obtained. Clearly, $\rho_X^{-1}(B_i) = \cup_{s \in S}[s] \cap \rho_X^{-1}(X)$. For the converse observe that for any finite word s, the set $\cap_{i \leq |s|} B_{s_i}$ is recursively open, uniformly in s. Then U is the union of a uniform collection of constructively open sets and thus it is constructively open itself. □

□

It follows that computable functions are continuous. Since $\hat{T}$ is a morphism, T is computable iff $T^{-1}(B_n)$ is constructively open uniformly in n, or equivalently, the preimages of ideal balls are uniformly constructively open. Proposition 1.6.1.2 directly implies the following usefull characterization.

Corollary 1.6.2.1. *A function $T : X \to Y$ is computable if and only if there is a partial recursive φ: $\mathcal{S} \times \mathbb{N} \to \mathbb{Q}$ such that:*

(i) $\varphi(s, e)$ halts iff $T(s) \in U_e^Y$ and

(ii) $B(s, \varphi(s, e)) \subset T^{-1}(U_e^Y)$ holds.

As an elementary application, let us show that:

Proposition 1.6.2.3. *For any ideal point $s_n \in \mathcal{S}$, the distance function $d : X \to \mathbb{R}$ defined by $d(x) = d(s_n, x)$ is computable, uniformly in n.*

Proof. Let $I_{i,j} = (q_i, q_j)$ be a rational interval. As $d(s_n, s_k)$ is computable, given an ideal point s_k we can semidecide whether $d(s_n, s_k) \in I_{i,j}$ and compute some ϵ for which $(d(s_n, s_k)-\epsilon, d(s_n, s_k)+\epsilon) \subset I_{i,j} \Leftrightarrow B(s_k, \epsilon) \subset d^{-1}I_{i,j}$. The results follows from corollary 1.6.2.1. $\square$

Partially computable functions Since we will work with functions which are not necessarily continuous everywhere, we shall consider functions which are computable on some subset $D \subset X$. We obtain such a notion by just replacing τ_X with $\tau_X \cap D = \{U \cap D : U \in \tau_X\}$ in Proposition 1.6.2.2. In other words, a function T is ***computable on D*** ($D \subset X$) if $T^{-1}(U_e^Y) \cap D = U_{e'}^X \cap D$ where $U_{e'}^X$ is constructively open uniformly in e.

1.6.3 Computable points

From definition 1.4.0.4 follows that a point x is computable if the collection of ideal balls containing x is recursively enumerable. There is another way in which computable points are usually described. The density of the set of ideal points allows to "reach" the whole space by means of *finite approximations*. Let us say that the sequence of points $(x_i)_i$ is ***fast*** if $d(x_i, x_{i+1}) < 2^{-i}$ for all $i \in \mathbb{N}$. As the space is complete, every fast sequence has a limit. The following proposition may be helpful.

Definition 1.6.3.1. A sequence of points x_n is said to ***converge effectively*** to a limit x if for all $n \geq 0$, $d(x_n, x) \leq f(n)$ where $f : \mathbb{N} \to \mathbb{R}$ is computable and decreases to zero. ⌟

Proposition 1.6.3.1. *The following are equivalent:*

1. *x is computable*

2. *there is a recursive fast sequence of ideal points converging x*

3. *There is a sequence of uniformly computable points x_n which converges effectively to x.*

Proof. Let x be computable. Hence, one can find ideal balls containing x with radius as small as we want. Taking their centers, one can enumerate a fast sequence of ideal points converging to x. A recursive fast sequence of ideal points converging to x, does it effectively. If x_n converges effectively to x, the we can extract a subsequence which is fast. Call it x_i. Now, by what has been proved before, one can construct a fast sequence of ideal points converging to x. Let it be s_i. We have $d(s_i, x) < 2^{-i+1}$ for all i. Then, given an ideal ball

$B(s,q)$ containing x, there exists some i such that $B(s_i, 2^{-i+1}) < B(s,q)$, which we can effectively find. □

Remarks 1.6.3.1.

1. If T is computable then the images of ideal points can be uniformly computed, that is: $T(s_i^X)$ is computable, uniformly in i.

2. More generally, if T is computable then there exists an algorithm which computes the image $T(x)$ of any x in the following sense: the user enters some rational ϵ to the algorithm which, after asking finitely many times the user for finite approximations of x, halts outputting a finite approximation of $T(x)$ up to ϵ.

⌟

1.7 Computability on some functions spaces

1.7.1 The enumerative lattice $C(X,L)$

Let $(X,d,\mathcal{S})$ be a computable metric space and $(L,\leq,\mathcal{P})$ an enumerative lattice. The partial order $\leq$ on L induce the *pointwise partial order* $\preceq$ over functions from X to L, defined by $f \preceq g \Leftrightarrow f(x) \leq g(x)\ \forall x \in X$. The relation $\preceq$ makes this space of functions a complete lattice. We shall be interested in some functions from this space. Let us consider the numbered set $\mathcal{F}$ of ***open step functions*** from X to L:

$$f_{B_i}^{p_j}(x) = \begin{cases} p_j & \text{if } x \in B_i \\ \perp & \text{otherwise} \end{cases}$$

We then define $C(X,L)$ as the closure of $\mathcal{F}$ under pointwise suprema, with the pointwise ordering. By definition, we have:

Proposition 1.7.1.1. *$(\mathcal{C}(X,Y),\preceq,\mathcal{F})$ is an enumerative lattice.*

Now we prove that (partial) constructive functions from X to L can always be extended to be total and that these are exactly the constructive elements of $C(X,L)$.

We remark that a recursive function $F : \mathbb{N}^{\mathbb{N}} \to \mathbb{N}^{\mathbb{N}}$ realizes some function $\hat{f} : X \to L$ computable on the set $D := \{x \in X : F \text{ is extensional on } x\}$. The following lemma shows that F also defines a constructive element $f \in C(X,L)$ which coincides with $\hat{f}$ on D.

Lemma 1.7.1.1. *Any recursive function $F : \mathbb{N}^{\mathbb{N}} \to \mathbb{N}^{\mathbb{N}}$ defines a constructive element $f \in C(X, L)$ such that for every $x \in X$ on which F is extensional, if σ describes x then $F(\sigma)$ describes $f(x)$.*

Proof. Let $F : \mathbb{N}^{\mathbb{N}} \to \mathbb{N}^{\mathbb{N}}$ be recursive. Then there exists φ such that $F^{-1}[w] = \cup_n[\varphi(w, n))]$. We define a constructive element $f_\varphi \in C(X, L)$ by enumerating a uniform sequence of constructive elements. Let $w \in \mathbb{N}^*$ and denote by $x_w = \sup_{i \leq |w|} p_{w_i}$. Denote by $B_{n(w)} \subset X$ the constructively open set represented by $[\varphi(w, n)]$. Then define $f_\varphi = \sup\{f^{x_w}_{B_{n(w)}} : w \in \mathbb{N}^*, n \in \mathbb{N}\}$. Suppose that F is extensional on x. Let σ be a description of x. Then by construction we have: $f_\varphi(x) = \sup_i x_{F(\sigma)_{1:i}}$. □

Lemma 1.7.1.2. *Any constructive element of $C(X, L)$ is a (total) constructive function.*

Proof. Let $f = \sup\{f^{p_j}_{B_i} : \langle i, j\rangle \in E\}$ with E r.e. From a description of x, we can semidecide wether $x \in B_i$ for any ideal B_i. Then dovetail $\langle i, j\rangle$ and output j if $x \in B_i$. The sequence so obtained is independent of the given description of x, up to permutations. Then all them describe the same element of L, namely $\sup_n F(\sigma)_n = f(x)$. □

Corollary 1.7.1.1. *For any $f : X \to L$ constructive on D there exists a total constructive $\hat{f}$ which coincides with f on D.*

Proof. Let F be the function realizing f. F is extensional on D. By lemma 1.7.1.1, F defines a constructive element $\hat{f} \in C(X, L)$ which coincides with f on D. By lemma 1.7.1.2, $\hat{f}$ is a total constructive function. □

The proof even shows that the equivalence is constructive: the evaluation of any $f : X \to L$ on any $x \in X$ can be achieved by an algorithm having access to any description of $f \in \mathcal{C}(X, L)$, and any recursive F realizing f can be converted into an algorithm describing $f \in C(X, L)$. More precisely:

Proposition 1.7.1.2. *Let X be a computable metric space, L be an enumerative lattice and Y be any represented space. Let $g : Y \times X \to L$ and $h : Y \to C(X, L)$ be constructive functions, then the following functions are constructive:*

$$\begin{aligned} Eval : \mathcal{C}(X, L) \times X &\longrightarrow L \\ (f, x) &\longmapsto f(x) \end{aligned} \tag{1.1}$$

$$\begin{aligned} Curry(g) : Y &\longrightarrow C(X, L) \\ y &\longmapsto g(y, \cdot) \end{aligned} \tag{1.2}$$

$$\begin{aligned} Decurry(h) : Y \times X &\longrightarrow L \\ (y, x) &\longmapsto h(y)(x) \end{aligned} \tag{1.3}$$

1.7.2 Lower semi-computability

Let X be a computable metric space. The following is a standard definition.

Definition 1.7.2.1. A real valued function $f : X \to \mathbb{R}$ is said to be ***lower semi-computable*** if $f^{-1}(q_i, \infty)$ is constructively open uniformly in i. f is ***upper semi-computable*** if $-f$ is lower semi-computable. ⌟

Proposition 1.7.2.1. *A function $f : X \to \mathbb{R}$ is computable iff it is upper and lower semi-computable.*

Proof. By proposition 1.6.2.2, f is computable iff $f^{-1}(q_i, q_j)$ is constructively open uniformly in i, j. The result follows from the relation $f^{-1}(q_i, q_j) = f^{-1}(q_i, +\infty) \cap f^{-1}(-\infty, q_j)$. □

In the same way that corollary 1.6.2.1, one can easily prove:

Proposition 1.7.2.2. *A function $f : X \to \mathbb{R}$ is lower semi-computable if and only if there is a partial recursive φ: $\mathcal{S} \times \mathbb{N} \to \mathbb{Q}$ such that:*

(i) $\varphi(s, i)$ halts iff $f(s) \in (q_i, \infty)$ and

(ii) $B(s, \varphi(s, i)) \subset T^{-1}(q_i, \infty)$ holds.

Now we analyze real valued functions from the point of view of enumerative lattices. Let us consider the real line as an enumerative lattice $(\overline{\mathbb{R}}, \leq, \mathbb{Q})$, and the space of functions $\mathcal{C}(X, \overline{\mathbb{R}})$ introduced above, which is also an enumerative lattice. In this case, the set $\mathcal{F}$ of step functions takes the form:

$$f_{B_i}^{q_j}(x) = \begin{cases} q_j & \text{if } x \in B_i \\ \perp & \text{otherwise} \end{cases}$$

Proposition 1.7.2.3. *$C(X, \overline{\mathbb{R}})$ is the set of lower semi-continuous functions. A lower semi-continuous function is constructive iff it is lower semi-computable.*

Proof. Let $f \in C(X, \overline{\mathbb{R}})$. Then there exists $E \subset \mathbb{N}$ such that $f = \sup\{q_j 1_{B_i} : \langle i, j \rangle \in E\}$. To see that f is lower semi-continuous, observe that $f^{-1}(q_k, \infty) = \cup_{i \in E'} B_i$ where $E' = \{i \in \mathbb{N} : q_j \geq q_k, \langle i, j \rangle \in E\}$. Moreover, if E is r.e then so is E'; constructive functions are then lower semi-computable. Conversely, if for any $k \in \mathbb{N}$, $f^{-1}(q_k, \infty)$ is open, then there is $E_k \subset \mathbb{N}$ such that $f^{-1}(q_k, \infty) = \cup_{i \in E_k} B_i$. Hence $f = \sup\{q_k 1_{B_i} : (i, k) \in \mathbb{N}^2, i \in E_k\}$. Again, if f is lower semi-computable then E_k is r.e uniformly in k. f is then constructive. □

As a corollary we have:

Corollary 1.7.2.1. *A function $f : X \to \mathbb{R}$ is computable iff there exists two recursive monotone sequences $f_n^{\uparrow}$ and $f_n^{\downarrow}$ of step functions such that for all $x \in X$, $f_n^{\uparrow}(x) \nearrow f(x)$ and $-f_n^{\downarrow}(x) \searrow f(x)$ (equivalently $f_n^{\downarrow}(x) \nearrow -f(x)$) holds.*

Proof. Direct from propositions 1.7.2.1 and 1.7.2.3 □

1.7.3 Computable *simple* functions

Following Gács, let us introduce a certain fixed, enumerated sequence of Lipschitz functions. Let $\mathcal{H}_0$ be the set of functions of the form:

$$g_{s,r,\epsilon} = |1 - |d(x, s) - r|^+/\epsilon|^+ \tag{1.4}$$

where $s \in S$, $r, \epsilon \in \mathbb{Q}$ and $|a|^+ = \max\{a, 0\}$. These are uniformly computable (composition of computable functions) Lipschitz functions equal to 1 in the ball $B(s, r)$, to 0 outside $B(s, r + \epsilon)$ and with intermediate values in between.

Let

$$\mathcal{H} = \{g_1, g_2, \ldots, \} \tag{1.5}$$

be the smallest set of functions containing $\mathcal{H}_0$ and the constant 1, and closed under max, min and finite rational linear combinations. We fix some enumeration $\nu_{\mathcal{H}}$ of $\mathcal{H}$ and we write g_n for $\nu_{\mathcal{H}}(n) \in \mathcal{H}$. Clearly, these is also a uniform family of computable functions.

Remark 1.7.3.1. These function are of the form $g_n = c_n + f_n$ where c_n is a constant (computable from n) and f_n has bounded support (and from n one can compute a bound for its diameter). ⌟

In particular, for any $f \in \mathcal{H}$, we have that f and $-f$ are constructive elements of $C(X, \overline{\mathbb{R}})$.

When working with positive real valued functions, the set $\mathcal{H}^+$ ($\mathcal{H}$ restricted to positive functions) can be seen as an alternative to step functions. Let us consider the enumerative lattice $C(X, \overline{\mathbb{R}^+})$ of positive functions. Here, $\overline{\mathbb{R}^+}$ denotes the enumerative lattice $(\mathbb{R}^+ \cup \{+\infty\}, \leq, \mathbb{Q}_+)$. In this space, functions in the set $\mathcal{F}$ (of step functions) are defined by:

$$f_{B_i}^{q_j}(x) = \begin{cases} q_j & \text{if } x \in B_i \\ 0 & \text{otherwise} \end{cases}$$

Proposition 1.7.3.1. *The enumerative lattices* $(C(X, \overline{\mathbb{R}^+}), \leq, \mathcal{F})$ *and* $(C(X, \overline{\mathbb{R}^+}), \leq, \mathcal{H}^+)$ *are isomorphic.*

Proof. The sets $\mathcal{F}$ and $\mathcal{H}^+$ are equivalent simple sets. Indeed, for $\epsilon_n = \frac{1}{n}$ we have that $f^q_{B(s,r)} = \sup\{q \cdot g_{s,r-\epsilon_n,\epsilon_n} : n \in \mathbb{N}\}$ and that $g_{s,r,\delta} = \sup\{f^{\epsilon_n}_{B(s,r,\delta(1-\epsilon_n))} : n \geq 1\}$. The identity is then an isomorphism between $(C(X, \overline{\mathbb{R}^+}), \leq, \mathcal{F})$ and $(C(X, \overline{\mathbb{R}^+}), \leq, \mathcal{H}^+)$. □

Hence, any positive lower semi-computable function, is the supremum of a computable sequence of functions from $\mathcal{H}^+$. The relation between computability and constructivity can only be stated for bounded functions.

Corollary 1.7.3.1. *Let* $f_i : X \to \mathbb{R}$ *be bounded by* M_i ($|f_i| \leq M_i$). *If* M_i *is computable uniformly in* i, *then* f_i *is uniformly computable iff* $f_i + M_i$ *and* $-f_i + M_i$ *are uniformly constructive elements of* $(C(X, \overline{\mathbb{R}^+}), \leq, \mathcal{H}^+)$.

Proof. First, we prove the result for positive functions. Let $M_i \geq f_i \in C(X, \overline{\mathbb{R}^+})$, where M_i is uniformly computable. The only if part follows directly from the above proposition. To prove the if part observe that, since $-f_i + M_i$ is constructive, $(-f_i + M_i)^{-1}(q_n, \infty) = f^{-1}(0, M_i - q_n)$ is uniformly constructively open. Then f_i is upper semi-computable. It is also lower semi-computable since $f_i + M_i$ is. To prove the result for bounded functions f_i which are not necessarily positive just apply this to $\hat{f} = f_i + M_i \leq 2M_i$. □

Chapter 2

Constructive Measure Theory

2.1 Introduction

Let X be a computable metric space. In this chapter, we investigate the notion of "computability" of probability measures over X (very simply defined in Cantor space). The computable structure of X induce a computable structure over the metric space of probability measures over X, $M(X)$. The computable elements of this space being the *computable measures*. We study the computability properties of the measure of opens sets as well as the integral of functions, obtaining that these quantities can be only partially computed. In order to properly state these results, we use the notion of enumerative lattice (1.5.1.2) which is more suited to express *semi-computability* in a general way. Then we give several characterizations of computable measure in terms of these last computability properties, and show the following result:

Theorem *Let $\mu \in M(X)$ be a computable probability measure. Then there exists a collection of Borel sets $\mathcal{A} = \{A_1, A_2, ...\}$ with the following properties:*

- *$\mathcal{A}$ is a generating algebra.*
- *the union, intersection and complementation are* constructive *operations,*
- *the map $i \mapsto \mu(A_i)$ is computable.*

This theorem will proof very useful since it is the main tool in many constructions. The notion of ***computable probability space*** (a computable metric space with a computable probability measure) is then introduced as well as the corresponding notion of ***isomorphism*** between these objects, namely, a measure preserving bijection which is moreover computable on a *constructive* G_δ set of measure one. We obtain a first classification, given by:

Theorem *Let (X, μ) be a computable probability space. Then:*

- *there exists a computable measure ν over the space $\{0,1\}^{\mathbb{N}}$ such that (X, μ) and $(\{0,1\}^{\mathbb{N}}, \nu)$ are isomorphic.*
- *if μ is non-atomic, then (X, μ) is isomorphic to $([0,1], \lambda)$ where λ is the Lebesgue measure.*

In particular, this result allows to directly transfer many tools developed in Cantor space to any computable probability space.

2.2 Computability on the space $\mathcal{M}(X)$ of probability measures

2.2.1 $\mathcal{M}(X)$ as a computable metric space

Given a metric space (X, d), the set $\mathcal{M}(X)$ of Borel probability measures over X can be endowed with the *weak topology*, which is the finest topology for which

$$\mu_n \longrightarrow \mu \text{ if and only if } \int f d\mu_n \longrightarrow \int f d\mu$$

for all continuous bounded functions $f : X \to \mathbb{R}$.

This is a metrizable topology and, when X is separable and complete, $\mathcal{M}(X)$ is also separable and complete (see [Bil68]). Moreover, a computable metric structure on X induces in a canonical way a computable metric structure on $\mathcal{M}(X)$.

Let $\mathcal{D} \subset \mathcal{M}(X)$ be the set of those probability measures that are concentrated in finitely many points of $\mathcal{S}$ and assign rational values to them. It can be shown that this is a dense subset ([Bil68]). The numberings $\nu_{\mathcal{S}}$ of ideal points of X and $\nu_{\mathbb{Q}}$ of the rationals numbers induce a numbering $\nu_{\mathcal{D}}$ of ideal measures: $\mu_{\langle\langle n_1,..,n_k\rangle,\langle m_1,..,m_k\rangle\rangle}$ is the measure concentrated over the finite set $\{s_{n_1}, .., s_{n_k}\}$ where q_{m_i} is the weight of s_{n_i}.

The Prokhorov metric

Let us consider the particular metric on $\mathcal{M}(X)$:

Definition 2.2.1.1. The ***Prokhorov metric*** ρ on $\mathcal{M}(X)$ is defined by:

$$\rho(\mu,\nu) := \inf\{\epsilon \in \mathbb{R}^+ : \mu(A) \leq \nu(A^\epsilon) + \epsilon \text{ for every Borel set } A\}. \tag{2.1}$$

where $A^\epsilon = \{x : d(x,A) < \epsilon\}$. ⌟

It is known that it is indeed a metric, which induces the weak topology on $\mathcal{M}(X)$ (see [Bil68]). The following theorem will prove useful later:

Theorem 2.2.1.1. *Let $g\colon X \to \mathbb{R}$ be a β-Lipschitz function with support included in $B = B(x,r)$. If μ and ν in $\mathcal{M}(X)$ are such that $\rho(\mu,\nu) < \delta$ then $|\mu g - \nu g| \leq \delta\beta(r+1)$*

Proof. By the coupling theorem A.1.0.3, there exist a measure P over the product space $X \times X$ with marginals μ and ν such that $P[d(x,y) > \delta] \leq \delta$. Then we have that:

$$\begin{aligned}
\Big| \int_X g(x)d\mu - \int_X g(y)d\nu \Big| &\leq \int_{B\times B} |\, g(x) - g(y) \,|\, dP(x,y) \\
&\leq \beta \int_{B\times B} d(x,y)dP(x,y).
\end{aligned}$$

putting $D = \{d(x,y) > \delta\}$ we have that this quantity is equal to

$$\begin{aligned}
\beta\Big(\int_{B\times B\cap D} d(x,y)dP(x,y) &+ \int_{B\times B\cap D^c} d(x,y)dP(x,y)\Big) \\
&\leq \beta r\delta + \beta\delta = \delta(\beta(r+1)).
\end{aligned}$$

□

Moreover, we have that:

Proposition 2.2.1.1. *$(\mathcal{M}(X), \mathcal{D}, \rho)$ is a computable metric space.*

Proof. We have to show that the real numbers $\rho(\mu_i, \mu_j)$ are all computable, uniformly in $\langle i, j\rangle$. First observe that if U is a constructively open subset of X, $\mu_i(U)$ is lower semi-computable uniformly in i and U. Indeed, if $(s_{n_1}, q_{m_1}), \ldots, (s_{n_k}, q_{m_k})$ are the mass points of μ_i together with their weights (computable from i) then $\mu_i(U) = \sum_{s_{n_j} \in U} q_{m_j}$. As the s_{n_j} which belong to U can be enumerated from any description of U, this sum is lower-semi-computable. In particular, $\mu_i(B_{i_1} \cup \ldots \cup B_{i_k})$ is lower semi-computable and $\mu_i(\overline{B}_{i_1} \cup \ldots \cup \overline{B}_{i_k})$ is upper semi-computable, both of them uniformly in $\langle i, i_1, \ldots, i_k \rangle$. Now we prove that $\rho(\mu_i, \mu_j)$ is computable uniformly in $\langle i, j \rangle$.

Observe that if μ_i is an ideal measure concentrated over S_i, then (2.1) becomes $\rho(\mu_i, \mu_j) = \inf\{\epsilon \in \mathbb{Q} : \forall A \subset S_i,\ \mu_i(A) < \mu_j(A^\epsilon) + \epsilon\}$. Since μ_j is also an ideal measure and A^ϵ is a finite union of open ideal balls, the number $\mu_j(A^\epsilon)$ is lower semi-computable (uniformly) and then $\rho(\mu_i, \mu_j)$ is upper semi-computable, uniformly in $\langle i, j \rangle$. To see that $\rho(\mu_i, \mu_j)$ is lower-semicomputable, uniformly in $\langle i, j \rangle$, observe that $\rho(\mu_i, \mu_j) = \sup\{\epsilon \in \mathbb{Q} : \exists A \subset S_i,\ \mu_i(A) > \mu_j(A^{\overline{\epsilon}}) + \epsilon\}$, where $A^{\overline{\epsilon}} = \{x : d(x, A) \leq \epsilon\}$ (a finite union of closed ideal balls when $A \subset S_i$) and use the upper semi-computability of $\mu_j(A^{\overline{\epsilon}})$. $\square$

Definition 2.2.1.2. A measure μ is ***computable*** if it is a constructive point of $(\mathcal{M}(X), \mathcal{D}, \rho)$. $\lrcorner$

The effectivization of the space of Borel probability measures $\mathcal{M}(X)$ is of theoretical interest, and opens the question: what kind of information can be (algorithmically) recovered from a description of a measure as a point of the computable metric space $\mathcal{M}(X)$? Which we shall study in the next section.

The Wasserstein metric

In the particular case when the metric space X is bounded, an alternative metric can be defined on $\mathcal{M}(X)$. When f is a real-valued function, μf denotes $\int f d\mu$.

Definition 2.2.1.3. The *Wasserstein metric* on $\mathcal{M}(X)$ is defined by:

$$W(\mu, \nu) = \sup_{f \in 1-Lip(X)} (|\mu f - \nu f|) \tag{2.2}$$

where $1 - Lip(X)$ is the space of 1-Lipschitz functions from X to $\mathbb{R}$. $\lrcorner$

We recall (see [AGS05]) that W has the following properties:

Proposition 2.2.1.2.

1. *W is a distance and if X is separable and complete then $\mathcal{M}(X)$ with this distance is a separable and complete metric space.*

2. *The topology induced by W is the weak topology and thus W is equivalent to the Prokhorov metric.*

Moreover, if $(X, \mathcal{S}, d)$ is a computable metric space, then:

Proposition 2.2.1.3. *$(\mathcal{M}(X), \mathcal{D}, W)$ is a computable metric space.*

Proof. We have to show that the distance $W(\mu_i, \mu_j)$ between ideal measures is uniformly computable. From $\langle i, j \rangle$ we can compute the set $S_{i,j} = supp(\mu_i) \cup supp(\mu_j)$. Let $s_0 \in S_{i,j}$, then we can suppose that the supremum in (2.2) is taken over $1 - Lip^0_{s_0}(X) := \{f \in 1 - Lip(X) : 1 - Lip^0_{s_0}(X) f(s) = 0\}$. Given some precision ϵ we construct a finite set $\mathcal{N}_\epsilon \subset 1 - Lip^0_{s_0}(X)$ made of uniformly computable functions such that for each $f \in 1 - Lip^0_{s_0}(X)$ there is some $l \in \mathcal{N}_\epsilon$ satisfying $\sup\{|f(x) - l(x)| : x \in S_{i,j}\} < \epsilon$: compute an integer m such that $S_{i,j} \subset B(s, m)$; then $|f| < m$ for every $f \in 1 - Lip^0_s(X)$. Let n be such that $m/n < 2\epsilon$. For each $s \in S_{i,j}$ and $a \in \{\frac{lm}{n}\}^m_{l=-m}$ let us consider the functions defined by $\phi^+_{s,l}(x) := a + d(s, x)$ and $\phi^-_{s,l}(x) := a - d(s, x)$. Then it is not difficult to see that $\mathcal{N}_\epsilon$ defined as the set of all possible combinations of max and min made with the $\phi^{+-}_{s,l}(x)$ satisfy the required condition.

Therefore, since $\sup(|f - g|) < \epsilon$ implies $|\mu(f - g)| < \epsilon$ we have that:

$$W(\mu_i, \mu_j) \in [\sup_{g \in \mathcal{N}_\epsilon} (|\mu_i g - \mu_j g|), \sup_{g \in \mathcal{N}_\epsilon} (|\mu_i g - \mu_j g|) + 2\epsilon]$$

where the $\mu_i g$ are computable, uniformly in i. The result follows.

□

When X is bounded, the effectivization using the Prokhorov or the Wasserstein metrics turn out to be equivalent.

Theorem 2.2.1.2. *The Prokhorov and the Wasserstein metrics are computably equivalent. That is, the identity function $id : (\mathcal{M}(X), \mathcal{D}, \rho) \to (\mathcal{M}(X), \mathcal{D}, W)$ is a computable isomorphism, as well as its inverse.*

Proof. Let M be an integer such that $\sup_{x,y \in X} d(x, y) < M$. Suppose $\rho(\mu, \nu) < \epsilon/(M + 1)$. Then, by theorem 2.2.1.1, for every $f \in 1 - Lip(X)$ it holds $|\mu f - \nu f| \leq \epsilon$, then $W(\mu, \nu) < \epsilon$. Conversely, suppose $W(\mu, \nu) < \epsilon^2 < 1$. Let A be a Borel set and define $g^A_\epsilon := |1 - d(x, A)/\epsilon|^+$. Then $\epsilon g^A_\epsilon \in 1 - Lip(X)$. $W(\mu, \nu) < \epsilon^2$ implies $\mu \epsilon g^A_\epsilon < \nu \epsilon g^A_\epsilon + \epsilon^2$ and since $\mu(A) \leq \mu g^A_\epsilon$ and $\nu g^A_\epsilon \leq \nu(A^\epsilon)$, we conclude $\mu(A) \leq \nu(A^\epsilon) + \epsilon$ and then $\rho(\mu, \nu) < \epsilon$.

Therefore, given a fast sequence of ideal measures converging to μ in the Prokhorov metric, we can construct a fast sequence of ideal measures converging to μ in the W metric and vice-versa.

$\square$

2.2.2 Computing with probability measures

The purpose of this section is to study the computability of the integral of functions or the measure of sets, given a description of the measure. These quantities are in general not computable. The classical example is to consider the sequence of measures $\delta_{1/n}$ which converges to the computable measure δ_0. Let $1_{(0,1)}$ be the indicator function of the open interval. For all n, we have that $\int 1_{(0,1)}(x)d\delta_{1/n} = 1$ and $\int 1_{(0,1)}d\delta_0 = 0$. So the function $\mu \mapsto \int 1_{(0,1)}d\mu$ is in general not computable since it is not even continuous. Instead, we will show that it is lower-semicomputable.

Given an integrable function $f \in L^1$, we denote by $L_f : M(X) \to \mathbb{R}$ the function mapping μ into μf. Let g_i denotes the functions from the simple set $\mathcal{H}$ introduced in section 1.7.3. The regularity of these functions allows to show:

Theorem 2.2.2.1. *The functions L_{g_i}: $\mathcal{M}(X) \to \mathbb{R}$ are computable uniformly in i.*

Proof. We use corollary 1.6.2.1. Let $g_i \in \mathcal{H}$. From i we can compute a Lipschitz constant β and a bound r for the diameter of its support (we suppose $c = 0$ in remark 1.7.3.1). Let $I_{k,j} = (q_k, q_j)$ be an enumeration of the rational intervals. For each ideal measure μ_n we can semi-decide $\mu_n g \in I_{k,j}$ (since $\mu_n g_i$ is computable). Now compute some ϵ such that $(\mu_n g_i - \epsilon, \mu_n g_i + \epsilon) \subset I_{k,j}$ and take $\delta = \epsilon(\beta(r+1))^{-1}$. By theorem 2.2.1.1 we get $B(\mu_n, \delta) \subset L_{g_i}^{-1}(I_{k,j})$. $\square$

To be able to talk about integrals of more general functions, as the indicator functions of open sets, we will consider the functions spaces $C(X, \overline{\mathbb{R}^+}), C(M(X), \overline{\mathbb{R}^+})$ and $C(M(X), [0,1])$ (with its enumerative lattice structure) and the following two operators:

(i) the ***integral operator***,

$$L : C(X, \overline{\mathbb{R}^+}) \to C(M(X), \overline{\mathbb{R}^+})$$
$$f \mapsto L_f$$

and, (ii) the ***valuation operator***,

$$v : \tau_X \to C(M(X), [0,1])$$
$$U \mapsto L_{1_U}$$

We can now state:

Theorem 2.2.2.2. *The integral and the valuation operators are constructive.*

Proof. It is obvious that the integral operator is monotone. The fact that it commutes with suprema of increasing sequences follows from the monotone convergence theorem (A.1.0.4). Its constructivity is immediate from theorem 2.2.2.1 since the set $\mathcal{H}^+$ of simple functions is closed under suprema of finite collections. To see that the valuation operator is constructive, observe that given a finite union of ideal balls $B^k = \cup_{j \leq k} B_{i_j}$ we have that $1_{B^k} \in C(X, \overline{\mathbb{R}^+})$ is constructive uniformly in $\langle i_1, .., i_k \rangle$, and as the integral operator is constructive, so is $L_{1_B^k}$. □

So, we know how to (lower) semi-compute the integral of constructive functions. It is natural to ask whether one can also semi-compute the integral of functions which are constructive only on some subset $D \subset X$. The following proposition shows that this is the case at least when the domain of computability has full measure.

Proposition 2.2.2.1. *Let $D \subset X$ be a Borel set. If $f_i : X \to \mathbb{R}^+$ is a uniform sequence of functions which are lower semi-computable on D, then L_{f_i} is lower semi-computable on $M(D) := \{\mu : \mu(X \setminus D) = 0\}$, uniformly in i.*

Proof. By Thm. 1.7.1.1, from each i one can construct a lower semi-computable function $\hat{f}_i$ satisfying $\hat{f}_i = f_i$ on D. Since the function $\mu \mapsto \int_X \hat{f}_i d\mu$ is lower semi-computable, uniformly in i and $\mu(X \setminus D) = 0$, we have that on $M(D)$ it coincides with $\mu \mapsto \int f_i d\mu$, which is then lower semi-computable on $M(D)$, uniformly in i. □

For computable functions we have:

Corollary 2.2.2.1. *Let $D \subset X$ be measurable. Let $f_i : X \to \mathbb{R}$ be such that $|f_i| \leq M_i$ with M_i computable from i. If the f_i are computable on D uniformly in i, then $L_{f_i} : \mathcal{M}(X) \to \mathbb{R}$ is computable on $M(D) := \{\mu : \mu(X \setminus D) = 0\}$, uniformly in i.*

Proof. $L_{f_i+M_i} = L_{f_i} + M_i$ is bounded by $2M_i$. Since $f_i + M_i \in C(X, \overline{\mathbb{R}^+})$ is constructive on D uniformly in i, proposition 2.2.2.1 implies that on $M(D)$, so is $L_{f_i+M_i}$. The same applies to $-f_i+M_i$. We obtain that $L_{-f_i+M_i}$ is constructive on $M(D)$. Since $-L_{f_i}+M_i = L_{-f_i+M_i}$, the result follows from corollary 1.7.3.1. □

The above results can be stated in a more general fashion (we only state the version for total functions):

Corollary 2.2.2.2. *Let Y be a represented space and $f_i : Y \times X \to \overline{\mathbb{R}^+}$ be a constructive function, uniformly in i. Then $L_{f_i} : Y \to C(M(X), \overline{\mathbb{R}^+})$ mapping $y \mapsto L_{f_i(y,\cdot)}$ is constructive, uniformly in i.*

Proof. It is a composition of constructive functions: $L_{f_i} = L \circ [Curry(f_i)]$. □

Corollary 2.2.2.3. *Let $f_i : M(X) \times X \to \mathbb{R}$ be such that $|f_i| \leq M_i$. If f_i and M_i are computable uniformly in i, then $L_{f_i} : \mathcal{M}(X) \to \mathbb{R}$ mapping $\mu \mapsto L_{f_i(\mu,\cdot)}(\mu)$ is computable uniformly in i.*

Proof. Apply corollary 2.2.2.2 with $Y = M(X)$ to $f_i + M_i$ and to $-f_i + M_i$ as in the proof of corollary 2.2.2.1. □

Now we characterize computable measures:

Theorem 2.2.2.3. *The following are equivalent:*

(i) Measure μ is computable.

(ii) μg_i is computable uniformly in i.

(iii) The measure $\mu(B = \cup_{j=1}^{k} B_{i_j})$ of finite unions of ideal open balls is lower semi-computable uniformly in $\langle i_1, \ldots, i_k \rangle$.

Proof. [$(i) \Rightarrow (ii)$] It follows from theorem 2.2.2.1.
[$(ii) \Rightarrow (iii)$] It follows from theorem 2.2.2.2.
[$(iii) \Rightarrow (i)$] Given an ideal ball $B(\mu_n, \epsilon)$, since $\rho(\mu_n, \mu) < \epsilon$ iff $\mu_n(A) < \mu(A^\epsilon) + \epsilon$ for all $A \subset S_n$ (where S_n is the finite support of μ_n) and $\mu(A^\epsilon)$ is lower semi-computable (A^ϵ is a finite union of open ideal balls) we can semi-decide $\mu \in B(\mu_n, \epsilon)$. □

Of course, we also have that $\mu \in M(X)$ is computable iff $\mu(U_e)$ is lower semi-computable, uniformly in e (where (U_e) is a recursive enumeration of the constructively open sets). This directly gives a criteria to check the computability of an operator $L : M(X) \to M(X)$.

Corollary 2.2.2.4. *Let X be a computable metric space and $\mathcal{D}$ be a subset of $M(X)$. A transformation $L : M(X) \to M(X)$ is computable on $\mathcal{D}$ iff the function $\mu \mapsto L(\mu)(U_e)$ is lower semi-computable on $\mathcal{D}$, uniformly in e.*

Symbolic spaces and the unit interval as special cases

Examples 2.2.2.1.

1. On a symbolic space $\Sigma^{\mathbb{N}}$ (where Σ is a finite alphabet) with its natural computable metric space structure, the ideal balls are the cylinders. Any finite union of cylinders can always be expressed as a disjoint (and finite) union of cylinders, and the complement of a cylinder is a finite union of cylinders. Thus we have:

   Proposition 2.2.2.2. *A measure $\mu \in \mathcal{M}(\Sigma^{\mathbb{N}})$ is computable iff so is the function $\mu : \Sigma^* \to \mathbb{R}^+$ mapping w to $\mu([w])$.*

2. On the unit real interval, ideals balls are open rational intervals. Again, a finite union of such intervals can always be expressed as a disjoint (and finite) union of open rational intervals. Then:

   Proposition 2.2.2.3. *A measure $\mu \in \mathcal{M}([0,1])$ is computable iff the measures of the rational open intervals are uniformly lower semi-computable.*

   If μ has no atoms, a rational open interval is the complement of at most two disjoint open rational intervals, up to a null set. In this case, the measure μ is computable iff the measures of the rational intervals are uniformly *computable*.

$\lrcorner$

2.3 Computable probability spaces

In this section we adopt a measure theoretical view to study in more detail computable metric spaces equipped with a fixed computable measure. Let us then introduce:

Definition 2.3.0.1. A ***computable probability space (CPS)*** is a pair $(\mathcal{X}, \mu)$ where $\mathcal{X}$ is a computable metric space and μ is a computable Borel probability measure on X. ⌟

As already said, a computable function defined on the whole space is necessarily continuous. But a transformation or an observable need not be continuous at every point, as many interesting examples prove (piecewise-defined transformations, characteristic functions of measurable sets,...), so the requirement of being computable everywhere is too strong. In the (computable) measure-theoretical setting, a natural weaker condition is to require the function to be computable almost everywhere, but this seems to be too weak for our purposes. An intermediate definition which makes things work is the following:

Definition 2.3.0.2. A set $D \subset X$ is called a ***constructive*** $\boldsymbol{G_\delta}$***-set*** if it is the intersection of uniformly constructively open sets. ⌟

Definition 2.3.0.3. Let $(\mathcal{X}, \mu)$ and $\mathcal{Y}$ be a computable probability space and a computable metric space respectively. A function $f : (\mathcal{X}, \mu) \to Y$ is μ***-almost computable*** if it is computable on a constructive G_δ-set (denoted as $\mathrm{dom} f$ or D_f) of measure one. ⌟

Remark 2.3.0.1. Given a uniform sequence of μ-almost computable functions $(f_i)_i$, any computable operation $\odot_{i=0}^{n} f_i$ (addition, multiplication, composition, etc...) is μ-almost computable, uniformly in n. ⌟

We recall that $F : (\mathcal{X}, \mu) \to (\mathcal{Y}, \nu)$ is measure-preserving if $\mu(F^{-1}(A)) = \nu(A)$ for all Borel sets A. To classify computable probability spaces, we need to define their morphisms and isomorphisms.

Definition 2.3.0.4. A ***morphism of computable probability spaces*** $F : (\mathcal{X}, \mu) \to (\mathcal{Y}, \nu)$, is a μ-almost computable measure-preserving function $F : D_F \subseteq X \to Y$.
An ***isomorphism*** $(F, G) : (\mathcal{X}, \mu) \rightleftarrows (\mathcal{Y}, \nu)$ is a pair (F, G) of morphisms such that $G \circ F = \mathrm{id}$ on $F^{-1}(D_G)$ and $F \circ G = \mathrm{id}$ on $G^{-1}(D_F)$. ⌟

Remark 2.3.0.2. To every isomorphism (F, G) one can associate the canonical invertible morphism $\varphi = F|_{D_\varphi}$ with $\varphi^{-1} = G|_{D_{\varphi^{-1}}}$, where $D_\varphi = F^{-1}(G^{-1}(D_F))$ and $D_{\varphi^{-1}} = G^{-1}(D_F)$. Of course, (φ, φ^{-1}) is an isomorphism. ⌟

Computable probability structures can be easily transferred:

Proposition 2.3.0.4. *Let $(\mathcal{X}, \mu)$ be a computable probability space, $\mathcal{Y}$ be a computable metric space and $F : X \to Y$ a function which is computable on a constructive G_δ-set of*

μ-measure one. Then the induced measure μ_F on Y defined by $\mu_F(A) = \mu(F^{-1}(A))$ is computable and F is a morphism of computable probability spaces.

2.3.1 Borel sets of computable measure.

The measure of most sets is not computable. The measure of constructively open sets for example, is in general only lower semi-computable. Nevertheless, sets whose measure is computable play an important role. In order to handle these sets we will use a collection of more elementary sets (of computable measure) which have interesting algebraic properties. They are given by the following theorem:

Theorem 2.3.1.1. *Let (X, μ) be a computable probability space. Then there exists a collection of* constructive *Borel sets $\mathscr{A} = \{A_1, A_2, ...\}$ with the following properties:*

- *$\mathscr{A}$ is a generating algebra.*
- *the union, intersection and complementation are* constructive *operations,*
- *the map $i \mapsto \mu(A_i)$ is computable.*

This theorem will be our main tool in many later constructions. To prove it, we need some preparation.

Definition 2.3.1.1. A measurable set A is said to be a ***μ-continuity set*** if $\mu(\partial A) = 0$ where $\partial A = \overline{A} \cap \overline{X \setminus A}$ is the boundary of A. ⌟

Definition 2.3.1.2. A set A is said to be ***almost decidable*** if there is a constructive G_δ set D of measure one and two constructively open sets U and V such that:

$$U \cap D \subset A, \quad V \cap D \subseteq A^{\mathcal{C}}, \quad \mu(U) + \mu(V) = 1$$

⌟

Remarks 2.3.1.1.

1. The collection of almost decidable sets is an algebra.
2. An almost decidable set is always a continuity set.
3. A μ-continuity ideal ball is always almost decidable (with $D = X$).

4. Unless the space is disconnected (i.e. has non-trivial clopen subsets), no set can be *decidable*, i.e. semi-decidable (constructively open) and with a semi-decidable complement (such a set must be clopen[1]). Instead, a set can be decidable *with probability* 1: there is an algorithm which decides if a point belongs to the set or not, for almost every point. This is why we call it *almost decidable*.

⌟

Ignoring computability, the existence of open μ-continuity sets directly follows from the fact that the collection of open sets is uncountable and μ is finite. The problem in the computable setting is that there are only countable many constructively open sets.

The following will be an important tool.

Theorem 2.3.1.2. *Let $(\mathcal{X}, \mu)$ be a CPS and $(f_i)_i$ be a sequence of uniformly computable real valued functions on X. Then there is a sequence of uniformly computable reals $(x_n)_n$ which is dense in $\mathbb{R}$ and such that $\mu(\{f_i^{-1}(x_n)\}) = 0$ for all i, n.*

The proof uses the following lemma:

Lemma 2.3.1.1. *Let X be $\mathbb{R}$ or $\mathbb{R}^+$ or $[0,1]$. Let μ be a computable probability measure on X. Then there is a sequence of uniformly computable reals $(x_n)_n$ which is dense in X and such that $\mu(\{x_n\}) = 0$ for all n.*

Proof. Let I be a closed rational interval. We construct $x \in I$ such that $\mu(\{x\}) = 0$. To do this, we construct inductively a nested sequence of closed intervals J_k of measure $< 2^{-k+1}$, with $J_0 = I$. Suppose $J_k = [a, b]$ has been constructed, with $\mu(J_k) < 2^{-k+1}$. Let $m = (b-a)/3$: one of the intervals $[a, a+m]$ and $[b-m, b]$ must have measure $< 2^{-k}$, and we can find it effectively—let it be J_{k+1}. From a constructive enumeration $(I_n)_n$ of all the dyadic intervals, we can construct $x_n \in I_n$ uniformly. □

Proof of theorem 2.3.1.2. Consider the uniformly computable measures $\mu_i = \mu \circ f_i^{-1}$ and define $\nu = \sum_i 2^{-i} \mu_i$. By theorem 2.2.2.3, ν is a computable measure and then, by Lemma 2.3.1.1, there is a sequence of uniformly computable reals $(x_n)_n$ which is dense in $\mathbb{R}$ and such that $\nu(\{x_n\}) = 0$ for all n. Since $\nu(A) = 0$ iff $\mu_i(A) = 0$ for all i, we get $\mu(\{f_i^{-1}(x_n)\}) = 0$ for all i, n. □

[1]In the Cantor space for example (which is totally disconnected), every cylinder (ball) is a decidable set. Indeed, to decide if some infinite sequence belongs to some cylinder it suffices to compare the finite word defining the cylinder to the corresponding finite prefix of the infinite sequence.

The following result will be used many times in the sequel.

Corollary 2.3.1.1. *There is a sequence of uniformly computable reals $(r_n)_{n\in\mathbb{N}}$ such that $(B(s_i, r_n))_{i,n}$ is a basis of almost decidable balls.*

Proof. Apply Theorem 2.3.1.2 to $(f_i)_i$ defined by $f_i(x) = d(s_i, x)$. □

Note that different algorithmic descriptions of the same μ may yield different sequences $(r_n)_{n\in\mathbb{N}}$. It is understood that some algorithmic description of μ has been chosen and fixed. This can be done only because the measure μ is computable, which is then a crucial hypothesis. Moreover, it can be shown that there exists (non-computable) measures for which there is no almost decidable set. For a counterexample we refer to [Hoy08]. We remark that every ideal ball can be expressed as a recursively enumerable union of almost decidable balls, and vice-versa. So the two bases are constructively equivalent.

We are able to prove theorem 2.3.1.1.

Proof of theorem 2.3.1.1. Define $\mathscr{A}$ to be the algebra generated by the collection of almost decidable balls constructed above. As almost decidable balls are a basis, they generate $\mathscr{B}$. Then, so does $\mathscr{A}$. It is easy to see that elements of $\mathscr{A}$ are almost decidable. Moreover, the numbering of the collection of almost decidable balls induce a numbering of $\mathscr{A}$ such that A_n is almost decidable uniformly in n. To prove that $\mu(A_n)$ is computable uniformly in n, let U_n and V_n be the sets making A_n almost decidable. We have $\mu(A_n) = \mu(U_n) = \mu(V_n)$. As both, U_n and V_n are constructively open, their measures are lower semi-computable. Since $\mu(U_n) + \mu(V_n) = 1$, their measures are also upper semi-computable. The result follows. □

Constructive open sets with computable measure

Let us now introduce a class of sets with computable measure, which are not necessarily almost decidable,

Definition 2.3.1.3. A set A is called ***computably open*** if it is constructively open and $\mu(A)$ is computable. ⌟

Lemma 2.3.1.2. *If V is computably open, then there is a recursives equence (A_n) in $\mathscr{A}$ satisfying $A_n \nearrow V$ and $\mu(V \triangle A_n) \leq 2^{-n}$.*

Proof. As V is constructively open, there is a recursive sequence of almost decidable balls (B_i) such that $V = \cup_i B_i$. Let $B^k = \cup_{i \leq k} B_i$. As $\mu(V)$ is computable, given any $\epsilon > 0$, a k can be computed such that $\mu(V \bigtriangleup B^k) = \mu(V) - \mu(B^k) < \epsilon$. □

Corollary 2.3.1.2. *Let V_i and V_j be computably open. Then $V_i \cup V_j$ and $V_i \cap V_j$ are also computably open, uniformly in i, j.*

Proof. Let $A^i_n \nearrow V_i$ and $A^j_n \nearrow V_j$ be as in the above lemma. It is easy to see that $\mu(V_i \cup V_j) - \mu(A^i_{n+1} \cup A^j_{n+1}) \leq 2^{-n}$ and that $\mu(V_i \cap V_j) - \mu(A^i_{n+1} \cap A^j_{n+1}) \leq 2^{-n}$. □

Given a subset A, it is usual to consider the ***normalized*** measure μ_A, defined by $\mu_A(B) = \frac{\mu(B \cap A)}{\mu(A)}$ for any Borel set B. The computability of this measure is an important question. The following proposition gives a positive answer when the set A is good enough.

Proposition 2.3.1.1. *Let μ be a computable measure. Let A be an almost decidable subset of X or a computably open set. Then the induced measures μ_A and $\mu_{A^{\mathcal{C}}}$ are computable.*

Proof. let $W = B_{n_1} \cup \ldots \cup B_{n_k}$ be a finite union of ideal balls. Suppose A is almost decidable. Then there exists a constructively open set $U = A(\text{mod } 0)$ (hence $\mu(A) = \mu(U)$). If A is computably open, then put $U = A$ and the same holds. We have that

$$\mu_A(W) = \frac{\mu(W \cap A)}{\mu(A)} = \frac{\mu(W \cap U)}{\mu(A)}$$

and that

$$\mu_{A^{\mathcal{C}}}(W) = \frac{\mu(W \cap A^{\mathcal{C}})}{\mu(A^{\mathcal{C}})} = \frac{\mu(W \cap U^{\mathcal{C}})}{\mu(A^{\mathcal{C}})}.$$

$W \cap U$ and $W \cup U$ are constructively open sets, so their measure is lower semi-computable. $\mu(W \cap U^{\mathcal{C}}) = \mu(W \cup U) - \mu(U)$, so $\mu(W \cap U^{\mathcal{C}})$ is lower semi-computable too. As $\mu(A)$ and $\mu(A^{\mathcal{C}}) = 1 - \mu(A)$ are computable, $\mu_A(W)$ and $\mu_{A^{\mathcal{C}}}(W)$ are lower semi-computable. Note that everything is uniform in $\langle n_1, \ldots, n_k \rangle$. The result follows from theorem 2.2.2.3. □

Recursively measurable functions

Let (X, μ) and (Y, ν) be computable probability spaces. The following is a natural definition:

Definition 2.3.1.4. A function $f : X \to Y$ is said to be ***recursively measurable*** provided that $f^{-1}(A_n)$ is almost decidable uniformly in n for every almost decidable set $A_n \in \mathscr{A}_Y$. ⌟

The following facts are proved exactly as in the classical setting.

Proposition 2.3.1.2. *Let $f : X \to Y$ be a function.*

- *If $f^{-1}(B_i)$ is (uniformly) almost decidable for every almost decidable ball $B_i \subset Y$, then f is recursively measurable.*
- *If $Y = \overline{\mathbb{R}}$ and there is a* dense *sequence (r_n) of uniformly computable reals such that $\nu(\{r_n\}) = 0$ and $f^{-1}((r_n, \infty])$ is (uniformly) almost decidable, then f is recursively measurable.*

Recursive measurability is a stronger property that μ-almost computability, as the following proposition shows. We recall that a measurable function $f : X \to Y$ is said to be ***non singular*** if for every set $A \subset Y$ of measure zero, the preimage of A, $f^{-1}(A)$ has measure zero too.

Proposition 2.3.1.3. *Any recursively measurable function $f : X \to Y$ is in particular μ-almost computable. If f is non singular and μ-almost computable then it is recursively measurable.*

Proof. Suppose f is recursively measurable. Let (B_i^Y) the class of almost decidable balls over Y. Then $f^{-1}(B_i^Y)$ is almost decidable. Denote by U_i and V_i the associated constructively open sets. We have $U_i \subset f^{-1}(B_i^Y)$ and $V_i \subset f^{-1}({B_i^Y}^{\mathcal{C}}) = (f^{-1}B_i^Y)^{\mathcal{C}}$. Define the constructive G_δ set $D_f = \cap_i(V_i \cup U_i)$. Clearly, $f^{-1}(B_i^Y) \cap D_f = U_i \cap D_f$. f is then computable on D_f. Conversely, suppose f is computable on the constructive G_δ set D_f and non singular. Then $f^{-1}B_i^Y = U_i \cap D_f$ with U_i uniformly constructively open and $f^{-1}(\overline{B}_i^Y)^{\mathcal{C}} = V_i \cap D_f$ with V_i uniformly constructively open. Since f is non singular, $f^{-1}B_i^Y$ is (uniformly) almost decidable. $\square$

Another characterization of the computability of measures

The existence of a basis of almost decidable sets also leads to another characterization of the computability of measures, which is reminiscent of what happens on the Cantor space (see corollary 2.2.2.2). Let us say that two bases $(U_i)_i$ and $(V_i)_i$ of the topology τ are constructively equivalent if both $id_\tau : (\tau, \subseteq, \mathcal{U}) \to (\tau, \subseteq, \mathcal{V})$ and its inverse are constructive functions between enumerative lattices.

Corollary 2.3.1.3. *A measure $\mu \in \mathcal{M}(X)$ is computable if and only if there is a basis $\mathcal{U} = (U_i)_{i\in\mathbb{N}}$ of uniformly almost decidable open sets which is constructively equivalent to $\mathcal{B}$ and such that all $\mu(U_{i_1} \cup \ldots \cup U_{i_k})$ are computable uniformly in $\langle i_1, \ldots, i_k\rangle$.*

Proof. If μ is computable, the almost decidable balls $U_{\langle i,n\rangle} = B(s_i, r_n)$ are a basis which is constructively equivalent to $\mathcal{B}$: indeed, $B(s_i, r_n) = \bigcup_{q_j<r_n} B(s_i, q_j)$ and $B(s_i, q_j) = \bigcup_{r_n<q_j} B(s_i, r_n)$, and r_n is computable uniformly in n. For the converse note that a finite union of ideal balls, being a constructively open set relative to the basis $\mathcal{U}$, have a lower semi-computable measure. The result follows from theorem 2.2.2.3. □

Generalized binary representations

The Cantor space $\{0,1\}^{\mathbb{N}}$ is a privileged place for computability. This can be understood by the fact that it is the countable product (with the product topology) of a finite space (with the discrete topology). A consequence of this is that membership of a basic open set (cylinder) boils down to a pattern-matching and is then *decidable*. As decidable sets must be clopen, this property cannot hold in connected spaces. As a result, a computable metric space is not in general constructively homeomorphic to the Cantor space.

Nevertheless, the real unit interval $[0,1]$ is not so far away from the Cantor space. The binary numeral system provides a correspondence between real numbers and binary sequences, which is certainly not homeomorphic, unless we remove the small set of dyadic numbers. In particular, the remaining set is totally disconnected, and the dyadic intervals form a basis of clopen sets.

Actually, this correspondence makes the computable probability space $[0,1]$ with the Lebesgue measure isomorphic to the Cantor space with the uniform measure. This fact has been implicitly used, for instance, to extend algorithmic randomness from the Cantor space with the uniform measure to the unit interval with the Lebesgue measure.

We extend this to any computable probability space defining the notion of *binary representation*, and show that every computable probability space has a binary representation, which implies in particular that every computable probability space is isomorphic to the Cantor space with a computable measure. To carry out this generalization, let us briefly scrutinize the binary numeral system on the unit interval:

$\delta : \{0,1\}^{\mathbb{N}} \to [0,1]$ is a total surjective morphism. Every non-dyadic real has a unique expansion, and the inverse of δ, defined on the set D of non-dyadic numbers, is computable.

Moreover, D is large both in a topological and measure-theoretical sense: it is a residual (a countable intersection of dense open sets) and has measure one. (δ, δ^{-1}) is then an isomorphism.

In our generalization, we do not require every binary sequence to be the expansion of a point, which would force X to be compact.

Definition 2.3.1.5. A ***binary representation*** of a computable probability space $(\mathcal{X}, \mu)$ is a pair (δ, μ_δ) where μ_δ is a computable probability measure on $\{0,1\}^{\mathbb{N}}$ and $\delta : (\{0,1\}^{\mathbb{N}}, \mu_\delta) \to (\mathcal{X}, \mu)$ is a surjective morphism such that, calling $\delta^{-1}(x)$ the set of ***expansions*** of $x \in X$:

(i) there is a constructive dense full-measure G_δ-set D of points having a unique expansion and,

(ii) $\delta^{-1} : D \to \delta^{-1}(D)$ is computable.

⌟

We remark that when the support of the measure (the smallest closed set of full measure) is the whole space X, like the Lebesgue measure on the interval, a full-measure G_δ-set is always dense but, in general, it is only dense on the support of the measure: this is why we explicitly require D to be dense. Also remark that a binary representation δ always induces an isomorphism (δ, δ^{-1}) between the Cantor space and the computable probability space.

The sequel of this section is devoted to the proof of the following result:

Theorem 2.3.1.3. *Every computable probability space $(\mathcal{X}, \mu)$ has a binary representation.*

The space, restricted to the domain D of the isomorphism, is then totally disconnected: the preimages of the cylinders form a basis of clopen and even *decidable* sets. In the whole space, they are not decidable any more. Instead, they are *almost decidable*.

In order to prove this theorem, we need the existence of a class of almost decidable balls $B(s,r)$ with an additional property.

Definition 2.3.1.6. An almost decidable set A is said to be ***exact*** if $U \cup V$ is dense, where U and V are the constructively open sets making A almost decidable. ⌟

Let $B(s,r)$ be a μ-continuous ball with computable radius: in general it is not exact (for instance, isolated points may be at distance exactly r from s). But if there is no ideal point at distance r from s, then $B(s,r)$ is exact: take $U = B(s,r)$ and $V = X \setminus \overline{B}(s,r)$.

Lemma 2.3.1.3. *There is a sequence $(r_n)_{n\in\mathbb{N}}$ of uniformly computable reals such that $(B(s_i, r_n))_{\langle i,n\rangle}$ is a basis of uniformly almost decidable balls which are furthermore exact.*

Proof. define $U_{\langle i,k\rangle} = \{r \in \mathbb{R}^+ : \mu(\overline{B}(s_i, r)) < \mu(B(s_i, r)) + 1/k\}$: by computability of μ, this is a constructively open subset of $\mathbb{R}^+$, uniformly in $\langle i, k\rangle$. It is furthermore dense in $\mathbb{R}^+$: the spheres $S_r = \overline{B}(s_i, r) \setminus B(s_i, r)$ form a partition of the space when r varies in $\mathbb{R}^+$ and μ is finite, so the set of r for which $\mu(S_r) \geq 1/k$ is finite. Define $V_{\langle i,j\rangle} = \mathbb{R}^+ \setminus \{d(s_i, s_j)\}$: this is a dense constructively open set, uniformly in $\langle i, j\rangle$. Then by the computable Baire Category Theorem (see [YMT99], [Bra01]), the dense constructive G_δ-set $\bigcap_{\langle i,k\rangle} U_{\langle i,k\rangle} \cap \bigcap_{\langle i,j\rangle} V_{\langle i,j\rangle}$ contains a sequence $(r_n)_{n\in\mathbb{N}}$ of uniformly computable real numbers which is dense in $\mathbb{R}^+$. In other words, all r_n are computable, uniformly in n. By construction, for any s_i and r_n, $B(s_i, r_n)$ is exact. □

We denote by $(B_i^\mu)_i$ the class of exact balls and by C_i^μ the set $X \setminus \overline{B_i^\mu}$ and define:

Definition 2.3.1.7. For $w \in \{0, 1\}^*$, the ***cell*** $\Gamma(w)$ is defined by induction on $|w|$:

$$\Gamma(\epsilon) = X, \quad \Gamma(w0) = \Gamma(w) \cap C_i^\mu \quad \text{and} \quad \Gamma(w1) = \Gamma(w) \cap B_i^\mu$$

where ϵ is the empty word and $i = |w|$. ⌟

Cells are exact sets, uniformly in w.

Proof. (of theorem 2.3.1.3). We construct an encoding function $b : D \to 2^\omega$, a decoding function $\delta : D_\delta \to X$, and show that δ is a binary representation, with $b = \delta^{-1}$.

Encoding. Let $D = \bigcap_i B_i^\mu \cup C_i^\mu$: this is a dense full-measure constructive G_δ-set. Define the computable function $b : D \to 2^\omega$ by:

$$b(x)_i = \begin{cases} 1 & \text{if } x \in B_i^\mu \\ 0 & \text{if } x \in C_i^\mu \end{cases}$$

Let $x \in D$: $\omega = b(x)$ is also characterized by $\{x\} = \bigcap_i \Gamma(\omega_{0..i-1})$. Let μ_δ be the image measure of μ by b: $\mu_\delta = \mu \circ b^{-1}$. b is then a morphism from (X, μ) to $(2^\omega, \mu_\delta)$.

Decoding. Let D_δ be the set of binary sequences ω such that $\bigcap_i \overline{\Gamma(\omega_{0..i-1})}$ is a singleton. We define the decoding function $\delta : D_\delta \to X$ by:

$$\delta(\omega) = x \text{ if } \bigcap_i \overline{\Gamma(\omega_{0..i-1})} = \{x\}$$

ω is called an expansion of x. Remark that $x \in B_i^\mu \Rightarrow \omega_i = 1$ and $x \in C_i^\mu \Rightarrow \omega_i = 0$, which implies in particular that if $x \in D$, x has a unique expansion, which is $b(x)$. Hence, $b = \delta^{-1} : \delta^{-1}(D) \to D$ and $\mu_\delta(D_\delta) = \mu(D) = 1$. We now show that $\delta : D_\delta \to X$ is a surjective morphism. For seek of clarity, the center and the radius of the ball B_i^μ will be denoted s_i and r_i respectively. Let us call i an *n-witness* for ω if $r_i < 2^{-(n+1)}, \omega_i = 1$ and $\Gamma(\omega_{0..i}) \neq \emptyset$.

• D_δ is a constructive G_δ-set: we show that $D_\delta = \bigcap_n \{\omega \in 2^\omega : \omega \text{ has a } n\text{-witness}\}$. Let $\omega \in D_\delta$ and $x = \delta(\omega)$. For each n, $x \in B(s_i, r_i)$ for some i with $r_i < 2^{-(n+1)}$. Since $x \in \overline{\Gamma(\omega_{0..i})}$, we have that $\Gamma(\omega_{0..i}) \neq \emptyset$ and $\omega_i = 1$ (otherwise $\overline{\Gamma(\omega_{0..i})}$ is disjoint of B_i^μ). In other words, i is an n-witness for ω. Conversely, if ω has a n-witness i_n for all n, since $\overline{\Gamma(\omega_{0..i_n})} \subseteq \overline{B_{i_n}^\mu}$ whose radius tends to zero, the nested sequence $(\overline{\Gamma(\omega_{0..i_n})})_n$ of closed cells has, by completeness of the space, a non-empty intersection, which is a singleton.

• $\delta : D_\delta \to X$ is computable. For each n, find some n-witness i_n of ω: the sequence $(s_{i_n})_n$ is a fast sequence converging to $\delta(\omega)$.

• δ is surjective: we show that each point $x \in X$ has at least one expansion. To do this, we construct by induction a sequence $\omega = \omega_0\omega_1 \dots$ such that for all i, $x \in \overline{\Gamma(\omega_0 \dots \omega_i)}$. Let $i \geq 0$ and suppose that $\omega_0 \dots \omega_{i-1}$ (empty when $i = 0$) has been constructed. As $B_i^\mu \cup C_i^\mu$ is open dense and $\Gamma(\omega_{0..i-1})$ is open, $\overline{\Gamma(\omega_{0..i-1})} = \overline{\Gamma(\omega_{0..i-1}) \cap (B_i^\mu \cup C_i^\mu)}$ which equals $\overline{\Gamma(\omega_{0..i-1}0)} \cup \overline{\Gamma(\omega_{0..i-1}1)}$. Hence, one choice for $\omega_i \in \{0, 1\}$ gives $x \in \overline{\Gamma(\omega_{0..i})}$. By construction, $x \in \bigcap_i \overline{\Gamma(\omega_{0..i-1})}$. As $(B_i^\mu)_i$ is a basis and $\omega_i = 1$ whenever $x \in B_i^\mu$, ω is an expansion of x.

□

Computable Lebesgue spaces

In the theory of algorithmic randomness, most result and constructions are done in the Cantor space with the uniform measure or, equivalently, in the unit interval with Lebesgue measure λ. In this section we present a tool to transfer some of these result to more general computable probability spaces .

Definition 2.3.1.8. A computable probability space is a ***computable Lebesgue space*** if it is isomorphic to the computable probability space $([0,1],\lambda)$ where λ is the Lebesgue measure. ⌟

Theorem 2.3.1.4. *Every computable probability space with no atoms is a computable Lebesgue space.*

We first prove the result for $I=([0,1],\mu)$.

Lemma 2.3.1.4. *The interval endowed with a non-atomic computable probability measure is a computable Lebesgue space.*

Proof. We define $F:[0,1]\to[0,1]$ by $F(x)=\mu([0,x])$. As μ has no atom and is computable, F is computable and surjective. As F is surjective, it has right inverses. Two of them are $G_<(y)=\sup\{x : F(x)<y\}$ and $G_>(y)=\inf\{x : F(x)>y\}$, and satisfy $F^{-1}(\{y\})=[G_<(y),G_>(y)]$. They are increasing and respectively left- and right-continuous. As F is computable, they are even lower- and upper semi-computable respectively. Let us define $D=\{y : G_<(y)=G_>(y)\}$: every $y\in D$ has a unique pre-image by F, which is then injective on $F^{-1}(D)$. The restriction of F on $F^{-1}(D)$ has a left-inverse, which is given by the restriction of $G_<$ and $G_>$ on D. Let us call it $G:D\to I$. By lower and upper semi-computability of $G_<$ and $G_>$, G is computable. Now, D is a constructive G_δ-set: $D=\bigcap_n\{y : G_>(y)-G_<(y)<1/n\}$. We show that $I\setminus D$ is a countable set. The family $\{[G_<(y),G_>(y)] : y\in I\}$ indexed by I is a family of disjoint closed intervals, included in $[0,1]$. Hence, only countably many of them have positive length. Those intervals correspond to points y belonging to $I\setminus D$, which is then countable. It follows that D has Lebesgue measure one (it is even dense). (F,G) is then an isomorphism between (I,μ) and (I,λ). □

Proof of the Theorem 2.3.1.4. We know from Theorem 2.3.1.3 that every CPS $(\mathcal{X},\mu)$ has a binary representation, which is in particular an isomorphism with the Cantor space (C,μ_C). The latter is isomorphic to (I,μ_I) where μ_I is the induced measure μ_I. If μ is non-atomic, so is μ_I. By the previous lemma, (I,μ_I) is isomorphic to (I,λ). □

We also note that:

Corollary 2.3.1.4. *Every computable probability space with no atoms is isomorphic to Cantor space with the uniform measure.*

2.3.2 Borel sets as a computable metric space

When one identify sets up to measure zero, the space of Borel sets can be metrized. Let $[\mathscr{B}]$ denote the class of Borel sets quotiented by the equivalence relation $A \sim B \Leftrightarrow A = B(\text{mod } 0)$. Let us consider over $[\mathscr{B}]$ the metric $d_\mu(A, B) = \mu(A \triangle B)$[2].

Theorem 2.3.2.1. *The triple $([\mathscr{B}], d_\mu, \mathscr{A})$ is a computable metric space.*

Proof. By theorem 2.3.1.1, $\mathscr{A}$ generates $\mathscr{B}$. The density of $\mathscr{A}$ in $[\mathscr{B}]$ follows then from theorem A.1.0.2. Let us show that for any A_i and A_j in $\mathscr{A}$, $d_\mu(A, B)$ is computable, uniformly in i, j. Since $\mathscr{A}$ is an algebra, $A_i \triangle A_j \in \mathscr{A}$. Say $A_i \triangle A_j = A_k$ and the number k can be computed from i, j. The result now follows from theorem 2.3.1.1. □

It follows that a computable element of $[\mathscr{B}]$ has a computable measure. Moreover,

Theorem 2.3.2.2. *The collection of computable elements of $[\mathscr{B}]$ is an algebra.*

Proof. If U is a computable Borel set, then U^c is also computable since $A \triangle U = A^c \triangle U^c$. Let U and V be computable Borel sets. Let us show that $U \cup V$ is computable. To compute $C_n \in \mathscr{A}$ such that $d_\mu(C_n, U \cup V) < 2^{-n}$, just take A_n and B_n in $\mathscr{A}$ such that $d_\mu(A_n, U) < 2^{-n-1}$ and $d_\mu(B_n, V) < 2^{-n-1}$ holds. This can be done since U and V are computable elements of $[\mathscr{B}]$. Then put $C_n = A_n \cup B_n$. The result follows from the relation $(A \cup B) \triangle (U \cup V) \subset (A \triangle U) \cup (B \triangle V)$. As $U \cap V = U^c \cup V^c$, the intersection is also a computable Borel set. □

Proposition 2.3.2.1. *If (φ, φ^{-1}) is an isomorphism, then the function $F : [\mathscr{B}^X] \to [\mathscr{B}^Y]$ mapping E into $\varphi(E)$ is an isometry.*

Proof. Since φ is an isomorphism, $\varphi(A \triangle B) = \varphi(A) \triangle \varphi(B)$ and the same holds for φ^{-1}. □

Remark 2.3.2.1. Let us say that f is a ***recursively simple function*** from X to $\overline{R}$ if it is of the form:

$$f = \sum_{i \leq k} q_i 1_{A_i}$$

[2]The triangular inequality follows from the relation $A \triangle B = (A \triangle C) \triangle (C \triangle B) \subset (A \triangle C) \cup (C \triangle B)$

where the A_i are almost decidable sets and $q_i \in \mathbb{Q}$. Recursively simple functions can be made a numbered set in a way that the q_i and the A_i can be recovered from the number, and vice-versa. Let us denote this set by $\mathfrak{S}$. Among recursively simple functions are the nonnegative ones, whose range is a subset of $[0, \infty)$. With the help of Theorem 2.3.1.2 one can prove, exactly as in the classical setting, that every measurable function $f : X \to [0, \infty]$ is a supremum of recursively simple functions. The integral of recursively simple functions is computable and the integral of any measurable function can be defined using recursively simple functions just as in the classical setting. Identifying functions which are equal almost everywhere gives rise to the computable metric space $(\mathcal{L}^1, \|\cdot\|_1, \mathfrak{S})$, whose computable points have a computable integral. ⌟

Part II

Randomness and Dynamical Systems

Chapter 3

Algorithmic randomness over general metric spaces

3.1 Introduction

The roots of algorithmic randomness go back to the work of Von Mises in the early 20th century. He suggested a notion of individual infinite random sequence based on *limit-frequency* properties invariant under the action of *selection* functions from some "acceptable" set. The problem was then to properly define what an "acceptable" selection function could be. Some years later, the concept of *computable* function was formalized, providing a natural class of functions to be considered as acceptable. This gave rise to Church's notion of *computable randomness* ([Chu40]). Nevertheless, substantial understanding was achieved only with the works of Kolmogorov [Kol65], Martin-Löf [ML66, ML71], Levin [ZL70, Lev84], Schnorr [Sch71, Sch72] and Chaitin [Cha75], since then, many efforts have contributed to the development of this theory (see for example [Asa88, Cha87, Cha90, Vov01, Dav01, Dav04]) which is now well established and intensively studied. Standard reference books are [Cal94, Cal02] and [LV93].

There are several different possible definitions, but it is Martin-Löf's one which has received most attention. This notion can be defined, at least, from three different points of view:

1. *measure theoretic*. This was the original presentation by Martin-Löf ([ML66]). Roughly, an infinite sequence is random if it satisfies all "effective" probabilistic laws (see defi-

nition 3.3.1.1).

2. *compressibility.* This characterization of random sequences, due to Schnorr and Levin (see [ZL70, Sch71, Cha75]), uses the prefix-free Kolmogorov complexity: random sequences are those which are maximally complex.

3. *predictability.* In this approach (started by Ville [Vil39] and reintroduced to the modern theory by Schnorr [Sch72]) a sequence is random if, in a fair betting game, no "effective" strategy ("martingale") can win an unbounded amount of money against it.

In [Sch71], a somewhat broader notion of algorithmic randomness (narrower notion of probabilistic law) was proposed: Schnorr randomness. This notion received less attention over the years: Martin-Löf's definition is simpler, leads to universal tests, and many equivalent characterizations (besides, Schnorr's book is not in English...). Recently, Schnorr randomness has begun to receive more attention. The work [DG02] for instance, characterizes it in terms of Kolmogorov complexity (with respect to *computable machines*).

In the measure theoretic presentation of the theory of algorithmic randomness, an infinite binary sequence is said to be ***random*** with respect to a given measure (not necessarily computable) if it *passes* all the *effective statistical tests*. An effective test can be defined in many ways, for now let us say that an ***effective test*** is a lower *semi-computable* integrable function $T : X \to \mathbb{R}^+$, and that x ***passes*** the test T if $T(x) < \infty$. The success of this theory lies in some outstanding results among which the existence of a ***universal test***[1] is a fundamental one. The problem of the extension of algorithmic randomness (in the Martin-Löf version) to a more general setting has been studied by different authors ([HW98, HW03, Gác05]), the question of existence of such *universal test*, being of particular interest. The strongest result (established by Gács in [Gác05]) guarantee the existence of a universal test under an additional computability condition on the space considered.

With the tools developed in the first part of the thesis we show that the existence of a universal test is guaranteed in any computable metric space (without any further condition).

Then we study algorithmic randomness in the particular case of computable probability spaces. We give some basic properties of Martin-Löf random points as well as (the gen-

[1] That is, a test t_u such that for every test t, there exists $c \in N$ for which $t \leq ct_u$. Hence x is random $\Leftrightarrow$ x pass t_u.

eralization of) Schnorr-random points, a slightly weaker notion, introduced by C. Schnorr ([Sch71]).

3.2 Martin-Löf randomness for arbitrary probability measures

On the Cantor space with a computable measure μ, Martin-Löf originally defined the notion of an individual random sequence as a sequence passing all μ-randomness tests. A μ-randomness test *à la Martin-Löf* is a sequence of uniformly constructively open sets $(U_n)_n$ satisfying $\mu(U_n) \leq 2^{-n}$. The set $\bigcap_n U_n$ has null measure, in an effective way: it is then called an effective null set. Equivalently, a μ-randomness test can be defined as a positive lower semi-computable function $t : 2^\omega \to \mathbb{R}$ satisfying $\int t d\mu \leq 1$ (see [VV93] for instance). The associated effective null set is $\{x : t(x) = +\infty\} = \bigcap_n \{x : t(x) > 2^n\}$. Actually, every effective null set can be put in this form for some t. A point is then called μ-random if it lies in no effective null set. We begin by studying Martin-Löf randomness with respect to arbitrary measures. Following Gács, we will use the second presentation of randomness tests and prove that over any computable metric space, a universal uniform randomness test always exists.

Definition 3.2.0.1. Given a measure $\mu \in \mathcal{M}(X)$, a ***μ-randomness test*** is a μ-constructive element t of $\mathcal{C}(X, \mathbb{R}^+)$, such that $\int t d\mu \leq 1$. Any subset of $\{x \in X : t(x) = +\infty\}$ is called a ***μ-effective null set***.
A ***uniform randomness test*** is a constructive function T from $\mathcal{M}(X)$ to $\mathcal{C}(X, \mathbb{R}^+)$ such that for all $\mu \in \mathcal{M}(X)$, $\int T^\mu d\mu \leq 1$ where T^μ denotes $T(\mu)$. ⌟

Note that T can be also seen as a lower-semi-computable function from $\mathcal{M}(X) \times X$ to $\mathbb{R}^+$ (see section 1.7).

A presentation *à la Martin-Löf* can be directly obtained using the functions below:

$$\begin{array}{rccc} F: & \mathcal{C}(X,\mathbb{R}^+) & \to & \tau^{\mathbb{N}} \\ & t & \mapsto & (t^{-1}(2^n, +\infty))_n \end{array} \qquad \begin{array}{rccc} G: & \tau^{\mathbb{N}} & \to & \mathcal{C}(X,\mathbb{R}^+) \\ & (U_n)_n & \mapsto & (x \mapsto \sup\{n : x \in \bigcap_{i\leq n} U_i\}) \end{array}$$

which are constructive, satisfy $F \circ G = id : \tau^{\mathbb{N}} \to \tau^{\mathbb{N}}$ and preserve the corresponding effective null sets.

A uniform randomness test T induces a μ-randomness test T^μ for all μ. We show two important results which hold on any computable metric space:

- the two notions are actually equivalent (theorem 3.2.0.3),
- there is a *universal* uniform randomness test (theorem 3.2.0.4).

The second result was already obtained by Gács, but only on spaces which have recognizable Boolean inclusions, which is an additional computability property on the basis of ideal balls.

By lemma 1.7.1.1, constructive functions from $\mathcal{M}(X)$ to $\mathcal{C}(X,\mathbb{R}^+)$ can be identified to constructive elements of the enumerative lattice $\mathcal{C}(\mathcal{M}(X),\mathcal{C}(X,\mathbb{R}^+))$. Let $(H_i)_{i\in\mathbb{N}}$ be an enumeration of all its constructive elements (proposition 1.5.2.1).

Lemma 3.2.0.1. *There is a constructive function $T:\mathbb{N}\times\mathcal{M}(X)\to\mathcal{C}(X,\mathbb{R}^+)$ satisfying:*

- *for all i, $T_i=T(i,\cdot)$ is a uniform randomness test,*
- *if $\int H_i(\mu)d\mu<1$ for some μ, then $T_i(\mu)=H_i(\mu)$.*

Proof. By proposition 1.7.1.2, the constructive functions $H_i:\mathcal{M}(X)\to\mathcal{C}(X,\mathbb{R}^+)$ can be seen as constructive functions $\hat{H}_i=Decurry(H_i)$ from $M(X)\times X$ to $\mathbb{R}^+$ which, by lemma 1.7.1.1, can be identified to constructive elements of the enumerative lattice $C(M(X)\times X,\mathbb{R}^+)$. Over this lattice, let us consider the set $\mathcal{H}^+=\{h_1,h_2,...\}$ of simple functions introduced in section 1.7.3. By proposition 1.7.3.1 the constructive elements of $C(M(X)\times X,\mathbb{R}^+)$, $\hat{H}_i$, are the supremum of recursive sequences of these simple functions, say: $\hat{H}_i=\sup_n h_{\varphi(i,n)}$ for some recursive φ. Let $\hat{H}_i^k=\sup_{n\leq k}h_{\varphi(i,n)}$ and $H_i^k=Curry(\hat{H}_i^k)$.

We are now able to define T: $T(i,\mu)=\sup\{H_i^k(\mu):\int H_i^k(\mu)d\mu<1\}$. By corollary 2.2.2.3, $L_{\hat{H}_i^k}:M(X)\to\mathbb{R}^+$ mapping μ into $\int\hat{H}_i^k(\mu,\cdot)d\mu=\int H_i^k(\mu)d\mu$ is computable uniformly in i,k. Hence T is a constructive function from $\mathbb{N}\times\mathcal{M}(X)$ to $\mathcal{C}(X,\mathbb{R}^+)$. □

As a consequence, every randomness test for a particular measure can be extended to a uniform test:

Theorem 3.2.0.3 (Uniformity vs non-uniformity)**.** *Let μ_0 be a measure. For every μ_0-randomness test t there is a uniform randomness test $T:\mathcal{M}(X)\to\mathcal{C}(X,\mathbb{R}^+)$ with $T(\mu_0)=\frac{1}{2}t$.*

Proof. let μ_0 be a measure and t a μ_0-randomness test: $\frac{1}{2}t$ is then a μ_0-constructive element of the enumerative lattice $\mathcal{C}(X,\mathbb{R}^+)$, so by corollary 1.7.1.1 there is a constructive element

H of $\mathcal{C}(\mathcal{M}(X), \mathcal{C}(X, \mathbb{R}^+))$ such that $H(\mu_0) = \frac{1}{2}t$ and hence, $H = H_i$ for some i. Since $\int H_i(\mu_0)d\mu_0 = \frac{1}{2}\int t d\mu_0 < 1$, the uniform test T_i constructed in lemma 3.2.0.1 satisfy $T_i(\mu_0) = \frac{1}{2}t$. □

Theorem 3.2.0.4 (Universal uniform test)**.** *There is a universal uniform randomness test, that is a uniform test T_u such that for every uniform test T there is a constant c_T with $T_u \geq c_T T$.*

Proof. it is defined by $T_u := \sum_i 2^{-i-1}T_i$: as every T_i is a uniform randomness test, T_u is also a uniform randomness test, and if T is a uniform randomness test, then in particular $\frac{1}{2}T$ is a constructive element of $\mathcal{C}(\mathcal{M}(X), \mathcal{C}(X, \mathbb{R}^+))$, so $\frac{1}{2}T = H_i$ for some i. As $\int H_i(\mu)d\mu = \frac{1}{2}\int T(\mu)d\mu < 1$ for all μ, $T_i(\mu) = H_i(\mu) = \frac{1}{2}T(\mu)$ for all μ, that is $T_i = \frac{1}{2}T$. So $T_u \geq 2^{-i-2}T$. □

Definition 3.2.0.2. Given a measure μ, a point $x \in X$ is called ***μ*-random** if $T_u^\mu(x) < \infty$. Equivalently, x is μ-random if it lies in no μ-effective null set. ⌟

The set of μ-random points is denoted by R_μ. This is the complement of the maximal μ-effective null set $\{x \in X : T_u^\mu(x) = +\infty\}$.

3.3 Algorithmic randomness on computable probability spaces

We study the particular case of a computable measure. Let $(\mathcal{X}, \mu)$ be then a computable probability space.

3.3.1 Martin-Löf randomness

In this section we will rather use the *à la Martin-Löf* presentation of randomness:

Definition 3.3.1.1. A Martin-Löf test (ML-test) is a sequence of uniformly constructively open sets $(A_n)_n$ such that $\mu(A_n) \leq 2^{-n}$. We say that x ***fails*** the ML-test if $x \in A_n$ for all n. Any subset of $\cap_n A_n$ is called an effective null set. x is called ML-random if it fails no ML-test. ⌟

Remarks 3.3.1.1.

1. In the previous section, a ML-random point was called μ-random. This was to emphasize the fact that μ was a parameter, and that we dealt with a *uniform* notion. This will be no more the case, and then we do not need to mention μ explicitly.

2. As the intersection of two (or any finite number of) constructively open sets is again constructively open, we will suppose that a ML-test $(A_n)_n$ always satisfy $A_{n+1} \subset A_n$.

⌟

The following definition corresponds to the effective version of the well known Borel-Cantelli lemma.

Definition 3.3.1.2. A ***Borel-Cantrell test*** (BC-test) is a uniform sequence $(C_n)_n$ of constructively open sets such that $\sum_n \mu(C_n) < \infty$. We say that x ***fails*** the BC-test if $x \in C_n$ infinitely often (i.o.). ⌟

The following proposition was proved in Cantor setting by Solovay. The same proof works in our general setting.

Proposition 3.3.1.1. *x fails a* ML*-test iff x fails a* BC*-test.*

Proof. Let $(A_n)_n$ be a ML-test. $\sum_n \mu(A_n) \leq \sum_n 2^{-n} < \infty$. Then it is a BC-test. Conversely, let $(C_n)_n$ be a BC-test. Let c be such that $\sum_n \mu(C_n) < 2^c$. Define the constructively open set $A_k := \{x : |\{n : x \in C_n\}| \geq 2^{k+c}\}$. Observe that $\mu(A_k) \leq \frac{2^c}{2^{k+c}} = 2^{-k}$ since there are at least 2^{k+c} repetitions. Since $x \in C_n$ i.o. if and only if $x \in A_k$ for all k, the result follows. □

Now we generalize and study a somewhat broader notion of algorithmic randomness, due to C. Schnorr.

3.3.2 Schnorr randomness

Definition 3.3.2.1. A ***Schnorr test*** (Sch-test) is a ML-test $(A_n)_n$ such that the sequence of reals $(\mu(A_n))_n$ is uniformly computable. We say that x fails the Sch-test if $x \in A_n$ for all n. Any subset of $\cap_n A_n$ is called a ***strong effective null set***. A point x is called Sch-***random*** if it fails no Sch-test. ⌟

Of course, a Sch-random point is in particular ML-random and hence, all results which hold for Sch-random points also hold for ML-random points.

Remark 3.3.2.1. From corollary 2.3.1.2 follows that the intersection of two (or any finite number of) constructively open sets with computable measure is again constructively open with computable measure. We will suppose that a Sch-test $(A_n)_n$ always satisfy $A_{n+1} \subset A_n$. ⌟

An important difference between Sch-tests and ML-tests is the fact that there is no recursive enumeration of all Sch-tests and hence, there is no *maximal strong effective null set*. Nevertheless, if $(T_k)_k$ is a uniform sequence of Sch-tests, put $\Lambda := \{x : \exists$ k such that x fails $T_k\}$. Then:

Proposition 3.3.2.1. *Λ is a strong effective null set.*

Proof. Let $T_k = (C^k_n)_n$. We define the Sch-test $T_{\max} = (A_i)_i$ where $A_i := \cup_{k \leq i} C^k_i$. The measure of A_i is computable and satisfy $\mu(A_i) \leq 2^{-i}i$, which converges effectively to 0. We can then extract a subsequence A_{i_n} such that $\mu(A_{i_n}) \leq 2^{-n}$ which is then a Sch-test satisfying $\cap_i A_i = \cap_n A_{i_n}$. Suppose x fails $T_k = (C^k_n)_n$ for some k. That is $x \in \cap_n C^k_n$. By remark 3.3.2.1 this is the same as $x \in \cap_{n \geq k} C^k_n$ which is included in $\cap_{i \geq k} A_i$. Hence x fails $T_{\max}$ too. □

Hence even if there is no *universal test*, we can still have a single test which detects all the regularities we need, in a given problem. This is since, usually, all objects concerned with a given problem are *uniformly computable*. This will be an important tool in chapter 5.

The following is the strong version of Borel-Cantelli tests.

Definition 3.3.2.2. A ***strong*** BC-***test*** is a BC-test $(C_n)_n$ such that $\sum_n \mu(C_n)$ is computable. ⌟

Which also characterize (Schnorr) randomness.

Proposition 3.3.2.2. *An element x fails a* Sch-*test if and only if x fails a strong* BC-*test.*

Proof. Let $(C_n)_n$ be a strong BC-test. Let c be such that $2^c > \sum_n \mu(C_n)$. Define the constructively open set $A_k := \{x : |\{n : x \in C_n\}| \geq 2^{k+c}\}$. Then $\mu(A_k) < 2^{-k}$. Observe that A_k is the union of all the (2^{k+c})-intersections of C_n's. Since $\mu(C_k) = \sum_n \mu(C_n) - \sum_{n \neq k} \mu(C_n)$ and the C_n's are constructively open, we have that $\mu(C_n)$ is computable (uniformly in n). We choose a basis $(B^i)_i$ of almost decidable balls to work with. Recall that finite unions or intersections of almost decidable sets are almost decidable too and that the measure of an almost decidable set is computable. Now we show that $\mu(A_k)$ is computable uniformly in k. Let $\epsilon > 0$ be rational. Let n_0 be such that $\sum_{n \geq n_0} \mu(C_n) < \frac{\epsilon}{2}$. Then $\mu(\bigcup_{n \geq n_0} C_n) < \frac{\epsilon}{2}$. For each C_n with $n < n_0$ we construct an almost decidable set $C_n^\epsilon \subset C_n$ (a finite union of almost decidable balls) such that $\mu(C_n) - \mu(C_n^\epsilon) < \frac{1}{n_0}\frac{\epsilon}{2}$. Then $\sum_{n<n_0}[\mu(C_n) - \mu(C_n^\epsilon] < \frac{\epsilon}{2}$. Define A_k^ϵ to be the union of the (2^{k+c})-intersections of the C_n^ϵ's for $n < n_0$. Then A_k^ϵ is almost decidable and then has a computable measure. Moreover $A_k \subset A_k^\epsilon \cup (\bigcup_{n \geq n_0} C_n) \cup (\bigcup_{n<n_0} C_n \setminus C_n^\epsilon)$, hence $\mu(A_k) - \mu(A_k^\epsilon) < \epsilon$. □

3.3.3 Some properties of random points

Let (X, μ) be a computable probability space . In the following we state some properties which hold for both notions of randomness. Let us then denote just by R_μ the set of random points (Martin-Löf or Schnorr).

Theorem 3.3.3.1. *Let A be an almost decidable set or a computably open set. Then*

1. *If $\mu(A) = 1$ then $R_\mu \subset A$.*

2. *$R_{\mu_A} = R_\mu \cap A$, where μ_A is the normalized measure.*

Proof. We prove 1: fix a basis of almost decidable balls to work with and show that the complement of a constructively open set of measure one is a strong effective null set. Let $U = \cup_i B_i$ be a constructively open set of measure one and define $U_k := \cup_{i \leq k} B_i$ and $\overline{U}_k := (\cup_{i \leq k} \overline{B}_i)^{\mathcal{C}}$. Both are constructively open almost decidable sets. Since $U^{\mathcal{C}} \subset \overline{U}_k$ for all k, from $(\overline{U}_k)_k$ we can extract the Sch-test we are looking for. We prove 2: By proposition 2.3.1.1 μ_A is a computable measure. If A is almost decidable and U and V are the associated constructively open sets, by 1 we have $R_\mu \subset U \cup V$ and $R_{\mu_A} \subset U \subset A$. So we can suppose $A = U$, where U is constructively open with a computable measure. We can then compute n_0 such that $2^{-n_0} \leq \mu(U)$. Hence, if V_n is a uniform sequence of constructively open sets satisfying $\mu(V) \leq 2^{-n}$, then $\hat{V}_n = V_{n+n_0}$ satisfy $\mu_A(\hat{V}_n) \leq 2^{-n}$. Moreover, if the V_n are computably open for μ, since $\mu_A(V_n) = \mu(V_n \cap A)\mu(A)^{-1}$, V_n is also computably open for μ_A. Then, any test T (Martin-Löf or Schnorr) for μ can be converted in a test (of the same kind) for μ_A. This proves $R_{\mu_A} \subset R_\mu \cap A$. Conversely, if V_n is a test for μ_A then it is easy to see that $\hat{V}_n = V_n \cap A$ is a test (of the same kind) for μ. The result follows. □

Morphisms of computable probability spaces behave well with respect to algorithmic randomness:

Proposition 3.3.3.1. *Morphisms of computable probability spaces are defined on Sch-random points and preserve randomness.*

Proof. To prove it, we shall use the following lemma:

Lemma 3.3.3.1. *The complement of a constructive G_δ set of measure one, is a strong effective null set.*

Proof. As the sets involved in the constructive G_δ-set of measure one are in particular computably open, the result follows from theorem 3.3.3.1 and proposition 3.3.2.1. □

As a consequence, morphisms are defined on random points. To see that a morphism $\phi : X \to Y$ preserves randomness, observe that if $(C_n)_n$ is a test in Y, then $(\phi^{-1}(C_n))_n$ is a test (of the same kind) in X intersected with $\mathrm{dom}(\phi)$. □

The following corollaries are straightforward.

Corollary 3.3.3.1. *Let $(F, G) : (\mathcal{X}, \mu) \rightleftarrows (\mathcal{Y}, \nu)$ be an isomorphism of computable probability spaces. Then there are two single strong effective null sets N_X and N_Y such that $F_{|X\setminus N_X}$ and $G_{|X\setminus N_Y}$ are total computable bijections between $X \setminus N_X$ and $X \setminus N_Y$, and $(F_{|X\setminus N_X})^{-1} = G_{|X\setminus N_Y}$. In particular, the same holds for R_μ and R_ν instead of $X \setminus N_X$ and $X \setminus N_Y$.*

In particular:

Corollary 3.3.3.2. *Let δ be a binary representation on a computable probability space $(\mathcal{X}, \mu)$. Each point having a Schnorr random expansion, is Schnorr random, and each Schnorr random point has a unique expansion, which is Schnorr random. The same holds for Martin-Löf random points.*

This proves that algorithmic randomness over a computable probability space could have been defined encoding points into binary sequences using a binary representation: this would have led to the same notion of randomness. Using this principle, a notion of Kolmogorov complexity characterizing randomness comes for free. For $x \in D$, define:

$$H_n(x) = H(\omega_{0..n-1}) \text{ and } \Gamma_n(x) = \delta([\omega_{0..n-1}])$$

where ω is the expansion of x and H is the prefix Kolmogorov complexity.

Corollary 3.3.3.3. *Let δ be a binary representation on a computable probability space $(\mathcal{X}, \mu)$. Then x is ML-random if and only if $x \in \mathrm{dom}(\delta)$ and:*

$$(\exists c)(\forall n) H_n(x) \geq -\log \mu(\Gamma_n(x)) - c$$

For Schnorr random points, it seems that a machine characterization is only available when the measure is uniform. Theorem A.3.0.10 directly implies:

Corollary 3.3.3.4. *Let φ be an isomorphism between (X, μ) and $(\{0,1\}^{\mathbb{N}}, \lambda)$ where λ is the uniform measure. Then x is Schnorr random if and only if $x \in \mathrm{dom}(\varphi)$ and for every computable machine M (see definition A.3.0.2),*

$$(\exists c)(\forall n) K_M(\varphi(x)_{1:n}) \geq n - c$$

All this allows to treat algorithmic randomness within probability theory over general metric spaces. In the next chapter it is applied, for instance, to easily show that in ergodic systems over metric spaces, ML-random points are well-behaved: they are *typical* with respect to any computable measure preserving transformation, generalizing what has been proved in [V'y97] for the Cantor space.

Let us make some comments on the definition of complexity used here to characterize randomness. For an infinite binary sequence ω, the knowledge of the prefix of length n can be understood in two ways:

- As the knowledge of a set of *small* measure to which ω belongs.
- As the knowledge of ω itself at finite *precision* 2^{-n}. This *precision* is actually relative to the standard metric of the symbolic space.

In the first, how *small* is the set we know ω belongs to, depends on the underlying measure. Instead, the *finite precision* 2^{-n} corresponds to a finite approximation of ω, given by the n-prefix, and which depends only on n.

Over a computable metric space, if complexity is defined using a coding into sequences as we did, the second interpretation is lost: the knowledge of x at *finite precision* 2^{-n}, $H_n(x)$, says a priori nothing about how far (in distance) we are from x. To get such a definition, the complexity of a point x at *precision* n should be defined as the complexity of some finite approximation (that is, an ideal point) lying in $B(x, 2^{-n})$ or something like that.

In the next chapter, in order to study the relations between random points and entropy, we shall be interested in both notions of complexity: via coding into sequences, and via finite approximations. There will be, however, a new ingredient under consideration: dynamics. It will be shown that, for our purposes, the two notions lead to the same results.

3.3.4 Random points and convergence of random variables

Let (X, μ) be a computable probability space . We will denote by ***MLR*** and ***SR*** the set of random points according to Martin-Löf and Schnorr respectively. We recall that a ***Stochastic process*** X_n is a sequence of random variables on (X, μ). Many statements of probability theory are about the convergence of some quantity $f(X_n)$ associated to a given stochastic process. For example, the law of large numbers is about the convergence of the time average $\frac{1}{n}\sum_{i=0}^{n-1} X_i$. In its strong form this law asserts that for i.i.d. processes, the sequence $S_n : X \to \mathbb{R}$ defined by $S_n(x) = \frac{1}{n}\sum_{i=0}^{n-1} X_i(x)$ converge to $E(X_0) = \int X_0 d\mu$ for almost every x.

More generally, consider a sequence (f_n) of random variables whose limit exists almost surely. That is, $\mu\{x \in X : \lim_n f_n(x) \text{ exists}\} = 1$. This is often understood as

“if we pick some x at random” then $\lim_n f_n(x)$ exists,

and we would like to replace “pick some x at random” with “pick a *random* x”. In the following we give conditions to do this. Let us introduce some notation. Given a sequence of measurable sets (A_n), let $\{A_n \text{ i.o.}\} := \cap_{k\geq 0} \cup_{n\geq k} A_n$ be the set of those x belonging to infinitely many A_n's. If (f_n) is a sequence of random variables, for any $\delta > 0$ let $D_{i,j}(\delta) = \{x : |f_i(x) - f_j(x)| > \delta\}$. Let $A_n(\delta) := \{x : \exists\, i, j \geq n,\, x \in D_{i,j}(\delta)\}$ be the set of points with at least one deviation of δ after n. Note that for these sets we have $\{A_n(\epsilon) \text{ i.o.}\} = \cap_n A_n(\epsilon)$.

Proposition 3.3.4.1. *Let f_n be a sequence of (arbitrary) random variables. Let $\delta_k \to 0$. Then the following statements are equivalent.*

i) $\lim_n f_n(x)$ *exists for each* $x \in MLR$ *(or* $x \in SR$*).*

ii) For all $k \geq 1$, $\bigcap_n A_n(\delta_k) \subset MLR^{\mathcal{C}}$ *(or* $\subset SR^{\mathcal{C}}$ *respectively).*

Proof. i) implies that the set $N = \{x : f_n(x) \text{ does not converge}\} \subset MLR^{\mathcal{C}}$ $(SR^{\mathcal{C}})$. Thus, $N = \cup_{k\geq 1} \cap_n A_n$. Then $\cap_n A_n \subset ML^{\mathcal{C}}$ $(SR^{\mathcal{C}})$ for all $k \geq 0$. Conversely, as for each k, $\bigcap_n A_n(\delta_k) \subset MLR^{\mathcal{C}}$ $(SR^{\mathcal{C}})$, their union too. Then f_n converge for each random x. □

When the random variables f_n are almost computable, the sets $A_n(\delta_k)$ are constructively open sets (intersected with some constructive G_δ set where the whole sequence is computable). As $\mu(\cap_n A_n(\delta_k)) = 0$ is equivalent to $\mu(A_n(\delta_k)) \to 0$, we will say that the sequence of random variables f_n ***converges effectively*** if $\mu(A_n(\delta_k))$ converges effectively to 0, for all k.

Proposition 3.3.4.2. *Let f_n be a sequence of almost computable random variables. If the sets $\mu(A_n(\delta_k))$ converges effectively to 0, then $\lim_n f_n(x)$ exists for each Schnorr-random point x.*

Proof. Let $\delta_k \to 0$. For any subsequence n_i we have that

$$\bigcap_{n\geq 0} A_n(\delta_k) \subset \bigcap_{i\geq 0} A_{n_i}(\delta_k)$$

As $\mu(A_n(\delta_k))$ converges effectively to 0, we can choose n_i such that $\mu(A_{n_i}) \leq 2^{-i}$. Then $\hat{A}_i = A_{n_i}$ is an effective null set. Let us show that is it a strong effective null set. In order to compute $\mu(\hat{A}_i)$ up to ϵ, find l such that $\mu(A_l) \leq \epsilon$. Hence, $\mu(\hat{A}_i) - \epsilon \leq \mu(\bigcup_{0\leq i,j\leq l} D_{i,j}(\delta_k))$ and by Theorem 2.3.1.2, we can choose δ_k such that $\mu(D_{i,j}(\delta_k))$ is computable. □

Let f_n be a sequence of random variables which converges almost everywhere to a function f. Suppose that f_n satisfy the conditions of Proposition 3.3.4.2. Then if x is random, we know that $\lim_n f_n(x)$ exists but... what about its value? can we assure it is equal to $f(x)$?. If f is almost computable, we can.

Proposition 3.3.4.3. *Let f_n be a sequence of almost computable functions which converge to f almost everywhere. If f is almost computable, then for each $x \in SR$,*

$$\limsup_n f_n(x) \geq f(x) \geq \liminf_n f_n(x) \tag{3.1}$$

holds.

Proof. We only show $\liminf_n f_n(x) \leq f(x)$ for each $x \in SR$. The other inequality can be proved in a similar way. We prove that the set $\{x : \liminf_n f_n(x) > f(x)\}$ is a strong effective null-set. Let δ_k be a sequence of uniformly real numbers converging to 0. Theorem 2.3.1.2 applied to the sequence $\hat{f}_n := f_n - f$ allows to choose δ_k such that $\mu(f_n^{-1}\{f(x) + \delta_k)\}) = 0$ for all n, k. Hence the sets $A_n(k) := \{x : f_n(x) > f(x) + \delta_k\}$ are almost decidable uniformly in n, k. On the other hand, the set $\{x : \liminf_n f_n(x) > f(x)\}$ is equal to

$$\{x : \exists k \geq 0 \text{ and } N \geq 0, \forall n \geq N, f_n(x) > f(x) + \delta_k\} = \bigcup_k \bigcup_N \bigcap_{n \geq 0} A_{n+N}(k)$$

Let $\hat{A}_n(k) = \bigcap_{i \leq n} A_i(k)$. Then $\bigcap_n A_n(k) = \bigcap_n \hat{A}_n(k)$.. Since $\mu(A_{n+N}(k))$ is computable uniformly in n (and k), so is $\mu(\hat{A}_n(k))$. As $f_n \to f$ almost everywhere, $\mu(\bigcap_{n \geq 0} A_{n+N}(k)) = 0$ for each N and then we can extract a subsequence n_i such that $\mu(A_{n_i+N}(k)) \leq 2^{-i}$. Hence $C_i^{N,k} := \hat{A}_{n_i+N}(k)$ is a Schnorr test for each N, k and satisfy $\bigcap_{n \geq N}\{x : f_n(x) > f(x) + \delta_k\} = \bigcap_i C_i^{N,k}$. The result follows. $\square$

Chapter 4

Random points and ergodic theorems

4.1 Introduction

The randomness of a particular outcome is always relative to some statistical test. The notion of algorithmic randomness, defined by Martin-Löf in 1966, is an attempt to have an "absolute" notion of randomness. This absoluteness is actually relative to all "effective" statistical tests, and lies on the hypothesis that this class of tests is sufficiently wide.

Martin-Löf's original definition was given for infinite symbolic sequences. With this notion each single random sequence behave as a generic sequence of the probability space for each effective statistical test. In this way many probabilistic theorems having almost everywhere statements can be translated to statements which hold *for each* random sequence. As an example we cite the fact that in each infinite string of 0's and 1's which is random for the uniform measure, all the digits appear with the same limit frequency. This can be seen as corollary of the V'yugin ergodic theorem for individual random sequences (see [V'y97] and lemma 4.3.2.2 below).

A particularly interesting class of stationary stochastic processes is constituted by those generated by a measure preserving map on a metric space (these are the objects studied by ergodic theory).

Let $(\mathcal{X}, \mu)$ be a computable probability space and let MLR and SR be the set of random points according to Martin-Löf and Schnorr respectively. The aim of this chapter is to study the set of random points from a dynamical point of view. That is, we will put a dynamic T on $(\mathcal{X}, \mu)$ (an endomorphism of computable probability spaces), and look at the abilities of random points (which are *a priori* independent of T) to describe the statistical properties of T. We recall that a Borel set A is called T-***invariant*** if $T^{-1}(A) = A(\text{mod } 0)$ and that the transformation T is said to be ***ergodic*** if every T-invariant set has measure 0 or 1.

In the classical ergodic theory, a powerful technique (symbolic dynamics) allows to associate to a general system as above (X, T, μ) a shift on a space of infinite strings having similar statistical properties. In section 4.2 we use the algorithmic features of computable metric spaces and its random points to define and construct effective symbolic model for the dynamics. In this models random points are associated to random infinite strings, and we will use this tool to generalize theorems which are proved in the symbolic setting to the more general setting.

We will first consider two main results of ergodic theory, namely the Poincaré recurrence theorem and the Birkhoff ergodic theorem.

In section 4.3 we prove that each random point is recurrent (a sort of Poincaré recurrence theorem for random points) and the generalization (thm. 4.3.2.1) of the above mentioned V'yugin ergodic theorem for random points to computable measure preserving transformations on computable metric spaces. Concerning Schnorr's concept of randomness, we prove that Schnorr random points are exactly those following simultaneously the statistical behaviour of a certain class of *mixing* systems. This will also be used in the last sections of the chapter to investigate the orbit complexity of random points in such dynamical systems.

In section 4.4 and following, we consider the orbit complexity of random points and its relations with the entropy of the system.

The well known notion of measure theoretic entropy (see section 4.4) of a dynamical system (also called Kolmogorov-Sinai entropy) was inspired by Shannon theory of information. The entropy of a system is a measure of the rate of Shannon information which is necessary to describe the dynamics. We remark that Shannon information is a global average notion, which depends on the probability measure which is considered on the space.

In 1965, Kolmogorov defined an algorithmic notion of information content of a single string. This information does not depend on the measure and was actually intended to provide an absolute notion of information and individual randomness. In this setting a sequence will be called random if it contains maximal information. But Martin-Löf proved that no sequence could be random in this sense, which lead him to propose his definition. Later, the original idea of Kolmogorov was refined, and was proved to give the notion of Martin-Löf randomness (see theorem A.3.0.9).

The orbit complexity of a point x is a measure of the information rate which is necessary to describe the behavior of the orbit of x. In this pointwise definition the information is measured by the Kolmogorov information content. In [Bru83] orbit complexity is defined for dynamical systems acting on metric spaces and it is proved that if the system is ergodic, the orbit complexity of almost each point equals the entropy of the system. In section 4.4 we introduce a definition of orbit complexity using effective symbolic dynamics, we compare this notion with the classical one obtaining (thm. 4.4.2.1) that they coincide at each random point (and hence on a total measure set). By this we prove (thm. 4.4.3.3) that in an ergodic computable measure preserving system, the orbit complexity of *each random* point coincides with the entropy of the system.

All these statements require that the dynamics and the invariant measure are computable.

The first assumption can be easily checked on concrete systems if the dynamics is given by a map which is effectively defined.

The second is more delicate: it is well known that given a map on a metric space, there can be a continuous (even infinite dimensional) space of probability measures which are invariant for the map, and many of them will be non computable. An important part of the theory of dynamical systems is devoted to select measures which are particularly meaningful. From this point of view, an important class of these measures is the class of SRB invariant measures, which are measures being in some sense the "physically meaningful ones"(for a survey on this topic see [You02]). In the next chapter we shall prove that in several classes of dynamical systems where SRB measures are proved to exist, these measures are also computable from our formal point of view, hence providing several classes of nontrivial concrete examples where our results can be applied.

4.2 Symbolic dynamics: the Computable Viewpoint

Let T be an endomorphism of the (Borel) probability space (X, μ). In the classical construction, one considers access to the system given by a finite measurable partition, that is a finite collection of pairwise disjoint Borel sets $\mathcal{P} = \{p_1, \ldots, p_k\}$ such that $\mu(\cup_i p_i) = 1$. Then, to (X, μ, T) a ***symbolic dynamical system*** $(X_\mathcal{P}, \sigma)$ is associated (called the symbolic model of $(X, T, \mathcal{P})$). The set $X_\mathcal{P}$ is a subset of $\{1, 2, \ldots, k\}^\mathbb{N}$. To a point $x \in X$ corresponds an infinite sequence $\omega = (\omega_i)_{i\in\mathbb{N}} = \phi_\mathcal{P}(x)$ defined as

$$\phi_\mathcal{P}(x) = \omega \Leftrightarrow \forall j \in \mathbb{N}, T^j(x) \in p_{\omega_j}.$$

The transformation $\sigma : X_\mathcal{P} \to X_\mathcal{P}$ is the ***shift*** defined by $\sigma((\omega_i)_{i\in\mathbb{N}}) = (\omega_{i+1})_{i\in\mathbb{N}}$.

As $\mathcal{P}$ is a measurable partition, the map $\phi_\mathcal{P}$ is measurable and then the measure μ induces the measure $\mu_\mathcal{P}$ (on the associated symbolic model) defined by $\mu_\mathcal{P}(B) = \mu(\phi_\mathcal{P}^{-1}(B))$ for all measurable $B \subset X_\mathcal{P}$.

The requirement of $\phi_\mathcal{P}$ being measurable makes the symbolic model appropriate from the measure-theoretic view point, but is not enough to have a symbolic model compatible with the computational approach:

Definition 4.2.0.1. Let T be an endomorphism of the computable probability space $(\mathcal{X}, \mu)$ and $\mathcal{P} = \{p_1 \ldots, p_k\}$ a finite measurable partition. The associated symbolic model $(X_\mathcal{P}, \mu_\mathcal{P}, \sigma)$ is said to be ***an effective symbolic model*** if the map $\phi_\mathcal{P} : X \to \{1, \ldots, k\}^\mathbb{N}$ is a morphism of CPS (here the space $\{1, \ldots, k\}^\mathbb{N}$ is endowed with the standard computable structure). ⌟

The sets p_i are called the ***atoms*** of $\mathcal{P}$ and we denote by $\mathcal{P}(x)$ the atom containing x (if there is one). Observe that $\phi_\mathcal{P}$ is computable on its domain only if the atoms are constructively open sets (in the domain):

Definition 4.2.0.2 (computable partitions)**.** A measurable partition $\mathcal{P}$ is said to be a ***computable partition*** if its atoms are constructively open sets. ⌟

Conversely:

Proposition 4.2.0.4. *Let T be an endomorphism of the CPS (X, μ) and $\mathcal{P} = \{p_1 \ldots, p_k\}$ a finite computable partition. Then the associated symbolic model is effective.*

Proof. Define the full-measure constructive G_δ-set:

$$X^\mathcal{P} = \bigcap_{n\in\mathbb{N}} T^{-n}(p_1 \cup \ldots \cup p_k)$$

Since the sets $p_i \in \mathcal{P}$ are constructively open, for all $x \in X^{\mathcal{P}}$ we can decide which of p_i, x belongs to. This proves that $\phi_{\mathcal{P}}$ is computable over $X^{\mathcal{P}}$. Proposition 2.3.0.4 allows to conclude. □

After the definition an important question is: are there computable partitions?

Corollary 4.2.0.1. *On every computable probability space, there exists a family of uniformly computable partitions which generates the Borel σ-algebra.*

Proof. Take $\mathcal{P}_k = \{B^k, X \setminus B^k\}$: as the almost decidable balls form a basis of the topology, the σ-algebra generated by the P_k is the Borel σ-field. □

4.3 The statistics of random points

With the tools developed so far, it is possible to translate many results of the form

$$\mu\{x : P(x)\} = 1,$$

with P some predicate, into an "individual" result of the form:

"If x is μ-random, then $P(x)$".

We start by a result holding for both notions of randomness. It is the pointwise version of the Poincaré recurrence theorem.

4.3.1 Random points are recurrent

One property that is enjoyed by all measure preserving systems is recurrence:

Definition 4.3.1.1. Let X be a metric space. A point $x \in X$ is said to be ***recurrent*** for a Borel-measurable transformation $T : X \to X$, if $\liminf_n d(x, T^n x) = 0$. ⌟

Poincaré recurrence theorem asserts that in a separable metric space, almost every point is recurrent. Here is the pointwise version.

Proposition 4.3.1.1 (Random points are recurrent)**.** *Let (X, μ) be a computable probability space. If x is random, then it is recurrent with respect to* every *endomorphism T on (X, μ).*

Proof. Let x be (Schnorr-)random. We prove that for each almost decidable neighborhood B of x, x returns infinitely often to B under positive iterations by T. We note that for each such a neighborhood B, $\mu(B) > 0$ holds (since x is random). For each $N \geq 0$ let $U_N = \bigcup_{n \geq N} T^{-n}B$. Then $\bigcap_{N \geq 0} U_N$ is the set of all points of X which enter B infinitely often under positive iterations by T. Hence the set $B_0 = B \cap \bigcap_{N \geq 0} U_N$ consists of all points of B that enter B infinitely often. Then, by the classical Poincaré's recurrence theorem, $\mu(B_0) = \mu(B)$. Now, the set $B_0 \cap D$ where D is the domain of computability of T is a constructive G_δ-set. Moreover, it is a full measure set for the normalized measure μ_B. Hence by Theorem 3.3.3.1, $R_\mu \cap B = R_{\mu_B} \subset B_0 \cap D$. Thus, $x \in B_0$. $\square$

4.3.2 Algorithmic randomness v/s typicalness

We now compare our two randomness notions to a property stronger than recurrence: typicality. Let us then introduce this concept. Let (X, μ) be a probability space and T an ergodic continuous transformation on X. Let $C_b(X)$ be the space of bounded real-valued continuous functions on X. In such systems the famous Birkhoff ergodic theorem says that the time average computed along the orbit $\{x, T(x), T^2(x), ..\}$ coincide with the space average with respect to μ, for almost every orbit. More precisely, for any $f \in L^1(X)$ it holds

$$\lim_{n \to \infty} \frac{S_n^f(x)}{n} = \int f d\mu, \tag{4.1}$$

for μ-almost each x, where $S_n^f = f + f \circ T + \ldots + f \circ T^{n-1}$.

Given the transformation T, if a point x satisfies equation (4.1) for a certain ***observable*** f, then we say that x is (T, f)***-typical***.

Definition 4.3.2.1. If x is (T, f)-typical w.r. to every bounded continuous function $f : X \to \mathbb{R}$, then we call it a ***T-typical point***. ⌟

First of all, let us prove a useful lemma. We say that a collection $\mathcal{O} = \{f_1, f_2, ..\}$ of integrable functions is an ***essential family of observables*** if for every open set U there is a sequence $(f_i)_i$ in $\mathcal{O}$ such that $f_i \leq 1_U$ (where 1_U denotes its indicator function) for all i and $\lim_i \int f_i = \mu(U)$. We remark that any essential family of events (see definition A.1.0.1) induce an essential family of observables taking their indicator functions.

Lemma 4.3.2.1. *Let $\mathcal{O}$ be an essential family of observables. If x is typical w.r. to every $f \in \mathcal{O}$, then x is a T-typical point.*

Proof. We have to show that equation (4.1) holds for any bounded continuous observable f. First, we extend equation (4.1) to every continuity open set C. Let $(f_k)_k$ be a sequence of elements of $\mathcal{O}$ such that $f_k \leq 1_{\mathrm{Int}(C)}$ for all k and $\lim_k \int f_k = \mu(C)$. Then for all k:

$$\liminf_n \frac{1}{n} \sum_{i=0}^{n-1} 1_C \circ T^i(x) \geq \lim_n \frac{1}{n} \sum_{i=0}^{n-1} f_k \circ T^i(x) = \int f_k$$

so

$$\liminf_n \frac{1}{n} \sum_{i=0}^{n-1} 1_C \circ T^i(x) \geq \mu(C).$$

Applying the same argument to $X \setminus C$ gives the result. Now we extend the result to bounded continuous functions. Let f be continuous and bounded ($|f| < M$) and let $\epsilon > 0$ be a real number. Then, since the measure μ is finite, there exist real numbers $r_1, \ldots, r_k \in [-M, M]$ (with $r_1 = -M$ and $r_k = M$) such that $|r_{i+1} - r_i| < \epsilon$ for all $i = 1, \ldots, k-1$ and $\mu(f^{-1}(\{r_i\})) = 0$ for all $i = 1, \ldots, k$. It follows that for $i = 1, \ldots, k-1$ the sets $C_i = f^{-1}(]r_i, r_{i+1}[)$ are all continuity open sets. Hence the function $f_\epsilon = \sum_{i=1}^{k-1} r_i 1_{C_i}$ satisfies $\|f - f_\epsilon\|_\infty \leq \epsilon$ and then the result follows by density. □

Mathematically, T-typical points are those whose orbits under T reproduce the main statistical features of μ. Actually, from the orbit (under T) of any T-typical point (which form a total measure set) one can (weakly) recover the measure. In a more philosophical sense, under the hypothesis that the evolution of the observed physical system is actually given by T, these points represent "physically plausible" initial conditions: they follow the "expected" behaviour of the system. Still, these points may be "non physical" with respect to others systems having similar statistical properties. Let us explain this by the following simple example.

Example 4.3.2.1. We restrict ourself to one dimension. Suppose that "space" is modeled by the unit interval $[0, 1]$, and consider the Lebesgue measure λ on it. Is a dyadic number "physically plausible"?. Imagine we are observing a rather "rigid" process, modeled by $T(x) = x + \alpha (\mathrm{mod}\ 1)$, with irrational α. Lebesgue is the only measure that T preserves, and it is then ergodic. This implies that every point $x \in [0, 1]$ (in particular any dyadic number) is typical (and hence physically plausible) for this system. On the other hand, consider a rather "random" process for which Lebesgue measure is also ergodic, modeled for instance by $D(x) = 2x (\mathrm{mod}\ 1)$. For this system one expect orbits to be dense (and even equally distributed) in the space whereas the orbit of any dyadic number is eventually constant (equal to 0). Hence, dyadic numbers do not really represent "physical" points. ⌟

Following this idea, one could call a point *physical* provided it is T-typical with respect to all dynamics but, if we consider all possible dynamics, then there are no such points. Hence, we should restrict our attention to a smaller set of transformations. Of course, here we will choose "computable" ones.

Definition 4.3.2.2. Let (X, μ) be a computable probability space. We say that a point $x \in X$ is $\boldsymbol{\mu}$***-typical***[1] (or just typical) if it is T-typical for every ergodic endomorphism T. ⌟

In the rest of this section we study the relations between "physicalness" and algorithmic randomness.

...for Martin-Löf random points

For Martin-Löf random points, this problem has already been studied by V'yugin ([V'y97]) in the particular case of the Cantor space and for computable observables. We prove a general version which applies to computable dynamics on any CPS, for any observable. The strategy is simple: we use computable partitions to construct effective symbolic models and then apply the following particular case of V'yugin's main theorem.

Lemma 4.3.2.2. *Let μ be a computable shift-invariant ergodic measure on the Cantor space $\{0,1\}^\omega$. Then for each μ-random sequence ω:*

$$\lim_n \frac{1}{n} \sum_{i=0}^{n} \omega_i = \mu([1]) \tag{4.2}$$

We are now able to prove:

Theorem 4.3.2.1. *Let (X, μ) be a computable probability space. Then each ML-random point x is typical.*

Proof. First, let us show that if A is an almost decidable set then for all ML-random point x:

[1] These points represent "physical" points, in a sense. We use the word *typical* for two reasons: i) it seems to be the standard terminology in ergodic theory (although in the classical theory there is no natural way to define typicalness with respect to a whole class of dynamics) and ii) the word "physical" may be too controversial and then it must be used with prudence.

$$\lim_n \frac{1}{n}\sum_{i=0}^{n} f_A \circ T^i(x) = \mu(A) \tag{4.3}$$

Indeed, consider the *computable partition* defined by $\alpha := \{A, X \setminus A\}$ and the associated symbolic model $(X_\alpha, \sigma, \mu_\alpha)$. By proposition 4.2.0.4, $\phi_\alpha(x)$ is a well defined μ_α-random infinite sequence, so lemma 4.3.2.2 applies and gives the result. As almost decidable sets form an essential family of events, lemma 4.3.2.1 allows to conclude. □

Remark 4.3.2.1. We remark that V'yugin's proof do not really use the particular features of the Cantor space, nor the computability of the measure, so it can be easily adapted to hold on any computable metric space, with an arbitrary measure. ⌟

After this, a natural question is:

are typical points ML-random ?

This remains an open problem. Instead, we are able to prove that if a point is typical, then it is Schnorr random.

...for Schnorr randomness

In [V'y97], to prove that ML-random points are typical w.r. to some f, a BC-test is constructed such that if a point fails the Birkhoof theorem then it fails the test. To have the result for Sch-random points, we should be able to compute the measure of the sets involved in the BC-test. This problem can be overcome if the system has certain "mixing" or "loss of memory". This is naturally expressed by means of the ***correlation functions***. For integrable functions f, g let

$$\begin{aligned} C(f,g) &= \mu(f \cdot g) - \mu f \cdot \mu g, \\ C_n(f,g) &= C(f \circ T^n, g). \end{aligned}$$

which measures the dependence between the observables f and g (possibly with $f = g$) at times $n \gg 1$ and 0 respectively. Note that $C_n(f,g) = 0$ corresponds, in probabilistic terms, to $f \circ T^n$ and g being uncorrelated random variables. We will be interested in systems for which observables become more and more independent in time. More precisely, we say that T has ***polynomial decay of correlations w.r. to the observables*** f and g if there are computable constants $c_{f,g} > 0$ and $\alpha > 0$ such that:

$$|C_n(f,g)| \leq \frac{c_{f,g}}{n^\alpha} \qquad \text{for all } n \geq 1.$$

Definition 4.3.2.3. We say that a system (X, T, μ) is (polynomially) ***mixing*** if there is $\alpha > 0$ and an essential family $O = \{f_1, f_2, ...\}$ of uniformly μ-almost computable observables such that for each i, j there is $c_{i,j} > 0$ computable in i, j such that

$$|C_n(f_i, f_j)| \leq \frac{c_{i,j}}{n^\alpha} \qquad \text{for all } n \geq 1.$$

We say that the system is ***independent*** if all correlation functions $C_n(f_i, f_j)$ are 0 for sufficiently large n. ⌟

Examples of non-mixing but still ergodic systems are given for instance by irrational circle rotations with the Lebesgue measure[2]. Examples of mixing but not independent systems are given by piecewise expanding maps or uniformly hyperbolic systems which have a distinguished ergodic measure (called SRB measure and which is "physical" in some sense) with respect to which the correlations decay exponentially (see [Via97]). An example of a mixing system for which the decrease of correlations is only polynomial and not exponential, is given by the class of *Manneville-Pomeau* type maps (non uniformly expanding with an indifferent fixed point, see [Iso03]). For a survey see [You02]. All these examples will be treated in some detail in the last chapter. In particular, we shall prove that in each case, the physical invariant ergodic measure is computable.

Now we prove:

Theorem A. *Let $(\mathcal{X}, \mu)$ be a computable probability space which no atoms. The following properties of a point $x \in X$ are equivalent.*

1. *x is Schnorr random.*
2. *x is T-typical for every mixing endomorphism T.*
3. *x is T-typical for every independent endomorphism T.*

Remark 4.3.2.2. If the measure μ is atomic, it is easy to see that:

1. (X, μ) admits a mixing endomorphism if and only if $\mu = \delta_x$ for some x. In this case the theorem still holds, the only random point being x.
2. (X, μ) admits an ergodic endomorphism if and only if $\mu = \frac{1}{n}(\delta_{x_1} + ... + \delta_{x_n})$ (where $x_i \neq x_j$, for all $i \neq j$). In this case, a point x is Schnorr random if and only if it is typical for every ergodic endomorphism if and only if it is an atom.

[2]These systems are uniquely ergodic, so that every point is typical, which from our point of view is a less interesting situation.

$\lrcorner$

Proof of (1) $\Rightarrow$ (2) We shall use the following theorem:

Theorem 4.3.2.2. *If T has polynomial decay of correlations w.r. to a μ-almost computable observable ϕ, then the set of points which are not typical w.r to ϕ is contained in a strong effective null set.*

Proof. For $\delta_n > 0$, define the deviation sets:

$$A_n^{\phi}(\delta_n) = \left\{ x \in X : \left| \frac{S_n^{\phi}(x)}{n} - \int \phi d\mu \right| > \delta_n \right\}.$$

By Theorem 2.3.1.2 we can choose δ_n such that $A_n^{\phi}(\delta_n)$ is almost decidable. Then their measures are computable, uniformly in n.

By the Tchebytchev inequality,

$$\mu(A_n^{\phi}(\delta_n)) \leq \frac{1}{\delta_n^2} \left\| \frac{S_n^{\phi}(x)}{n} - \int \phi d\mu \right\|_{L^2}^2 .$$

Let us change ϕ by adding a constant to have $\int \phi d\mu = 0$. This does not change the above quantity. Then

$$\left\| \frac{S_n^{\phi}(x)}{n} - \int \phi d\mu \right\|_{L^2}^2 = \int \left(\frac{S_n^{\phi}(x)}{n} \right)^2 d\mu = \int \left(\frac{\phi + \phi \circ T + ... + \phi \circ T^{n-1}}{n} \right)^2 d\mu.$$

By invariance of μ this is equal to

$$\frac{1}{n^2} \int n\phi^2 d\mu + \frac{2}{n^2} \int \big(\sum_{i<j<n} \phi \circ T^{j-i} \phi d\mu \big) \, d\mu$$

hence

$$\begin{aligned} \delta_n^2 \mu(A_n^{\phi}(\delta_n)) &\leq \frac{\|\phi\|_{L^2}^2}{n} + \frac{2}{n} \sum_{k<n} |C_k(\phi, \phi)| \\ &\leq \frac{\|\phi\|_{L^2}^2}{n} + \frac{2c_{\phi,\phi}}{(1-\alpha)n^{\alpha}}. \end{aligned}$$

(Observe that α can be replaced by any smaller positive number, so we assume $\alpha < 1$.) Let us chose $\delta_n^2 \sim n^{\gamma}$, with $0 < \gamma < \alpha$. Hence, $\mu(A_n^{\phi}(\delta_n)) \leq Cn^{-\alpha'}$ for some constants C and $0 < \alpha' = \alpha - \gamma < 1$. Now, it is easy to find a sequence $(n_i)_{i \in \mathbb{N}}$ such that the subsequence

$(n_i^{-\alpha'})_i$ is effectively summable and $\frac{n_i}{n_{i+1}} \to 1$ (take for instance $n_i = i^\beta$ with $\alpha'\beta > 1$). This shows that the sequence $A^\phi_{n_i}(\delta_{n_i})$ is a strong BC-test. Therefore, if x is not in the associated strong null set, the subsequence $\frac{S^\phi_{n_i}(x)}{n_i}$ converges to $\int \phi d\mu$.

To show that for such points the whole sequence $\frac{S^\phi_n(x)}{n}$ converges to $\int \phi = \mu(E)$, observe that if $n_i \leq n < n_{i+1}$ and $\beta_i := \frac{n_i}{n_{i+1}}$ then we have:

$$\frac{S^\phi_{n_i}}{n_i} - 2(1-\beta_i)M \leq \frac{S^\phi_n}{n} \leq \frac{S^\phi_{n_{i+1}}}{n_{i+1}} + 2(1-\beta_i)M,$$

where M is a bound of ϕ. To see this, for any k, l, β with $\beta \leq k/l \leq 1$:

$$\begin{aligned} \frac{S^\phi_k}{k} - \frac{S^\phi_l}{l} &= \left(1 - \frac{k}{l}\right)\frac{S^\phi_k}{k} - \frac{S^\phi_{l-k} \circ T^{l-k}}{l} \\ &\leq (1-\beta)M + \frac{(l-k)M}{l} = 2(1-\beta)M, \end{aligned}$$

Taking $\beta = \beta_i$ and $k = n_i$, $l = n$ first and then $k = n$, $l = n_{i+1}$ gives the result. □

Corollary 4.3.2.1 ((1) ⇒ (2))**.** *Let $(\mathcal{X}, \mu)$ be a CPS. If a point x is Schnorr-random then it is T-typical for any mixing endomorphism T.*

Proof. If x is Schnorr random (it is outside any strong effective null set), by theorem 4.3.2.2 it is typical w.r.t. to each $\phi \in \mathcal{O}$. The result then follows from lemma 4.3.2.1 □

Proof of (2) ⇒ (3) Any independent dynamic is in particular mixing.

Proof of (3) ⇒ (1) The following proposition is a modification of a Schnorr's result [Sch71]. The proof we include here is taken from [GHR08a].

Proposition 4.3.2.1. *If the infinite binary string $\omega \in (\{0,1\}^{\mathbb{N}}, \lambda)$ is not Schnorr random (w.r. to the uniform measure), then there exists an isomorphism $\Phi : (\{0,1\}^{\mathbb{N}}, \lambda) \to (\{0,1\}^{\mathbb{N}}, \lambda)$ such that $\Phi(\omega)$ is not typical for the shift transformation σ.*

Proof. In order to prove this statement, we recall an equivalent definition of Schnorr-randomness.

Definition 4.3.2.4 (Martingale)**.** Let $(\Sigma^{\mathbb{N}}, \mu)$ be a Symbolic space with a probability distribution μ over it, as in section 2.2.2. A ***martingale*** for μ is a function $V : \Sigma^* \to \mathbb{R}^+$ with the property

$$\sum_{z\in\Sigma} \mu(xz)V(xz) = \mu(x)V(x).$$

It is a ***supermartingale*** if we have $\leq$ here. ⌟

The following inequality is well-known and easy to prove.

Proposition 4.3.2.2 (Martingale inequality)**.** *For any $\alpha > 0$ and any supermartingale V we have*

$$\{\omega : \exists n\, V(\omega_{[n]}) \geq \alpha V(\Lambda)\} \leq \alpha. \tag{4.4}$$

From now on we restrict our attention to the Cantor space $\{0,1\}^{\mathbb{N}}$ with the uniform measure λ. Then a martingale for λ is a function $V : 2^* \to \mathbb{R}^+$ with the property

$$\frac{1}{2}(V(x0) + V(x1)) = V(x).$$

For a string $x = x_1 x_2 \cdots \in \Sigma^* \cup \Sigma^{\mathbb{N}}$ let us denote $x_{1:n}$ just by $x_{[n]}$.

Definition 4.3.2.5. Let V be a computable supermartingale, and $f : \mathbb{N} \to \mathbb{N}$ an unbounded monotonic computable function. Define the set $\mathfrak{N}_{V,f}$ as the set of all sequences x with $\limsup_n V(x_{[n]})/f(n) > 0$. ⌟

It is easy to see that each set of the form $\mathfrak{N}_{V,f}$ has measure 0. Moreover, the following theorem is proved in [Sch71].

Proposition 4.3.2.3. *A set has the form $\mathfrak{N}_{V,f}$ for a martingale V if and only if there is a Schnorr test T such that the infinite strings failing T are exactly the elements of $\mathfrak{N}_{V,f}$.*

Proof. See the appendix. □

Let $\mathfrak{N}_{V,f}$ be given, and let $f' = \lfloor\sqrt{f}\rfloor$. Then $x \in \mathfrak{N}_{V,f}$ implies $V(x_{[n]}) > f'(n)$ for infinitely many n. Because of this, we will give yet another definition of (Schnorr-) constructive null set.

Let V be a computable martingale for λ and $f : \mathbb{N} \to \mathbb{N}$ an unbounded monotonic computable function with $f > 4$. We define the set $\mathfrak{N}'_{V,f}$ as the set of all sequences x with $V(x_{[n]}) > f(n)$ infinitely often.

It is obvious that the sets $\mathfrak{N}'_{V,f}$ are also just the null sets found by Schnorr tests.

Theorem 12.1 of Schnorr's book [Sch71] says that for each such set there is a measure preserving computable function $\Phi : \{0,1\}^{\mathbb{N}} \to \{0,1\}^{\mathbb{N}}$ such that for all $z \in \mathfrak{N}_{V,f}$ the value $\Phi(z)$ is non-typical. By definition, such Φ is given by $\Phi = \overline{\varphi}$ where $\varphi : 2^* \to 2^*$ is a monotonic computable function. In what follows we modify Schnorr's construction in such a way that $\overline{\varphi}$ has a computable inverse $\overline{\varphi}^{-1}$. In this case $\overline{\varphi}$ becomes an isomorphism between computable measureable spaces.

To prepare the construction of $\overline{\varphi}$, we need some more definitions. First, we define a series of tests using V, f, having more and more special properties.

Given our unbounded computable function $f : \mathbb{N} \to \mathbb{R}^+$, there is an unbounded strictly increasing recursive function $g : \mathbb{N} \to \mathbb{N}$ such that for all n we have

$$f(g(n)) > 2^{2n \log n}. \tag{4.5}$$

Let

$$\begin{aligned} U_n &= \{x \in \{0,1\}^{g(n)} : \max_{i \le g(n)} V(x_{[i]}) > 2^n\}, \\ U'_n &= \bigcap_{j=n}^{n \log n} U_j\{0,1\}^{\mathbb{N}}, \end{aligned}$$

then of course we have

$$U'_n \supseteq \{y \in \{0,1\}^{\mathbb{N}} : \max_{i \le g(n)} V(y_{[i]}) > 2^{n \log n}\}. \tag{4.6}$$

By the martingale inequality 4.4 we have

$$\lambda(U_n\{0,1\}^{\mathbb{N}}) \le 2^{-n}. \tag{4.7}$$

Claim 4.3.2.1. *If $y \in \mathfrak{N}'_{V,f}$, then there are infinitely many n with $y \in U'_n$.*

Proof. We have $V(y_{[i]}) > f(i)$ for infinitely many i. For such an i let n be such that $g(n-1) < i \le g(n)$, then noting $2(n-1)\log(n-1) \ge 2(n-1)(\log n - 1) \ge n \log n$ we have

$$\begin{aligned} V(y_{[i]}) > f(i) > f(g(n-1)) > 2^{2(n-1)\log(n-1)} \qquad & \text{by (4.5)} \\ > 2^{n \log n} \qquad & \end{aligned}$$

if n is sufficiently large (independently of y). From here we conclude by the inequality (4.6). □

In what follows we break up the sets U'_n into parts W'_i belonging to different prefixes. For each n let us define the following sets of integers:

$$L_n = \{i : n \le 3^i < 3^{i+1} \le n \log n\}.$$

Claim 4.3.2.2. *There is a computable function $s : \mathbb{N} \to \{0,1\}^*$ with the following properties.*

1. *The integers $|s_i| \le i$ form a monotonically increasing sequence with $\lim_i |s_i| = \infty$.*
2. *For each n the set of strings $\{s_i : i \in L_n\}$ is a covering set.*

The proof is easy. Now we modify our test sets further. Assume that a function $s : \mathbb{N} \to \{0,1\}^*$ is given satisfying the requirement in the claim. For every positive integer m let $i = \lfloor \log_3 m \rfloor$, and

$$W_m = s_i\{0,1\}^* \cap U_m = \{x \in U_m : x \sqsupseteq s_i\},$$
$$W'_i = \bigcap_{m=3^i}^{3^{i+1}-1} W_m\{0,1\}^{\mathbb{N}} = s_i\{0,1\}^{\mathbb{N}} \cap \bigcap_{m=3^i}^{3^{i+1}-1} U_m\{0,1\}^{\mathbb{N}}.$$

Claim 4.3.2.3. *We have $U'_n = \bigcup_{i \in L_n} W'_i$. Therefore $\omega \in \mathfrak{N}_{V,f}$ implies that there are infinitely many i with $\omega \in W'_i$.*

Proof. Since $\{s_i : i \in L_n\}$ is covering, for each $y \in U'_n$ there is a $i \in L_n$ with $y \in s_i\{0,1\}^{\mathbb{N}} \cap U'_n$. On the other hand $i \in L_n$ implies $n \le 3^i < 3^{i+1} - 1 < n \log n$, hence by its definition $U'_n \subseteq \bigcap_{m=3^i}^{3^{i+1}-1} U_m\{0,1\}^{\mathbb{N}}$. □

We define a measure-preserving invertible map $\overline{\varphi}$ via a monotonic measure-preserving computable function $\varphi : \{0,1\}^* \to \{0,1\}^*$ with $\varphi(\{0,1\}^{g(n)}) = \{0,1\}^n$. Suppose that φ has been defined up to $\{0,1\}^{g(n)}$, we define it for $\{0,1\}^{g(n+1)}$. Let $y \in \{0,1\}^n$, $D = \varphi^{-1}(y)$. Let $W = D\{0,1\}^{\mathbb{N}} \cap W_{n+1}\{0,1\}^{\mathbb{N}}$, then (4.7) implies $\lambda(W) \le 2^{-n-1}$. If $W \ne \emptyset$ then let $\varphi(W) = y1$.

Let $i = \lfloor \log_3(n+1) \rfloor$, then as we know, all elements of W share the prefix s_i. If all elements of D also share the prefix s_i then extend φ further on $D\{0,1\}^{g(n+1)-g(n)}$ to $y0$ or $y1$ in an arbitrary measure-preserving way. Otherwise let r be the first index $\le |s_i|$ such that there are strings $x', x'' \in D$ with $x'_r \ne x''_r$. For $j \in \{0,1\}$ let

$$D_j = \{x \in D : x_r = j\}.$$

By definition one of these sets contains W, without loss of generality assume $W \subseteq D_1$. For $j \in \{0,1\}$ we will define the sets $D'_j = \varphi^{-1}(yj)$. Of course we need $\lambda(D'_j) = 2^{-n-1}$, and we have already set $D'_1 \supseteq W$. Now, if $\lambda(D_1) \ge 2^{-n-1}$ then let $D'_1 \subseteq D_1$, otherwise let $D'_0 \subseteq D_0$. The further details of the choice of D'_j are arbitrary.

This completes the definition of φ. The measure preserving property is immediate from the definition. Let us observe another important property. The numbers $\lambda(D_j)/\lambda(D)$ have the form $p/2^q$ for some integers p, q with odd $p < 2^q$ and $1 \leq q \leq g(n)$. Denote $q(y) = q$. The definition of the extension gives

$$q(yj) = q(y) - 1. \tag{4.8}$$

Now we show that the image of a nonrandom string is not typical.

Claim 4.3.2.4. *If $\omega \in \mathfrak{N}_{V,f}$ then $\overline{\varphi}(\omega)$ is not typical.*

Proof. Suppose $\omega \in \mathfrak{N}_{V,f}$, and let $\eta = \overline{\varphi}(\omega)$, then there are infinitely many indices i with $\omega \in W_i'$. Let i be such, this implies $\omega \in W_m\{0,1\}^{\mathbb{N}}$ for $3^i \leq m < 3^{i+1}$. The construction gives $\eta_m = 1$ for $3^i \leq m < 3^{i+1}$. Since this is true for infinitely many i, the sequence η is not typical. □

To show that $\overline{\varphi}$ is invertible, we will find for each k a value $n = n(k)$ such that $x_k' \neq x_k''$ implies $\varphi(x'_{[g(n)]}) \neq \varphi(x''_{[g(n)]})$. We define $n(k)$ recursively via

$$\begin{aligned} n(0) &= 1, \\ n(k+1) &= n(k) + g(n(k)). \end{aligned}$$

Claim 4.3.2.5. *Let $x', x'' \in \{0,1\}^{\mathbb{N}}$ be two different sequences. For all $k \geq 1$ with $n = n(k)$, the relation $x_k' \neq x_k''$ implies $\varphi(x'_{[g(n)]}) \neq \varphi(x''_{[g(n)]})$.*

Proof. Let $y = \overline{\varphi}(x')$, and let $k \geq 1$ be the first place where $x_k' \neq x_k''$. For $m = n(k-1)$ consider the map φ on $\{0,1\}^{g(m)}$. If $y_{[m]} = \varphi(x'_{[g(m)]}) \neq \varphi(x''_{[g(m)]})$ then we are done, suppose this is not the case. Let $D = \varphi^{-1}(y_{[m]})$. By the choice of m, all elements of D share the prefix $x'_{[k-1]}$. The definition above extends φ to $\{0,1\}^{g(m+1)}$. If $y_{[m+1]} = \varphi(x'_{[g(m+1)]}) \neq \varphi(x''_{[g(m+1)]})$ then we are done. Otherwise relation (4.8) implies $q(y_{[m+1]}) < q(y_{[m]})$. Therefore by repeating the extension we must get $\varphi(x'_{[g(i)]}) \neq \varphi(x''_{[g(i)]})$ for some $i < m + g(m)$ before getting to $q(y_{[i]}) = 0$. □

This completes the proof of Proposition 4.3.2.1. □

Now we are able to finish the proof of our main result:

Theorem 4.3.2.3 ((3) ⇒ (1))**.** *Let $(\mathcal{X}, \mu)$ be a CPS whith no atoms. If a point x is T-typical for any independent endomorphism T then x is Schnorr random.*

Proof. Suppose that x is not Schnorr random. We construct a dynamic T for which x is not T-typical. From Corollary 2.3.1.4 we know that there is an isomorphism $\phi : (\mathcal{X}, \mu) \to (C, \lambda)$. If $x \notin \mathrm{dom}(\phi)$ then we can take any independent dynamic and modify it in order to be the identity on x. It is clearly still an independent dynamic (maybe with a smaller domain of computability) and x, being a fixed point, can't be T-typical. So let $x \in \mathrm{dom}(\phi)$. Then $\phi(x)$ is not Schnorr random in (C, λ), since ϕ as well as its inverse preserve Schnorr randomness. Then, by Proposition 4.3.2.1, $\varphi(\phi(x))$ is not σ-typical, where σ is the shift which is clearly independent (cylinders being the essential events). Put $\psi = \varphi \circ \phi$. Define the dynamic T on X by $T = \psi^{-1} \circ \sigma \circ \psi$. It is easy to see that T is independent for events of the form $E = \psi^{-1}[w]$. Since $\{\psi^{-1}[w] : w \in \{0,1\}^*\}$ form an essential family of almost decidable events, T is independent. As $\psi(x)$ is not σ-typical, x is not T-typical either. □

4.4 Entropy and orbit complexity

Let (X, T, μ) be an ergodic dynamical system and $\xi = \{C_1, \ldots, C_N\}$ be a finite measurable partition of X. Let $T^{-1}\xi$ be the partition given by the pre-images $T^{-1}C_i$. Then let

$$\xi_n = \xi \vee T^{-1}\xi \vee T^{-2}\xi \vee \ldots \vee T^{-n}\xi$$

be the partition given by the sets of the form

$$C_{i_0} \cap T^{-1}C_{i_1} \cap \ldots \cap T^{-n+1}C_{i_{n-1}},$$

varying C_{i_j} among all the sets of ξ. Knowing to which atom ξ_n a point x belongs to (that is to perform an observation) corresponds to knowing the atoms of the partition ξ that the orbit of x visits up to time $n-1$.

4.4.1 Measure-theoretical entropy

The notion of measure-theoretical entropy was introduced by Kolmogorov as an indicator of "how random a dynamical system is". It can be thought as the rate (per unit time) of gained information (or removed uncertainty) when observations of the type "$T^n(x) \in C_i$" are performed.

The notion of information used in the original definition was *Shannon information* but others notions of information, as Kolmogorov's one, could also be used in principle.

We now recall the classical definition and some basic tools. Then we present the definition obtained using Kolmogorov's notion: algorithmic information.

Entropy with Shannon information

Let us consider the ***Shannon information function*** relative to the partition ξ_n (the information which is gained by observing that $x \in \xi_n(x)$):

$$I_\mu(x|\xi_n) := -\log \mu\xi_n(x)$$

and its mean, the entropy of the partition ξ_n:

$$H_\mu(\xi_n) := \int_X I_\mu(.|\xi_n)d\mu = \sum_{C\in\xi_n} -\mu(C)\log\mu(C)$$

The measure theoretic or Kolmogorov-Sinaï entropy of T relative to the partition ξ is defined as:

$$h_\mu(T,\xi) = \lim_{n\to\infty} \frac{1}{n} H_\mu(\xi_n)$$

(which exists and is an infimum, since the sequence $(H_\mu(\xi_n))_n$ is sub-additive). Now we take the supremum over all finite partitions in order to suppress the *partition-dependency*: the Kolmogorov-Sinaï entropy of (X,T,μ) is defined as:

$$h_\mu(T) = \sup\{h_\mu(T,\xi) : \xi \text{ finite measurable partition}\}.$$

With the Shannon information function, it is possible to define a kind of pointwise notion of entropy with respect to a partition ξ:

$$\limsup_n \frac{1}{n} I_\mu(x|\xi_n).$$

This notion, by the celebrated Shannon-MacMillan-Breiman theorem equals $h_\mu(T,\xi)$ for μ-almost every point (moreover, the limit exists almost every where). We recall the following two results that we will need later. The first proposition follows directly from the definitions.

Proposition 4.4.1.1. *If (X_ξ,σ,μ_ξ) is the symbolic model associated to (X,T,μ,ξ) then $h_\mu(T,\xi) = h_{\mu_\xi}(\sigma)$.*

The next proposition is taken from [Pet83]:

Proposition 4.4.1.2. *If $(\xi_i)_{i\in\mathbb{N}}$ is a family of finite partitions which generates the Borel σ-algebra up to sets of measure 0, then $h_\mu(T) = \sup_i h_\mu(T, \xi_1\vee, ..., \vee\xi_i)$.*

Entropy with Kolmogorov information

Now we show in a simple way how with the Kolmogorov information content it is possible to give a definition equivalent to the above one when the measure is computable. An atom C of the partition ξ_n can be seen as a word of length n on the alphabet ξ, which allows one to consider its Kolmogorov complexity $K(C)$.

We then define the ***Kolmogorov information function*** (which is independent of μ) relative to the partition ξ_n:

$$\mathcal{I}(x|\xi_n) := K(\xi_n(x))$$

and its mean, which can be called the algorithmic entropy of the partition ξ_n:

$$\mathcal{H}_\mu(\xi_n) := \int_X \mathcal{I}(.|\xi_n)d\mu = \sum_{C\in\xi_n} \mu(C)K(C)$$

When μ is computable and ξ is a computable partition, by propositions 4.2.0.4 and 3.3.3.1 every μ-random point x is mapped to a random sequence $\phi_\xi(x)$. By theorem A.3.0.9, there is a constant c such that for every μ-random point x and every n it holds:

$$I_\mu(x|\xi_n) - d_\mu(\phi_\xi(x)) < \mathcal{I}(x|\xi_n) < I_\mu(x|\xi_n) + K(n) + c \tag{4.9}$$

(where d_μ is the deficiency of randomness, see theorem A.3.0.9), and since random points have measure one, one also has:

$$H_\mu(\xi_n) - 1 \leq \mathcal{H}_\mu(\xi_n) \leq H_\mu(\xi_n) + K(n) + c$$

so we obtain an easy proof that the measure theoretic entropy with respect to a computable partition can be obtained also using Kolmogorov information:

$$h_\mu(T|\xi) = \lim_n \frac{H_\mu(\xi_n)}{n} = \lim_n \frac{\mathcal{H}_\mu(\xi_n)}{n}$$

which in turns, since computable partitions contains a generating family (corollary 4.2.0.1), offers an equivalent definition of Kolmogorov-Sinai entropy:

$$h_\mu(T) = \sup\{\lim_n \frac{\mathcal{H}_\mu(\xi_n)}{n} : \xi \text{ finite computable partition}\}.$$

With the Kolmogorov information function, it is also possible to define a pointwise notion of algorithmic entropy with respect to a partition ξ:

$$K^{sym}(x, T|\xi) := \limsup_n \frac{1}{n}\mathcal{I}(x|\xi_n).$$

In [Bru83], Brudno has shown:

Proposition 4.4.1.3. *$K^{sym}(x, T|\xi) = h_\mu(T, \xi)$ for μ-almost every point.*

Actually, one of his goals was to define (using Kolmogorov complexity) a notion of entropy for a single orbit. A natural choice was to use $K^{sym}(x, T|\xi)$ and take supremum over all partitions to have an absolute notion, but Brudno showed that this supremum could be infinite for a "large" set of points. So, he proposed a somewhat different definition: orbit complexity.

4.4.2 Orbit complexity

Since Kolmogorov complexity is defined and meaningfull for a single symbolic sequence it seems that with this notion it is possible to define entropy of a single orbit. This problem was firstly considered by Brudno in [Bru83].

Brudno showed that if orbits are coded using finite partitions, its complexity may become infinite when we take supremum over all partitions (this is true for every non eventually periodic orbit), so we can't easily suppress the *partition-dependency* in the definition.

To solve the problem, Brudno coded orbits using finite open covers, which yields to a proper definition in the compact case. He proved that if the space is compact, the orbit complexity so defined equals the metric entropy of the system almost everywhere. If the space is non compact, the definition of Brudno gives infinite complexity to any orbit which has "many" isolated points (translations on the real line for instance). In [Gal00] another solution is proposed: on effective metric spaces, orbit complexity is defined using shadowing sequences of ideal points, which can be seen as words in a canonical way. This definition make sense in the non compact case (but no relation with entropy is proved) and, when X is compact, it coincide with Brudno's definition for *each* point. Let us hence introduce the classical notion of orbit complexity by this equivalent definition:

Shadowing orbit complexity

Let T be an endomorphism of the CPS $(\mathcal{X}, \mu)$. For each point x we define the *shadowing orbit complexity* of x under T, denoted $K^{shad}(x, T)$, which quantifies in some sense the algorithmic information needed to describe the orbit of x with finite but arbitrarily accurate precision.

Given $\epsilon > 0$ and $n \in \mathbb{N}$, the algorithmic information needed to describe the n first iterates of x up to ϵ is:

$$K_n^{shad}(x,T,\epsilon) := \min\{K(i) : d(T^j s_i, T^j x) < \epsilon \quad \forall j < n\}$$

where $K(i)$ is the self-delimiting complexity of i (actually, of its binary expansion).

We then define the maximal growth-rate of this information:

$$K^{shad}(x,T,\epsilon) := \limsup_{n\to\infty} \frac{1}{n} K_n^{shad}(x,T,\epsilon)$$

As ϵ tends to 0, this quantity increases, hence it has a limit (which can be infinite).

Definition 4.4.2.1. The ***shadowing orbit complexity*** of x under T is defined by:

$$K^{shad}(x,T) := \lim_{\epsilon\to 0^+} K^{shad}(x,T,\epsilon)$$

⌟

We remark that the n first iterates of x could be ϵ-shadowed by a pseudo-orbit of n ideal points instead of the orbit of a single ideal point. Actually it is easy to see that it gives the same quantity[3]:

$$\begin{aligned} K_n^{shad}(x,T,2\epsilon) &\overset{+}{\leq} \min\{K(\langle i_0,\ldots,i_{n-1}\rangle) : d(s_{i_j}, T^j x) < \epsilon \quad \forall j < n\} \\ &\overset{+}{\leq} K_n^{shad}(x,T,\epsilon/2) + K(n) \end{aligned} \tag{4.10}$$

Indeed, from $\langle i_0,\ldots,i_{n-1}\rangle$ some ideal point can be algorithmically found in $B(s_{i_0},\epsilon)\cap \ldots \cap T^{-(n-1)}B(s_{i_{n-1}},\epsilon)$ which is a r.e open set, uniformly in $\langle i_0,\ldots,i_{n-1}\rangle$. Its n first iterates 2ϵ-shadow the orbit of x, which proves the first inequality. For the second inequality, some $\langle i_0,\ldots,i_{n-1}\rangle$ can be algorithmically constructed from n and a point s_i corresponding to $K_n^{shad}(x,T,\epsilon/2)$, taking any $s_{i_j} \in B(T^j s_i,\epsilon)$.

Symbolic orbit complexity

We now show that in the computable framework, finite partitions does lead to a proper definition of orbit complexity in a very simple way (moreover, in this approach compactness

[3] In the following, $\overset{+}{\leq}$ stands for inequality up to a constant which depends only on T

is not important). We will see that this definition is equivalent to the above one at each random point. This will allow us to prove in the next subsection that for *each* random point of the space the orbit complexity equals the entropy.

Let T be an endomorphism of the computable probability space (X, μ).

Definition 4.4.2.2. The ***symbolic orbit complexity*** of x under T is:

$$K^{sym}(x, T) := \sup_{\xi \text{ computable partition}} K^{sym}(x, T|\xi)$$

⌟

Since computable partitions *are a countable collection* which, in addition, contains a generating family (see corollary 4.2.0.1), we can apply propositions 4.4.1.3 and 4.4.1.2 to obtain:

Proposition 4.4.2.1. $K^{sym}(x, T) = h_\mu(T)$ *for μ-almost every point*

which can be seen as an extension of Brudno's main result to the non compact case.

Equivalence of the two orbit complexities for random points

Now we prove that the notions of $K^{shad}(x, T)$ and $K^{sym}(x, T)$ coincide at every random point.

Theorem 4.4.2.1. *Let $(\mathcal{X}, \mu)$ be a compact computable probability space. Then for every μ-random point x:*

$$K^{shad}(x, T) = K^{sym}(x, T),$$

for any ergodic endomorphism T.

The difficult inequality is $K^{sym}(x, T) \leq K^{shad}(x, T)$. To prove it, we need the following lemma.

First remind that for all natural number $k \geq 1$, the self-delimiting complexity of its binary expansion $\overline{k}$ satisfies:

$$K(\overline{k}) \overset{+}{\leq} J(k)$$

where $J(x) = \log x + 1 + 2\log(\log x + 1)$ for all $x \in \mathbb{R}, x \geq 1$. Notice that J is a concave increasing function and that $x \mapsto xJ(1/x)$ is an increasing function on $]0, 1]$ and tends to 0 as $x \to 0$.

Also remind that for all finite sequence of strings $(x_1, \ldots, x_n)$, one has

$$K(x_1, \ldots, x_n) \overset{+}{\leq} K(x_1) + \ldots + K(x_n)$$

(this is one reason to use the self-delimiting complexity instead of the plain complexity).

Lemma 4.4.2.1. *Let Σ be a finite alphabet and $n \in \mathbb{N}$. Let $u, v \in \Sigma^n$ and $0 < \alpha < 1/2$ such that the density of the set of positions where u and v differ is less than α, that is:*

$$\frac{1}{n}\#\{i < n : u_i \neq v_i\} < \alpha < 1/2$$

Then $\left|\frac{1}{n}K(u) - \frac{1}{n}K(v)\right| < \alpha J(1/\alpha) + \frac{c}{n}$ where c is a constant independent of u, v and n.

Proof. let $(i_1, \ldots, i_p)$ be the ordered sequence of indices where u and v differ. By hypothesis, $p/n < \alpha$. Put $i_0 = 0$ and $k_j = i_j - i_{j-1}$ for $1 \leq j \leq p$.

We now show that u can be recovered from v and roughly $\alpha J(1/\alpha)n$ bits more. Indeed u can be computed from $(v, \overline{k_1}, \ldots, \overline{k_p})$, constructing the string which coincides with v everywhere but at positions $k_1, k_1 + k_2, \ldots, k_1 + \ldots k_p$. Then $K(u) \overset{+}{\leq} K(v) + K(\overline{k_1}) + \ldots + K(\overline{k_p}) \overset{+}{\leq} K(v) + J(k_1) + \ldots + J(k_p)$.

Now, as J is a concave increasing function, one has:

$$\frac{1}{p}\sum_{j \leq p} J(k_j) \leq J\left(\frac{1}{p}\sum_{j \leq p} k_j\right) = J\left(\frac{i_p}{p}\right) \leq J\left(\frac{n}{p}\right)$$

As a result,

$$\frac{1}{n}K(u) \leq \frac{1}{n}K(v) + \frac{p}{n}J\left(\frac{n}{p}\right) + \frac{c}{n}$$

where c is some constant independent of u, v, n, p.

As $x \mapsto xJ(1/x)$ is increasing for $x \leq 1/2$ and $p/n < \alpha < 1/2$, one has:

$$\frac{1}{n}K(u) \leq \frac{1}{n}K(v) + \alpha J(1/\alpha) + \frac{c}{n}$$

Switching u and v gives the result (c may be changed).

$\square$

We are now able to prove the theorem.

Proof. (of thm. 4.4.2.1) $K^{shad}(x,T) \leq K^{sym}(x,T)$: let $\epsilon > 0$. Choose a computable partition $\mathcal{P}$ of diameter $< \epsilon$ (this is why we require X to be compact). To every cell of $\mathcal{P}$, associate an ideal point which is inside. This finite dictionary can be encoded into a finite string (which could also be algorithmically generated from $\mathcal{P}$ which consists of open r.e. cells, but we do not need that). Then the translation through this finite dictionary is computable, and transforms the symbolic orbit of x into a sequence of ideal points ϵ-close to the orbit of x. So $K^{shad}(x,T) \leq K^{sym}(x,T)$.

For the other inequality, fix some computable partition $\mathcal{P}$. We show that for any $\beta > 0$ there is some $\epsilon > 0$ such that for every μ-random point x, $K^{sym}(x,T|\mathcal{P}) \leq K^{shad}(x,T,\epsilon)+\beta$. As $K^{shad}(x,T,\epsilon)$ increases as $\epsilon \to 0^+$ and β is arbitrary, the inequality follows.

First take $\alpha < 1/2$ such that $\alpha J(1/\alpha) < \beta$, and remark that

$$\lim_{\epsilon \to 0^+} \mu\left(\overline{(\partial\mathcal{P})^\epsilon}\right) = \mu(\partial\mathcal{P}) = 0$$

Hence there is some ϵ such that $\mu\left(\overline{(\partial\mathcal{P})^\epsilon}\right) < \alpha$. From a sequence of ideal points we will reconstruct the symbolic orbit of a μ-random point with a density of errors less than α. Lemma 4.4.2.1 will then allow to conclude.

We define an algorithm $\mathcal{A}(\epsilon, n, i)$ with $\epsilon \in \mathbb{Q}_{>0}$ and $n, i \in \mathbb{N}$ which outputs a word $a_0 \ldots a_{n-1}$ on the alphabet $\mathcal{P}$. To compute a_j, $\mathcal{A}$ semi-decides in a dovetail picture:

- $T^j s_i \in C$ for every $C \in \mathcal{P}$,
- $s \in C$ for every $s \in B(T^j s_i, \epsilon/2)$ and every $C \in \mathcal{P}$.

The first test which stops provides some $C \in \mathcal{P}$: put $a_j = C$.

Let x be a μ-random point and s_i an ideal point whose orbit $\epsilon/2$-shadows the first n iterates of x. We claim that $\mathcal{A}$ will halt. Indeed, if $T^j s_i$ is in no $C \in \mathcal{P}$, as $T^j x$ is a random point it belongs to some $C \in \mathcal{P}$, so $T^j x \in C \cap B(T^j s_i, \epsilon/2)$ which is an open set and then contains at least an ideal point s, which will be eventually dealt with.

We now compare the symbolic orbit of x with the symbolic sequence computed by $\mathcal{A}$. A discrepancy at rank j can appear only if $T^j x \in (\partial\mathcal{P})^\epsilon$. Indeed, if this is not the case then $B(T^j x, \epsilon) \subseteq C$ where C is the cell $T^j x$ belongs to. As $d(T^j s_i, T^j x) < \epsilon/2$, $B(T^j s_i, \epsilon/2) \subseteq B(x, \epsilon) \subseteq C$, so the algorithm gives the right cell.

Now, using the Birkhoff ergodic theorem for random points (theorem 4.3.2.1), the following holds for any μ-random point x:

$$\limsup_{n\to\infty} \frac{1}{n}\#\{j < n : T^j x \in (\partial\mathcal{P})^\epsilon\} \leq \mu(\overline{(\partial\mathcal{P})^\epsilon}) < \alpha$$

so there is some n_0 such that for all $n \geq n_0$, $\frac{1}{n}\#\{j < n : T^j x \in (\partial\mathcal{P})^\epsilon\} < \alpha$. This implies that for all $n \geq n_0$, taking s_i an ideal point whose orbit ϵ-shadows the first n iterates of x and with minimal complexity, the algorithm $\mathcal{A}(\epsilon, n, i)$ produces a symbolic string u which differs from the symbolic orbit v of x of length n with a density of errors $< \alpha$. Applying lemma 4.4.2.1 gives:

$$\begin{aligned} \frac{1}{n}K(\varphi_{\mathcal{P}}(x)_{<n}) = \frac{1}{n}K(v) &\leq \frac{1}{n}K(u) + \alpha J(1/\alpha) + \frac{c}{n} \\ &\leq \frac{1}{n}\left(K_n^{shad}(x, T, \epsilon) + K(\epsilon) + K(n) + c'\right) + \beta + \frac{c}{n} \end{aligned}$$

where c' depends on ϵ.

Taking the lim sup as $n \to \infty$ gives:

$$K^{sym}(x, T|\mathcal{P}) \leq K^{shad}(x, T, \epsilon) + \beta$$

□

4.4.3 Orbit complexity of random points equals the entropy of the measure

In [V'y97], the following result is proved:

Theorem 4.4.3.1 (V'yugin)**.** *Let μ be a computable shift-invariant ergodic measure on $\Sigma^{\mathbb{N}}$. Then, for any Martin-Löf μ-random sequence ω,*

$$\limsup_{n\to\infty} -\frac{1}{n}\log\mu([\omega_{1:n}]) = h_\mu(\sigma).$$

Which corresponds to the Shannon-McMillan-Breiman theorem for individual random sequences on the space of symbolic sequences. Using the tools developed in section 4.2, we easily prove the general version, on any computable probability space.

Theorem 4.4.3.2. *Let ξ be a computable partition on the computable probability space (X, μ). If x is μ-random then for every ergodic endomorphism T:*

$$\limsup_{n\to\infty} I_\mu(x|\xi_n) = h_\mu(T, \xi)$$

where $I_\mu(x|\xi_n)$ is the Shannon function information (see section 4.4.1).

Proof. Since ξ is computable, by proposition 4.2.0.4, the symbolic model (X_ξ, σ, μ_ξ) is effective. Hence, proposition 3.3.3.1 implies that $\varphi(x)$ is μ_ξ-random in X_ξ. By definition, $I_\mu(x|\xi_n) = -\frac{1}{n}\log \mu_\xi([\varphi(x)_{1:n}])$. Then the result follows from theorem 4.4.3.1 and proposition 4.4.1.1. □

We finish by showing how this theorem allows to easily strengthen proposition 4.4.2.1, proving that it holds for each μ-random point:

Theorem 4.4.3.3. *Let $(\mathcal{X}, \mu)$ be a (not necessarily compact) computable probability space. Then for each random point $x \in X$:*

$$K^{sym}(x, T) = h_\mu(T),$$

for every ergodic endomorphism T.

Proof. Let ξ be a computable partition. By the Levin-Chaitin theorem, for each μ-random point x, there is a constant c such that for all n:

$$I_\mu(x|\xi_n) - c < \mathcal{I}(x|\xi_n) < I_\mu(x|\xi_n) + K(n) + c$$

so

$$\limsup_{n\to\infty} \frac{\mathcal{I}(x|\xi_n)}{n} = \limsup_{n\to\infty} \frac{I_\mu(x|\xi_n)}{n}$$

Hence, by theorem 4.4.3.2, $K^{sym}(x, T|\xi) = h_\mu(T, \xi)$. Since the collection of all computable partitions generates the Borel σ-algebra (see corollary 4.2.0.1), proposition 4.4.1.2 proves the theorem. □

Remark 4.4.3.1. In ergodic theory, the *variational principle* is an important result stating that the supremum of the metric entropy over all invariant measures equals the *topological* entropy of the system. It is a very interesting fact that the supremum of the shadowing orbit complexity over all $x \in X$ equals the topological entropy too. And this can be proved without the help of the variational principle. The interested reader can find all the details of this in [Hoy08]. ⌟

Chapter 5

Computable "pseudo-random" points

In this chapter we develop a general probabilistic method allowing to construct computable points verifying a given statistical law. More precisely, it applies to those probabilistic laws that can be put into a Schnorr's test form. The existence of computable points verifying this law will be a consequence of the following result (Thm. 5.1.0.4):

Theorem *Let $(T_n)_n$ be a uniform sequence of Sch-test. Then the set of points passing all these tests contains a dense set of computable points.*

We then apply this to construct computable reals which are absolutely normal, and to computable initial conditions which follow the statistical behaviour of uniform families of dynamical systems. Finally we study the problem of the computability of the invariant measure for some particular classes of transformations.

5.1 Computable points passing schnorr tests

In this section, $\mathrm{Comp}(X)$ will denote the set of computable points of X. We remark that this is an invariant set for any computable transformation T.

Definition 5.1.0.1. A set $A \subset X$ is said to be recursively closed if, denoting B_i the ideal ball of number i, the set $\{i \in \mathbb{N} : B_i \cap A \neq \emptyset\}$ is r.e. ⌟

Proposition 5.1.0.1. *Let A be a recursively closed subset of X. Then* $\mathrm{Comp}(X) \cap A$ *is dense in A. That is,* $\overline{\mathrm{Comp}(X) \cap A} = A$.

Proof. Since $\{i \in \mathbb{N} : B_i = (B(s_{n_i}, q_{n_i}) \cap A \neq \emptyset\}$ is r.e, given some ideal ball $B = B(s, q)$ intersecting A, the set $\{k \in \mathbb{N} : \overline{B_k} \subset B, q_{n_k} \leq 2^{-k}, B_i \cap A \neq \emptyset\}$ is also r.e. Then we can effectively construct an exponentially decreasing sequence of ideal balls intersecting A. Hence $\{x\} = \cap_k B_k$ is a computable point lying in A.

□

The following theorem will be our main tool:

Theorem 5.1.0.4. *Let $(T_n)_n$ be a uniform sequence of Sch-test. Then the set of points passing all these tests contains a dense set of computable points.*

Proof. We shall use the following lemma:

Lemma 5.1.0.1. *Let (X, μ) be a computable probability space. Suppose the measure of a constructively open set A satisfies:*

1. *$\mu(A) < 1$ and,*

2. *$\mu(A)$ is computable.*

Then there is a computable point $x \notin A$.

Proof. Since $\mu(A)$ is computable, so are μ_A and $\mu_{X \setminus A}$ (proposition 2.3.1.1). Observe that in general, given an ideal ball B and a computable measure ν, the relation $\nu(B) > 0$ is semi-decidable. Hence, the support of the measure is a recursively closed set. By proposition 5.1.0.1, the set of computable points is dense in this support, in particular in $\mathrm{supp}(\mu_{X \setminus A})$. □

Let $T = (C_n)_n$ be a Sch-test. Given an almost decidable ball B of positive measure, we can effectively find n such that $\mu(C_n) < \mu(B)$. By lemma 5.1.0.1, there is a computable point in $B \setminus C_n$. The result then follows from proposition 3.3.2.1. □

5.2 Computable typical points

Before to enter in the main theme of typical statistical behaviors let us see an easier topological result in this line. One of the features of undecomposable (topologically transitive) chaotic systems is that there are many dense orbits, the following shows that if the system is computable then there are computable dense orbits.

We remark that this result can also be obtained as a corollary of the effective Baire theorem [Bra01].

Theorem 5.2.0.5. *Let $\mathcal{X}$ be a computable complete metric space and $T : X \to X$ a transformation which is computable on a dense r.e open set. If T has a dense orbit, then it has a computable one which is dense.*

In other words, there is a computable point $x \in X$ whose orbit is dense in X. Actually, the proof is an algorithm which takes an ideal ball as input and computes a transitive point lying in this ball.

Proof. $(B_i)_{i\in\mathbb{N}}$ being an enumeration of all ideal balls, define $U_i = dom(f) \cap \bigcup_n T^{-n}B_i$ which is r.e uniformly in i. By hypothesis, U_i is also dense. $\bigcap_i U_i$ is the set of transitive points. From any ideal ball $B(s_0, r_0)$ we effectively construct a computable point in $B(s_0, r_0) \cap \bigcap_i U_i$.

If $B(s_i, r_i)$ has been constructed, as U_i is dense $B(s_i, r_i) \cap U_i$ is a non-empty constructively open set, so an ideal ball $B(s, r) \subseteq B(s_i, r_i) \cap U_i$ can be effectively found (any of them can be chosen, for instance the first coming in the enumeration of the r.e. set). We then set $B(s_{i+1}, r_{i+1}) := B(s, r/2)$.

The sequence of balls computed satisfies:

$$\overline{B}(s_{i+1}, r_{i+1}) \subseteq B(s_i, r_i) \cap U_0 \cap \ldots \cap U_i$$

As $(r_i)_{i\in\mathbb{N}}$ is a decreasing computable sequence converging to 0 and the space is complete, $(s_i)_{i\in\mathbb{N}}$ converges effectively to a computable point x. Then $\{x\} = \bigcap_i B(s_i, r_i) \subseteq \bigcap_i U_i$. □

We will use the results from the previous section to prove that computable typical points exist for a large class of dynamical systems.

Theorem 5.2.0.6. *If T has polynomial decay of correlations w.r. to a μ-almost computable observable ϕ, then the set of points which are typical w.r to ϕ contains a dense set of computable points.*

Proof. First observe that, by lemma 3.3.3.1 and proposition 3.3.2.1 there is a strong effective null-set outside of which everything is well defined. Then apply theorem 5.1.0.4 to the above test coupled with the test given by theorem 4.3.2.2 □

Remark 5.2.0.2. In the proof we see that the Sch-test depends in an effective way on T, ϕ and $c_{\phi,\phi}$. This gives the possibility to operate in a way to apply proposition 3.3.2.1 to find computable points contained in the set of points typical with respect to uniform families ϕ_i, T_i. From now and beyond, we will assume that all test to be constructed are coupled with the test outside of which everything is well defined. ⌟

5.2.1 Application: computable absolutely normal numbers

An absolutely normal (or just normal) number is, roughly speaking, a real number whose digits (in every base) show a uniform distribution, with all digits being equally likely, all pairs of digits equally likely, all triplets of digits equally likely, etc.

While a general, probabilistic proof can be given that *almost all* numbers are normal, this proof is not constructive and only very few concrete numbers have been shown to be normal. It is for instance widely believed that the numbers $\sqrt{2}$, π and e are normal, but a proof remains elusive. The first example of an absolutely normal number was given by Sierpinski in 1916, twenty years before the concept of computability was formalized. Its construction is quite complicate and is a priori unclear whether his number is computable or not. In [BF02] a recursive reformulation of Sierpinski's construction (equally complicate) was given, furnishing a computable absolutely normal number.

As an application of theorem 5.2.0.6 we give a simple proof that computable absolutely normal numbers are dense in $[0, 1]$.

Let b be an integer ≥ 2, and X_b the space of infinite sequences on the alphabet $\Sigma_b = \{0, \ldots, b-1\}$. Let $T = \sigma$ be the shift transformation on X_b, and λ be the uniform measure. A real number $r \in [0, 1]$ is said to be absolutely normal if for all $b \geq 2$, its b-ary expansion $r_b \in X_b$ satisfies:

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} 1_{[w]} \circ \sigma^i(r_b) = \frac{1}{b^{|w|}} \qquad \text{for all } w \in \Sigma_b^*. \tag{5.1}$$

Theorem 5.2.1.1. *The set of computable reals which are absolutely normal is dense in* $[0, 1]$.

Proof. for each base $b \geq 2$, consider the transformation $T_b : [0,1] \to [0,1]$ defined by $T_b(x) = bx(\text{mod } 1)$. The Lebesgue measure λ is T_b-invariant and ergodic. The intervals $[k/b, (k+1)/b[$ $(k = 0, 1, ...b-1)$ induces the symbolic model $(\Sigma_b^{\mathbb{N}}, \sigma, \lambda)$, which is isomorphic to $([0,1], T_b, \lambda)$: the interval $[k/b, (k+1)/b[$ is represented by $k \in \Sigma_b$. For any word $w \in \Sigma_b$ define $I(w) = [0.w, 0.w + b^{-|w|}] \subset [0,1]$ to be the corresponding interval. Defining $\text{dom}(T_b) := [0,1] \setminus \{\frac{k}{b} : 0 \leq k \leq b\}$ (the interior of the partition) makes T_b a λ-almost computable transformation. The observable $f_w := 1_{I(w)}$ is also λ-almost computable, with *dom* $f_w = [0,1] \setminus \partial I(w)$. Note that given $r \in [0,1]$, equation 5.1 is equivalent to

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} f_w \circ T_b^i(r) = \frac{1}{b^{|w|}} \qquad \text{for all } w \in \Sigma_b^*.$$

Since $1_{[w]} \circ \sigma^n$ and $1_{[w]}$ are independent for $n > |w|$, so are $f_w \circ T_b^n$ and f_w. Hence Theorem 5.2.0.6 applies to $([0,1], T_b, \lambda)$ and f_w. Therefore, the set of points (for the system (T_b, λ)) which are not typical w.r.t the observable f_w fail a Sch-test $S_{b,w}$. Furthermore, $S_{b,w}$ is effective *uniformly* in $b, w \in \Sigma_b$. Hence, by proposition 3.3.2.1 and corollary 5.1.0.4, the set of absolutely normal numbers, contains a dense set of computable points. □

5.3 Computing some ergodic "physical" measures

We will see that in a large class of dynamical systems which have a single *physically relevant* invariant measure, the computability of this measure and related c_{g_i}, for observables in $\mathcal{H}$ can be proved, hence we can apply Thm. 5.2.0.6 to find pseudorandom points in such systems.

5.3.1 Physical measures

In general, given (X, T) there could be infinitely many invariant measures (this is true even if we restrict to probability measures). Among this class of measures, some of them are particularly important. Suppose that we observe the behavior of the system (X, T) trough a class of continuous functions $f_i : X \to \mathbb{R}$. We are interested in the statistical behavior of f_i along typical orbits of the system. Let us suppose that the time average along the orbit of x exists

$$A_x(f_i) = \lim_{n\to\infty} \frac{1}{n} \sum f_i(T^n(x))$$

this is a real number for each f_i. Moreover $A_x(f_i)$ is linear and continuous with respect to small changes of f_i in the sup norm. Then the orbit of x acts as a measure μ_x and $A_x(f_i) = \int f_i d\mu_x$ (moreover this measure is also invariant for T). This measure is physically interesting if it is given by a "large" set of initial conditions. This set will be called the basin of the measure. If X is a manifold, it is said that an invariant measure is *physical* (or SRB from the names of Sinai, Ruelle and Bowen) if its basin has positive Lebesgue measure (see [You02] for a survey and more precise definitions).

In what follows we will consider SRB measures in the classes of systems listed below,

1. The class of uniformly hyperbolic system on submanifolds of $\mathbb{R}^n$.
2. The class of piecewise expanding maps on the interval.
3. The class of *Manneville-Pomeau* type maps (non uniformly expanding with an indifferent fixed point).

All these systems, which are rather well understood, have a unique physical measure with respect to which correlations decay is at least polynomial. Furthermore, in each case, the corresponding constants can be estimated for functions in $\mathcal{F}$. The computability of the physical measures is proved case by case, but it is always a consequence of the fact that, in one way or another, the physical measure is "approached" by iterates of the Lebesgue measure at a known speed.

5.3.2 Uniformly hyperbolic systems

To talk about SRB measures on a system whose phase space is a manifold, we have to introduce the Lebesgue measure on a manifold and check that it is computable.

Computable manifolds and the Lebesgue measure

For simplicity we will not consider general manifolds but submanifolds of $\mathbb{R}^n$.

Definition 5.3.2.1. Let M be a computable metric subspace of $\mathbb{R}^n$. We say that M is a m dimensional computable C^k submanifold of $\mathbb{R}^n$ if there exists a computable function $f : M \times B(0,1) \to M$ (where $B(0,1)$ is the unit ball of $\mathbb{R}^m$and $M \times B(0,1)$ with the euclidean distance is a CMS in a natural way) such that for each $x \in M$, $f_x = f(x,.)$ is a C^k diffeomorphism (satisfying $f(x,0) = x$) with all k derivatives being computable in the sense that the functions $D^k f$ associating to $(x,z) \in M \times B(0,1)$ the derivative of f_x at z, denoted by $Df_{x,z}$ are computable. ⌟

For each x, the above f_x is a map whose differential at any $z \in B(\mathbf{0},1)$ is a linear, rank m function $Df_{x,z} : \mathbb{R}^m \to \mathbb{R}^n$. This can be seen as a composition of two functions $Df_{x,z} = Df^2_{x,z} \circ Df^1_{x,z}$ such that $Df^1_{x,z} : \mathbb{R}^m \to \mathbb{R}^m$ is invertible and $Df^2_{x,z} : \mathbb{R}^m \to \mathbb{R}^n$ is an isometry.

Let us denote B_x the image of $B(0,1)$ by f_x. Then the Lebesgue measure of $D \subset B_x$ is defined as

$$m(D) = \int_{f_x^{-1}(D)} \det(Df^1_{x,z})dz.$$

This does not depend on the choice of B_x and f_x, and it give rise to a finite measure (Lebesgue measure) on M (see [GMS98] page 74). This measure is indeed the m dimensional Hausdorff measure on M. Moreover, as the following lemma shows, the Lebesgue measure m is computable.

Lemma 5.3.2.1. *The Lebesgue measure on a computable C^k submanifold of $\mathbb{R}^n$ is computable.*

Proof. Suppose that A is a constructively open subset of some B_s, where s is an ideal point of M. Since the function $\det(Df^1_{x,z})$ is computable and the function $1_{f^{-1}(A)}(z)$ is lower semi-computable, we can lower semi-compute the value $m(A)$. In particular, there is a base of ideal balls whose measures are lower semi-computable. Let B and B' be such balls. Since these balls have zero measure boundaries, we can compute the measure of their intersection (which is an r.e open included in B). Hence any constructively open set can be decomposed into a (same measure) disjoint union of r.e open sets whose measures can be lower semi-computed. By Theorem 2.2.2.3, m is computable. □

The SRB measure of uniformly hyperbolic systems

Let us consider a connected C^2 computable manifold M and a dynamical system (M,T) where T is a C^2 computable diffeomorphism on M. Let $Q \subset M$ be a constructively open forward invariant set (i.e. $T(\overline{Q}) \subset Q$) and consider the (attracting) set

$$\Lambda = \bigcap_{n \geq 0} T^n(Q).$$

Suppose that Λ contains a dense orbit and that it is an hyperbolic set for T, which means that the following conditions are satisfied: there is a splitting of the tangent bundle of M on Λ: $T_\Lambda M = E^s_\Lambda \oplus E^u_\Lambda$ (at each point x of Λ the tangent space at x can be splitted in a direct sum of two spaces, the stable directions and the unstable ones) and a $\lambda_0 < 1$ such that

- the splitting is compatible with T, that is: $DT_x(E^s_x) = E^s_{T(x)}$ and $DT_x^{-1}(E^u_x) = E^u_{T^{-1}(x)}$.
- The dynamics expand exponentially fast in the unstable directions and contracts exponentially fast in the stable directions in an uniform way, that is: for each $x \in \Lambda$ and for each $v \in E^s_x$ and $w \in E^u_x$, $|DT_x(v)| \leq |\lambda_0 v|$ and $|DT_x^{-1}(w)| \leq |\lambda_0 w|$.

Under these assumptions it is known that

Theorem 5.3.2.1. *(see [Via97] e.g.)There is a unique invariant SRB measure μ supported on Λ. Moreover the measure is ergodic and its basin has full Lebesgue measure on Q.*

This measure has many good properties: it has exponential decay of correlations and it is stable under perturbations of T (see [Via97] e.g.). Another good property of this measure is that it is computable.

Theorem 5.3.2.2. *If M and T are C^2, computable and uniformly hyperbolic as above, then the SRB measure μ is computable.*

Proof. Let m be the Lebesgue measure on Q normalized by $m(Q) = 1$, clearly it is a computable measure. From [Via97] (Prop. 4.9, Remark 4.2) follows that there is $\lambda < 1$ such that for each ν-Hölder ($\nu \in (0, 1]$) continuous observable ψ, it holds

$$|\int_X \psi \circ T^n dm - \int \psi d\mu| \leq \lambda^n \ c_\psi \tag{5.2}$$

where $c_\psi = C \int |\psi| dm + \|\psi\|_\nu$, where C is independent from ψ. Then c_{ψ_i} can be computed for each uniform sequence $\psi_i \in \mathcal{H}$ uniformly in i. Since $\int_X \psi_i \circ T^n dm$ is computable uniformly in n (by Corollary 2.2.2.1), equation 5.2 implies that this sequence converges effectively to $\int \psi_i d\mu$, which is then computable. By theorem 2.2.2.3, μ is computable. □

Corollary 5.3.2.1. *In an unif. hyp. computable system (M, T) equipped with its SRB measure as above, the set of computable T-typical points is dense in the support of μ..*

Proof. μ is computable by the previous theorem, and the correlations decay is given by proposition 4.9 in [Via97] from which follows that there is $\lambda < 1$ such that for each $(g_i, g_j) \in \mathcal{H}^2$ the relation

$$|\int_X g_i \circ T^n g_j dm - \int g_i d\mu \int g_j d\mu| \leq \lambda^n \; c_{g_i,g_j}$$

holds, where $c_{g_i,g_j} = C(\int |g_i| dm + \|g_i\|_1)(\int |g_j| dm + \|g_j\|_1)$ (C is a constant independent of $g_i \in \mathcal{F}$, $\| * \|_1$ is the Lipschitz norm, since functions in $\mathcal{H}$ are Lipschitz) are computable uniformly in i, j. Then the result follows from theorem 5.2.0.6 and remark 5.2.0.2. □

5.3.3 Piecewise expanding maps

We introduce a class of discontinuous maps on the interval having an absolutely continuous SRB invariant measure. The density of this measure has also bounded variation. We will show that this invariant measure is computable.

Let I be the unit interval. Let $T : I \to I$ we say that T is piecewise expanding if there is a finite partition $P = \{I_1, ..., I_k\}$ of I, such that I_i are disjoint intervals and:

1. the restriction of T to each interval I_i can be extended to a C^1 monotonic map defined on $\overline{I_i}$ and the function $h : I \to \mathbb{R}$ defined by $h(x) = |DT(x)|^{-1}$ has bounded variation.

2. there are constants $C > 0$ and $\sigma > 1$ such that $|DT^n(x)| > C\sigma^n$ for every $n \geq 1$ and every $x \in I$ for which the derivative is defined.

3. For each interval $J \subset I$ there is $n \geq 1$ such that $f^n(J) = I$.

We remark that by point 1), in each interval I_i the map is Lipschitz. A classical result say that this kind of map has an absolutely continuous invariant measure (see [Via97], chapter 3 e.g.).

Theorem 5.3.3.1. *If T a piecewise expanding map as above, then it has a unique ergodic absolutely continuous invariant measure μ. The basin of this SRB measure has full Lebesgue measure. Moreover μ can be written as $\mu = \phi m$ where ϕ has bounded variation and m is the Lebesgue measure.*

Moreover as before, the SRB measure is also computable

Proposition 5.3.3.1. *If T is a m-almost computable piecewise expanding map satisfying points 1),...,3) above then its SRB measure is computable.*

Proof. Let us consider $\psi \in \mathcal{H}$ and $\int_X \psi \circ T^n dm$. Since T is m-almost computable, the ends of the intervals $I_1, ..., I_k$ are computable, and so is (uniformly on n) for T^n. If T is l-Lipschitz in each interval I_i then T^n is l^n-Lipschitz in each of its continuity intervals and the Lipschitz constant of $\psi \circ T^n$ on each continuity interval can be computed. Anyway, the functions $\psi \circ T^n$ are almost computable and then by Proposition 2.2.2.1 the integrals $\int_X \psi \circ T^n dm$ are uniformly computable.

Now, from [Via97] proposition 3.8, remark 3.2 it holds that there are $\lambda < 1, C > 0$, such that for each $\psi \in L^1$

$$|\int_X \psi \circ T^n dm - \int \psi d\mu| \leq C \ \lambda^n \ ||\psi||_{L^1}.$$

Hence, as $\int_X \psi \circ T^n dm$ is uniformly (in n) computable, so is $\int \psi d\mu$. By Thm 2.2.2.3, μ is computable. □

As Unif. Hyperbolic systems, also Piecewise Expanding maps can be shown to have exponential decay of correlations on bounded variation observables (see [Via97] Remark 3.2) and BV norm of functions in $\mathcal{H}$ can be computed. Hence as in the previous section we obtain,

Corollary 5.3.3.1. *In an m-computable piecewise expanding system equipped with its SRB measure, the set of computable typical points is dense in* $[0, 1]$.

5.3.4 Manneville-Pomeau type maps

We say that a map $T : [0, 1] \to [0, 1]$ is a *Manneville-Pomeau type map (MP map)* with exponent s if it satisfies the following conditions:

1. there is $c \in (0, 1)$ such that, if $I_0 = [0, c]$ and $I_1 = (c, 1]$, then $T\big|_{(0,c)}$ and $T\big|_{(c,1)}$ extend to C^1 diffeomorphisms, which is C^2 for $x > 0$, $T(I_0) = [0, 1]$, $T(I_1) = (0, 1]$ and $T(0) = 0$;

2. there is $\lambda > 1$ such that $T' \geq \lambda$ on I_1, whereas $T' > 1$ on $(0, c]$ and $T'(0) = 1$;

3. the map T has the following behaviour when $x \to 0^+$

$$T(x) = x + rx^{1+s}(1 + u(x))$$

for some constant $r > 0$ and $s > 0$ and u satisfies $u(0) = 0$ and $u'(x) = O(x^{t-1})$ as $x \to 0^+$ for some $t > 0$.

In [Iso03] (see also [Gou04]) it is proved that for $0 < s < 1$ these systems have a unique absolutely continuous invariant measure, whose density f is locally Lipschitz in a neighborhood of each $x > 0$ (the density diverges at $x = 0$) the system has polynomial decay of correlations for $(1-s)$-Hölder observables. Moreover we have that:

Theorem 5.3.4.1. *If T is a computable MP map then its absolutely continuous invariant measure μ is computable.*

Proof. Let f be the density of μ. T is topologically conjugated to the doubling map $x \to 2x \pmod 1$ hence for each small interval I there is $k > 0$ such that $T^k(I) = [0,1]$. Since f is locally Lipschitz, there is a small interval J on which $f > \delta_1 > 0$. Let n be such that $T^n(J) = [0,1]$. Let I be some small interval, then there exist $J' \subset J$ such that $T^n(J') = I$. Since T is λ-Lipschitz, we have $m(J') \geq \frac{m(I)}{\lambda^n}$. By this, $\mu(J') \geq \frac{\delta_1 m(I)}{\lambda^n}$ and by the invariance of μ, $\frac{\mu(I)}{m(I)} \geq \frac{\delta_1}{\lambda^n}$ and then, as I is arbitrary, for each $x \in [0,1]$ we have $f(x) > \frac{\delta_1}{\lambda^n} > 0$. In particular, $\frac{1}{f}$ is $(1-s)$-Hölder. Now we use the fact that the system has polynomial decay of correlations for $(1-s)$-Hölder observables. Let us consider $\phi \in \mathcal{H}$. Then we have that $\frac{1}{f}d\mu = dm$ and $\int \frac{1}{f}d\mu = 1$, hence, by the decay of correlation of this kind of maps

$$|\int \phi \circ T^n dm - \int \phi d\mu| = |\int \phi \circ T^n \frac{1}{f} d\mu - \int \phi d\mu \int \frac{1}{f} d\mu| \leq C\, ||\phi||_{1-s} \left|\left|\frac{1}{f}\right|\right|_{1-s} n^{s-1}.$$

The norm $||\phi||_{1-s}$ can be estimated for functions in $\mathcal{H}$, and then, as in the previous examples we have a way to calculate $\int \phi d\mu$ for each $\phi \in \mathcal{H}$ and again by Thm 2.2.2.3, μ is computable. □

Corollary 5.3.4.1. *In a computable Manneville-Pomeau type system, the set of computable typical points is dense in $[0,1]$.*

Chapter 6

Some open questions and future work

6.1 Computing invariant measures

Let (X, T) be a dynamical system over the computable metric space X. Let $M(X)$ denote the set of Borel probability measures on X. The problem of computing the invariant measures can be handled in a more general way. The dynamics T induces a dynamics $L : M(X) \to M(X)$ called the Perron Frobenius operator. It is defined by duality in the following way: if $\mu \in PM(X)$ then $L(\mu)$ is such that

$$\int_X f \; dL(\mu) = \int_X f \circ T \; d\mu$$

for each $f \in C_b^0(X)$.

Then, invariant measures can be found as solutions of the equation

$$W_1(\mu, L(\mu)) = 0$$

where W_1 denotes the Wasserstein distance. It is not hard to prove that under the following conditions, then these solutions can be computed.

1. the operator L is computable,
2. the space $M(X)$ is *recursively compact*,
3. the solutions of this equation are isolated.

Condition 1 is easily satisfied when T is computable on the whole space X. If X itself is recursively compact, then it can be shows that so is $M(X)$. If T is not computable then L is not necessarily computable. Still, by Corollary 2.2.2.4 and Proposition 2.2.2.1 we have that L can be computed at all measures μ which are "far enough" from the discontinuity set D. This is technically expressed by the condition $\mu(D) = 0$. Hence, condition 1 and 2 can be replaced by "L is computable on some subset $V \subset \mathcal{M}(X)$ which is recursively compact". In order to find that V, one idea is to define a norm over probability measure and search among the subsets of measures whose norm is less than a certain rang.

6.2 Computing metric entropy

The problem of computing the metric entropy can be handled using the computable metric space of Borel sets. As the metric entropy of a computable partition is a computable real, the existence of a computable generating partition implies the upper semi-computability of the metric entropy. If moreover, this partition is *effectively generating*, then the metric entropy can be computed. Are there such partitions?

6.3 More abstract random objects

Sometimes it is not easy to define what the *uniform measure* is. This is typically the case when we consider infinite dimensional spaces (as functions spaces). Such a notion could be obtained with the help of Kolmogorov complexity through the identification *random-incompressible*. For example, the space of continuous functions over $[0, 1]$ can be turned into a computable metric space, and then the Kolmogorov complexity of this functions can be defined. For instance, the fact that *complex* functions are nowhere derivable should not be difficult to prove. It would be interesting to see what properties *random* dynamical systems have.

Part III

Appendix

Appendix A

Background

A.1 Measure theory

We briefly present the Kolmogorov axiomatization of probability theory [Kol33].

Let X be a set. A family $\mathscr{B}$ of subsets of X is called an ***algebra*** if

$$X \in \mathscr{B},$$
$$A \in \mathscr{B} \Rightarrow A^{\mathcal{C}} \in \mathscr{B},$$
$$A, B \in \mathscr{B} \Rightarrow A \cup B \in \mathscr{B}.$$

We say that $\mathscr{B}$ is a ***σ-algebra*** if moreover

$$A_i \in \mathscr{B}, i \geq 1 \Rightarrow \bigcup_i A_i \in \mathscr{B}.$$

If $\mathscr{B}_0$ is a family of subsets of X, the σ-algebra generated by $\mathscr{B}_0$ (denoted $\sigma(\mathscr{B}_0)$) is defined to be the smallest σ-algebra over X that contains $\mathscr{B}_0$. A ***separable sigma algebra*** is a sigma algebra that can be generated by a countable collection of sets.

We then call the pair $(X, \mathscr{B})$ a *measurable space*. A *finite measure* on $(X, \mathscr{B})$ is a function $\mu : \mathscr{B} \to \mathbb{R}^+$ satisfying $\mu(\emptyset) = 0$ and (additivity) $\mu(\cup_{n=1}^{\infty} B_n) = \sum_{n=1}^{\infty} \mu(B_n)$ whenever $\{B_n\}_1^{\infty}$ is a sequence of members of $\mathscr{B}$ which are pairwise disjoint subsets of X. A *finite measure space* is a triple $(X, \mathscr{B}, \mu)$ where $(X, \mathscr{B})$ is a measurable space and μ is a finite measure on $(X, \mathscr{B})$.

If X is a topological space, the ***Borel*** σ-algebra of X is defined as the σ-algebra generated by the family of open sets of X. Sets in the Borel σ-algebra are called Borel sets. If the topology on X has a countable basis, then the Borel σ-algebra is separable.

In this work, we shall usually only consider probability spaces, which are finite measure spaces $(X, \mathscr{B}, \mu)$ with $\mu(X) = 1$ and where $\mathscr{B}$ is the Borel σ-algebra of X. A set $A \subset X$ has ***measure zero*** if there is a Borel set A_1 such that $A \subset A_1$ and $\mu(A_1) = 0$. We call two sets $A_1, A_2 \subset X$ ***equivalent modulo zero***, and write $A_1 = A_2 \pmod 0$ if the symmetric difference $A \triangle B$ has measure zero. We say that the collection $\mathscr{B}_0$ of subsets of X ***generates*** $\mathscr{B}$ ***(mod 0)*** if for every $A \in \mathscr{B}$ there is some $B \in \sigma(\mathscr{B}_0)$ such that $A = B(\text{mod } 0)$. We shall use the following result, see [Mañ87]:

Theorem A.1.0.2. *[Approximation theorem]If $(X, \mathscr{A}, \mu)$ is a probability space, a sub-algebra $\mathscr{A}_0$ generates $\mathscr{A}$ (mod 0) if and only if, for every $A \in \mathscr{A}$ and $\epsilon > 0$, there exists $A_0 \in \mathscr{A}_0$ such that $\mu(A_0 \triangle A) \leq \epsilon$.*

We write $A_1 \subset A_2 \pmod 0$ if A_1 is a subset of A_2 and $A_1 = A_2 \pmod 0$.

Definition A.1.0.1. Let us say that a family of Borel sets $\mathcal{E}$ is ***essential***, if for every open set U there is a sequence $(E_i)_i \subset \mathcal{E}$ such that $\cup_i E_i \subset U \pmod 0$. ⌟

A *random variable* is a *measurable* function $f : X \to \mathbb{R}^*$ which means that for each interval I, the event which yields f to take values in I, namely $f^{-1}(I)$, is a measurable event, i.e., is a member of $\mathscr{B}$. For a random variable f, we denote by $\mu f = \int f d\mu$ its expectation value. Let us recall the following classical results:

Theorem A.1.0.3. *[Coupling Theorem [Str65]]Let μ and ν be two probability measures over a complete separable metric space X with $\rho(\mu, \nu) \leq \epsilon$. Then there is a probability measure P on the space $X \times X$ with marginals μ and ν such that for a pair of random variables f, g having joint distribution P we have $P[x : d(f(x), g(x)) > \epsilon] \leq \epsilon$.*

Theorem A.1.0.4 (Monotone convergence theorem)**.** *Let (f_n) be a sequence of real integrable functions on X such that the sequence $(f_n(x))$ is monotonically increasing for almost every point and assume*

$$\sup_n \int_X f_n d\mu < \infty$$

Then the function $f(x) := \lim f_n(x)$ is integrable, and

$$\int_X f d\mu = \lim \int_X f_n d\mu.$$

A.2 Ergodic Theory

We will say just a few words about the standard way to model both deterministic and stochastic physical systems within the same framework, that of ergodic theory. For an introduction see for example [Mañ87], [Wal82], [HK95], [HK02], [Pet83]. The set of all possible *states* of the system will be denoted by X, which will be usually a metric space. A *point* $x \in X$ is supposed to give all possible information about the system; absolutely perfect knowledge of, for example, the position and momentum of every particle constituting the system. Since such perfect information is unattainable, one only assumes to be able to know whether or not x is consistent which some *observable event* B subset of X, that is, whether or not $x \in B$. To permit countable-infinite set-theoretic operations, one suppose the collection of all observable events to form a σ-algebra $\mathfrak{B}$, over which a probability measure μ is considered. A *measurement* on the system is modeled by a random variable $f : X \to \mathbb{R}$.

Now we describe dynamics. Development in time is modeled by a *measure preserving* transformation $T : X \to X$. This is a function which preserves observability (T is measurable) and probability: $\mu(T^{-1}(B)) = \mu(B)$ (we also say that μ is an invariant probability measure). The idea is that if the system is in a state $x \in X$ at a given instant, then at next instant it will be in state $T(x)$. The invariance of μ under T reflects the fact that we are in an *equilibrium* situation: probabilities of observable events do not change in time. A typical picture is the following: an initial point is chosen according to Lebesgue measure (subject to errors in measurement); the *trajectory* (or *orbit*) of this point $O(x) = \{x, T(x), T^2(x), \ldots\}$ approaches an "attractor", on which the limiting dynamics take place, and that can be quite complicated; Lebesgue measure itself also evolves in time towards a limiting invariant measure, supported on the attractor, describing at least in statistical terms the dynamics of the equilibrium situation of the system being studied. Different initial condition may lead in long term to quite different particular behaviors, but identical in a qualitative sens. An invariant measure is called *ergodic* if for all $B \in \mathfrak{B}$ such that $T^{-1}(B) = B$, it holds $\mu(B) = 0$ or $\mu(B^c) = 0$.

Ergodic theory is firstly concerned with understanding "essentially different" measure preserving transformations, where two transformations are considered to be essentially the same if they are *isomorphic*. The approaches to the isomorphism problem usually involves searching for isomorphism invariant quantities, which are useful in distinguishing and classifying transformations.

A.3 Kolmogorov Complexity

There are various ways in which the intuition behind the definition of Kolmogorov complexity [Kol83] can be stated. Consider the following strings, supposed to be generated by a random process like coin tossing:

$$x = 000000000000000000000000$$

$$y = 001011001011001011001011$$

$$z = 100100110110001110110100$$

Under the hypothesis of random generation, the perfect regularity of x make it appears to us extremely extraordinary and possibly it wouldn't be accepted as a random outcome. A little less evident is the regularity of the second string y, which is also a periodic one. The third string z seems to have no evident regularity, and it would be possibly accepted as a random outcome. However, probability theory assigns the same probability $P = 2^{-24}$ to all of them, and allows no distinction.

In [Gác], Peter Gács remarks,

> "...this convinces us only that the axioms of Probability theory, as developed in [Kolmogorov], do not solve all mysteries that they are sometimes supposed to".

Pierre-Simon Laplace has pointed out the following reason why intuitively a regular outcome of a random event is unlikely,

> "We arrange in our thought all possible events in various classes; and we regard as *extraordinary* those classes which include a very small number. In the games of heads and tails, if head comes up a hundred times in a row this appears to us extraordinary, because the almost infinite number of combinations that can arise in a hundred throws are divided in regular sequences, or those in which we observe a rule that is easy to grasp, and irregular sequences, that are incomparably more numerous." ([Lap52])

But, how to formalize such a distinction between regular and irregular sequences? Laplace distinguishes also between the object itself and a cause of the object,

> "The regular combinations occur more rarely only because they are less numerous. If we seek a cause wherever we perceive symmetry, it is not that we regard the symmetrical event as less possible than the others, but, since this event ought to be the effect of a regular cause or that of chance, the first of these suppositions is more probable than the second."([Lap52].)

Or, in Lévy's words,

> "Si donc en présence d'une suite remarquable nous excluons la première hypothèse [of the random origin of the data] ce n'est pas que le hasard ait *a priori* moins de chance de la produire qu'une autre; c'est qu'une autre cause que le hasard a plus de chance de la produire." [Lév25]

So, we can expect regular sequences to be produced by other causes but chance. Be able to find such a cause, that is, to have a theory (or a model) explaining the regularities of large data, means to be able to reproduce the data using this theory (this model). Thus, we can use it to give a much shorter description of the data than the data itself. This identification between regularity and compressibility is the central idea in the definition of Kolmogorov Complexity as a minimal description length function.

Let $\mathcal{X}$ be a class of finite objects, and $\mathcal{C}$ a class of finite words (the codes). A minimal description length function corresponds to a (surjective) decoding procedure $f : \mathcal{C} \to X$ such that for all other decoding procedures $g : \mathcal{C} \to X$ and for all object $x \in \mathcal{X}$, f satisfy an optimality property of the form:

$$\min_{c \in \mathcal{C}} \{|c| : f(c) = x\} \leq \min_{c \in \mathcal{C}} \{|c| : g(c) = x\}$$

If we let the class of decoding procedures to be the class of all functions from $\mathcal{C}$ to $\mathcal{X}$, then there is no such a minimal procedure; the natural restriction to make here is to consider only algorithmic procedures.

One more thing, the decoding procedure should be able to separate different codes which have been concatenated in a single word without other information that the concatenated string itself, or equivalently, codes must include information about their own lengths. That is, decoding procedures must have prefix-free domains (see section 1.2). Such codes are called self-delimiting codes. Since a given code is supposed to contain all necessary information to reconstruct in an algorithmic way the coded finite object, the Kolmogorov Complexity is also called *Algorithmic Information Content*. We prefer to use the latter, since as we will see later, random objects have very high information contents. But randomness is the absence of any "structure" and "organization", so, in a sense, random objects should not be very complex.

Definition A.3.0.2. . A ***prefix free algorithm*** is a partial recursive function $A : \{0,1\}^* \to \{0,1\}^*$ which has a prefix-free domain. A prefix free algorithm is called ***computable*** iff

$$\sum_{w \in \mathrm{dom} A} 2^{-|w|}$$

is a computable real. Let $\mathcal{A}$ be the set of all prefix free algorithms. ⌟

Definition A.3.0.3. Let $A \in \mathcal{A}$ be a prefix free algorithm. The ***Kolmogorov Complexity*** (or ***Information Content***) $K_A(w)$ of $w \in \{0,1\}^*$ is defined to be

$$K_A(w) = \begin{cases} \infty & \text{if there is no } p \text{ such that } A(p) = w \\ |p| & \text{if } p \text{ is a shortest input such that } A(p) = w \end{cases}$$

⌟

To have the promised minimality property, we need the existence of a universal prefix algorithm.

Theorem A.3.0.5. *$\mathcal{A}$ is recursively enumerable.*

Proof. We construct an algorithm P which transforms any number e into a Gödelnumber $P(e) = e^{A_{P(e)}}$ for a prefix algorithm $A_{P(e)}$ and such that every prefix algorithm has at least one Gödelnumber in the range of P. Given e, P generates the domain of the partial recursive function with Gödelnumber e, ϕ_e. A partial recursive function $A_{P(e)}$ with Gödelnumber $P(e)$ is determined in the following way: $A_{P(e)}(q)$ equals $\phi_e(q)$ except for those $q \in dom\phi_e$ which are initial segments or prolongations of previously generated $p \in dom\phi_e$. For these q, $A_{P(e)}(q)$ is undefined. By construction, $A_{P(e)}$ is a prefix algorithm and if ϕ_e is a prefix algorithm then $P(e) = e$. Hence the set of prefix algorithms is recursively enumerable. □

As a consequence, we have the existence of a universal prefix algorithm.

Theorem A.3.0.6. *There exist a universal prefix free algorithm which is asymptotically optimal, that is an $U \in \mathcal{A}$ such that $\forall\ A \in \mathcal{A}\ \exists\ c_A \in \mathbb{N}$, which depends on A, such that $\forall\ p \in \{0,1\}^*\ \exists\ q \in \{0,1\}^*$ which satisfy $U(q) = A(p)$ and $|q| = |p| + c_A$.*

Proof. We define a universal prefix algorithm U as follows: on inputs of the form $q = 0^{e^A}1p$, U simulates the action of the prefix algorithm A with Gödel-number e^A on p. For this U, if A is a prefix algorithm with Gödel-number e^A and $p \in \{0,1\}^*$ is an input, then $q = 0^{e^A}1p$ is such that $U(q) = A(p)$ and $|q| = |p| + e^A + 1$. $\square$

We fix an asymptotically optimal universal prefix free algorithm U and we let $K(w) = K_U(w)$. We can now state the *invariance theorem*, which gives the optimality property.

Theorem A.3.0.7. *$\forall A \in \mathcal{A}$, $\exists\, c_A \in \mathbb{N}$ such that $\forall w \in \{0,1\}^*$ we have $K(w) \leq K_A(w) + c_A$.*

Proof. Let p^* a shortest input such that $A(p^*) = w$, then $q = 0^{e^A}1p^*$ is such that $U(q) = w$ and then $K(w) \leq K_A(w) + c_A$, where $c_A = e^A + 1$. $\square$

We have claimed that the central idea in the definition of Kolmogorov Complexity is the identification between regularity and compressibility, so an irregular object (a random one) should be incompressible, that is to say with a high Information Content. Consequently, we have to look at sharp upper bounds and ask for a random object to have an information content close to these bounds. Given a finite string w, it is not difficult to think in a self-delimiting code which contains the word w itself and the information about its own length. In fact, we have the following easy first upper bound.

Lemma A.3.0.1. *For some constant c and all w: $I(w) \leq |w| + I(|w|) + c$*

More generally we have the following:

Theorem A.3.0.8. *Let μ be a computable measure on $\{0,1\}^{\mathbb{N}}$. Then for some constant c and all w: $K(w) \leq -\log \mu[w] + K(|w|) + c$*

We state now the promised relation between randomness and incompressibility for infinite sequence, expressed as the incompressibility of their initial segments.

Theorem A.3.0.9 (Schnorr, Levin)**.** *Let μ be a computable measure. Then ω is a Martin Löf random sequence with respect to μ if and only if $\exists m\ \forall n\ K(\omega_{1:n}) > -\log \mu[\omega_{1:n}] - m$.*

The minimal such m, defined by $d_\mu(\omega) := \sup_n\{-\log \mu[\omega_{1:n}] - K(\omega_{1:n})\}$ and called the ***randomness deficiency*** of x w.r.t μ, in addition to be finite almost everywhere has also a finite mean, that is $\int d_\mu(\omega) d\mu \leq 1$. For a proof see [LV93].

For Schnorr random sequences, a characterization in terms of Kolmogorov complexity is only available for the uniform measure. (At least at our knowledge).

Theorem A.3.0.10 (Downey)**.** *An infinite sequence $\omega \in \{0,1\}^{\mathbb{N}}$ is Schnorr random iff for all computable prefix free algorithms C, there is a constant m such that, for all n $K_C(\omega_{1:n}) \geq n - m$.*

Appendix B

Some proofs

Proof of proposition 4.3.2.3. It follows from the following two lemmas.

Lemma B.0.0.2. *Let V be a computable martingale for a computable measure μ over $\Sigma^{\mathbb{N}}$, and $f : \mathbb{N} \to \mathbb{N}$ an unbounded monotonic computable function. There is a Schnorr test $(Y_i)_i$ such that all elements of $\mathfrak{N}_{V,f}$ fail it.*

Proof. For integer $i \geq 1$ let

$$Y_i = \{x \in \Sigma^* : V^2(x) > f(|x|) \text{ and } V(x) > 2^i V(\Lambda)\}.$$

The sets Y_i are uniformly recursively enumerable by the form of their definition. The requirement $V(x) > 2^i V(\Lambda)$ and the martingale inequality (4.3.2.2) imply $\mu(Y_i) \leq 2^{-i}$, so the sets Y_i form a Martin-Löf test. It is also easy to see that $\omega \in \mathfrak{N}_{V,f}$ implies $\omega \in Y_i\Sigma^{\mathbb{N}}$ for all i.

It remains to show that the value $\mu(Y_i)$ is computable, uniformly in i. In order to compute it to within 2^{-r}, let n be such that $f(n) > 2^{2r}$, and let $A_n = Y_i \setminus \Sigma^n\Sigma^*$. Then we have

$$Y_i \setminus A_n \subseteq \{x \in \Sigma^* : V(x) > 2^r\}$$

and by the martingale inequality $\mu(Y_i) - \mu(A_n) \leq 2^{-r}$. Since the set A_n is finite, we can compute $\mu(A_n)$ to arbitrary precision. □

Lemma B.0.0.3. *If $(Y_i)_i$ is a Schnorr test for the uniform distribution λ over $\{0,1\}^{\mathbb{N}}$, then there is a computable martingale $V(x)$ and an unbounded monotonic computable function f with the property that the set of sequences failing the test is contained in $\mathfrak{N}_{V,f}$.*

Proof. Let us show first some easy ways to create martingales for the uniform distribution. First note that the positive linear combination of martingales is a martingale. The following easy proposition allows to extend some partial functions to martingales.

Proposition B.0.0.1. *For given integer n let $V : \{0,1\}^n \to \mathbb{R}_+$ be some computable function. We can extend it to be a martingale for the uniform distribution λ as follows. For $|x| \geq n$ set $V(x) = V(x_{[n]})$. For $|x| < n$ set*

$$V(x) = \lambda(x)^{-1} \sum_{\substack{y\in\Sigma^n \\ y \sqsupseteq x}} \lambda(y)V(y).$$

We have $V(\Lambda) = \sum_{y\in\{0,1\}^n} \lambda(y)V(y)$.

Given a Schnorr test $(Y_i)_i$ it is easy to see that we can assume each set Y_i to consist of incompatible strings (to be "prefix-free"), and to be recursive (since only $Y_i\Sigma^{\mathbb{N}}$ matters, we can arrange that at stage t of the enumeration of Y_i we use only strings of length $\geq t$). Let $B = \bigcup_i Y_i$, then it is also easy to see that $\mu(B)$ is computable. Let $B_n = B \cap \Sigma^n\Sigma^*$, then we can define an unbounded monotonic unbounded recursive function $f(n)$ with $\mu(B_n) \leq 2^{-2f(n)}$. With this we have

$$\sum_{\substack{x\in B \\ f(|x|)=m}} \lambda(x)2^{f(|x|)} \leq 2^{-2m}2^m = 2^{-m},$$

$$\sum_{x\in B} \lambda(x)2^{f(|x|)} \leq \sum_{m=0}^{\infty} 2^{-m} = 2.$$

Also, the sum $\sum_{x\in B} \lambda(x)2^{f(|x|)}$ is computable, and we can define a monotonic unbounded recursive function $g(n)$ with

$$\sum_{x\in B\cap\Sigma^{g(n)}\Sigma^*} \lambda(x)2^{f(|x|)} \leq 2^{-n}. \tag{B.1}$$

With a new test $\overline{Y}_n = B \cap \Sigma^{g(n)}\Sigma^*$, it can be verified that the sequences failing the test $(Y_i)_i$ are the same as those failing $(\overline{Y}_n)_n$.

For nonnegative integers i, n, let us define the function $V_{i,n} : \{0,1\}^n \to \mathbb{R}_+$ by $V_{i,n}(x) = 2^{f(n)}1_{\overline{Y}_i}(x)$, and define from it a martingale using the construction of Proposition B.0.0.1. Then we have

$$V_{i,n}(\Lambda) = \sum_{x\in\{0,1\}^n\cap\overline{Y}_i} \lambda(x)2^{f(|x|)}.$$

Let us define the martingale $V_i(x) = \sum_n V_{i,n}(x)$. Because of the fast convergence it is easy to see that $V_i(x)$ is computable, and (B.1) implies $V_i(\Lambda) \le 2^{-i}$. Finally we define the martingale $V(x) = \sum_{i\ge 1} V_i(x)$. It is again computable because of the fast convergence, with $V(\Lambda) \le \sum_{i\ge 1} 2^{-i} \le 1$. Since $x \in \overline{Y}_i$ implies $V(x) \ge f(|x|)$, if ω fails the test $(\overline{Y}_i)_i$ then it is in $\mathfrak{N}_{V,f}$. $\square$

This completes the proof of proposition 4.3.2.3. $\square$

Index

Bibliography

[AGS05] L. Ambrosio, N. Gigli, and G. Savaré. *Gradient flows: in metric spaces and in the space of probability measures.* Birkhauser, Zürich, 2005.

[AJ94] Samson Abramsky and Achim Jung. Domain theory. In S. Abramsky, D. Gabbay, and T. S. E. Maibaum, editors, *Handbook of Logic in Computer Science Volume 3*, pages 1–168. Oxford University Press, 1994.

[AKM65] R. L. Adler, A. G. Konheim, and M. H. McAndrew. Topological entropy. *Transactions of the American Mathematical Society*, 114:309–319, February 1965.

[Asa88] Eugene A. Asarin. *Individual Random Signals: a Complexity Approach. PhD thesis. Academic advisor: Andrei N. Kolmogorov.* PhD thesis, Moscow State University, 1988.

[BF02] Verónica Becher and Santiago Figueira. An example of a computable absolutely normal number. *Theor. Comput. Sci.*, 270(1-2):947–958, 2002.

[BH03] Olivier Bournez and Mathieu Hoyrup. Rewriting logic and probabilities. In Robert Nieuwenhuis, editor, *RTA*, volume 2706 of *Lecture Notes in Computer Science*, pages 61–75. Springer, 2003.

[Bil] Patrick Billingsley. *Probability and Measure.* New York, Toronto, London: John Wiley and Sons.

[Bil65] P. Billingsley. *Ergodic Theory and Information.* Wiley, New York, 1965.

[Bil68] Patrick Billingsley. *Convergence of Probability Measures.* John Wiley, New York, 1968.

[Bor09] Emile Borel. Les probabilités dénombrables et leurs applications arithmétiques. *Rendiconti del Circolo Matematico di Palermo*, 27:247–271, 1909.

[Bow71] Rufus Bowen. Entropy for group endomorphisms and homogeneous spaces. *Transactions of the American Mathematical Society*, 153:401–414, January 1971.

[Bow73] Rufus Bowen. Topological entropy for noncompact sets. *Transactions of the American Mathematical Society*, 184:125–136, October 1973.

[BP03] Vasco Brattka and Gero Presser. Computability on subsets of metric spaces. *Theoretical Computer Science*, 305:43–76, 2003.

[Bra01] Vasco Brattka. Computable versions of baire's category theorem. In *MFCS: Symposium on Mathematical Foundations of Computer Science*, 2001.

[Bru83] A.A. Brudno. Entropy and the complexity of the trajectories of a dynamical system. *Trans. Moscow Math. Soc.*, 2:127–151, 1983.

[BW99] Vasco Brattka and Klaus Weihrauch. Computability on subsets of euclidean space i: Closed and compact subsets. *Theor. Comput. Sci.*, 219(1-2):65–93, 1999.

[Cal94] Cristian Calude. *Information and Randomness: An Algorithmic Perspective*. Springer-Verlag New York, Inc., Secaucus, NJ, USA, 1994.

[Cal02] Cristian S. Calude. *Information and randomness*. Texts in Theoretical Computer Science. An EATCS Series. Springer-Verlag, Berlin, second edition, 2002. An algorithmic perspective, With forewords by Gregory J. Chaitin and Arto Salomaa.

[CH94] Jin-Yi Cai and Juris Hartmanis. On Hausdorff and topological dimensions of the Kolmogorov complexity of the real line. In *Proceedings of the 30th IEEE symposium on Foundations of computer science*, pages 605–619, Orlando, FL, USA, 1994. Academic Press, Inc.

[Cha33] David G. Champernowne. The construction of decimals normal in the scale of ten. *Journal of the London Mathematical Society*, 8:254–260, 1933.

[Cha75] Gregory J. Chaitin. A theory of program size formally identical to information theory. *J. Assoc. Comput. Mach.*, 22:329–340, 1975.

[Cha87] Gregory J. Chaitin. *Algorithmic information theory*, volume 1 of *Cambridge Tracts in Theoretical Computer Science*. Cambridge University Press, Cambridge, 1987. With a foreword by J. T. Schwartz.

[Cha90] G. J. Chaitin. *Information, randomness & incompleteness*, volume 8 of *World Scientific Series in Computer Science*. World Scientific Publishing Co. Inc., River Edge, NJ, second edition, 1990. Papers on algorithmic information theory.

[CHJW01] Cristian S. Calude, Peter H. Hertling, Helmut Jürgensen, and Klaus Weihrauch. Randomness on full shift spaces. *Chaos Solitons Fractals*, 12(3):491–503, 2001.

[Chu40] Alonzo Church. On the concept of a random sequence. *Bulletin of American Mathematical Society*, 46(2):130–135, 1940.

[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.

[DG02] Rodney G. Downey and Evan J. Griffiths. Schnorr randomness. *Electr. Notes Theor. Comput. Sci.*, 66(1), 2002.

[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.

[Eda97] Abbas Edalat. Domains for computation in mathematics, physics and exact real arithmetic. *The Bulletin of Symbolic Logic*, 3(4):401–452, 1997.

[EH98] Abbas Edalat and Reinhold Heckmann. A computational model for metric spaces. *Theoretical Computer Science*, 193:53–73, 1998.

[ES99] Abbas Edalat and Philipp Sünderhauf. A domain-theoretic approach to computability on the real line. *Theoretical Computer Science*, 210:73–98, 1999.

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

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

[Gal99] Stefano Galatolo. Pointwise information entropy for metric spaces. *Nonlinearity*, 12(5):1289–1298, 1999.

[Gal00] Stefano Galatolo. Orbit complexity by computable structures. *Nonlinearity*, 13:1531–1546, September 2000.

[GHR07a] Stefano Galatolo, Mathieu Hoyrup, and Cristóbal Rojas. Algorithmically random points in measure preserving systems, statistical behaviour, complexity and entropy. *Submitted*, 2007.

[GHR07b] Stefano Galatolo, Mathieu Hoyrup, and Cristóbal Rojas. An effective Borel-Cantelli lemma. Constructing orbits with required statistical properties. *Submitted*, 2007.

[GHR08a] Peter Gács, Mathieu Hoyrup, and Cristóbal Rojas. Schnorr randomness–A dynamical point of view. *In preparation*, 2008.

[GHR08b] Stefano Galatolo, Mathieu Hoyrup, and Cristóbal Rojas. Computable invariant measures. *In preparation*, 2008.

[GMS98] Mariano Giaquinta, Giuseppe Modica, and Jiri Soucek. *Cartesian currents in the calculus of variations, I*, volume 37 of *Ergebnisse Math. Grenzgebiete (III Ser)*. Springer, 1998.

[Gou04] Sébastien Gouëzel. Sharp polynomial estimates for the decay of correlations. *Israel J. Math.*, 139:29–65, 2004.

[Hem02] Armin Hemmerling. Effective metric spaces and representations of the reals. *Theor. Comput. Sci.*, 284(2):347–372, 2002.

[HK95] Boris Hasselblatt and Anatole Katok. *Introduction to the Modern Theory of Dynamical Systems*, volume 54 of *Encyclopedia of Mathematics and Its Applications*. Cambridge University Press, 1995.

[HK02] Boris Hasselblatt and Anatole Katok. *Handbook of Dynamical Systems*, volume 1A, chapter Principle Structures. Elsevier, 2002.

[Hoy07] Mathieu Hoyrup. Dynamical systems: stability and simulability. *Mathematical Structures in Computer Science*, 17(2):247–259, 2007.

[Hoy08] Mathieu Hoyrup. *Computability, Randomness and Ergodic Theory on metric spaces*. PhD thesis, Université Denis Diderot, Paris., 2008.

[HR87] Jr. Hartley Rogers. *Theory of Recursive Functions and Effective Computability*. MIT Press, Cambridge, MA, USA, 1987.

[HR07] Mathieu Hoyrup and Cristóbal Rojas. Computability of probability measures and Martin-Löf randomness over metric spaces. *Submitted*, 2007.

[HW97] Peter Hertling and Yongge Wang. Invariance properties of random sequences. *Journal of Universal Computer Science*, 3(11):1241–1249, 1997.

[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.

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

[Iso03] Stefano Isola. On systems with finite ergodic degree. *Far east journal of dynamical systems*, 5:1, 2003.

[Kal03] Iraj Kalantari. Density & Baire category in recursive topology. In Vasco Brattka, Matthias Schröder, Klaus Weihrauch, and Ning Zhong, editors, *Computability and Complexity in Analysis*, volume 302 of *Informatik Berichte*, pages 15–18. FernUniversität in Hagen, August 2003. International Conference, CCA 2003, Cincinnati, USA, August 28–30, 2003.

[KF82] K.-I. Ko and H. Friedman. Computational complexity of real functions. *Theoretical Computer Science*, 20(3):323–352, 1982.

[Ko91] Ker-I Ko. *Complexity Theory of Real Functions.* Birkhauser Boston Inc., Cambridge, MA, USA, 1991.

[Kol33] Andrei N. Kolmogorov. *Grundbegriffe der Wahrscheinlichkeitsrechnung. English translation (1950): Foundations of the theory of probability.* Chelsea, New York, USA, 1933.

[Kol58] Andrei N. Kolmogorov. A new metric invariant of transitive dynamical systems, and Lebesgue space automorphisms. *Doklady Akademii Nauk*, 119(5):861–864, 1958. In Russian.

[Kol65] Andrei N. Kolmogorov. Three approaches to the quantitive definition of information. *Problems of Information Transmission*, 1:1–17, 1965.

[Kol68] Andrei N. Kolmogorov. Logical basis for information theory and probability theory. *IEEE Transactions of Information Theory*, 14:662–664, 1968.

[Kol83] Andrei N. Kolmogorov. Combinatorial foundations of information theory and the calculus of probabilities. *Russian Mathematical Surveys*, 38(4):29–40, 1983.

[Kol98] Andrei N. Kolmogorov. On tables of random numbers. *Theor. Comput. Sci.*, 207(2):387–395, 1998.

[KST94] Tyll Krüger, Peter Seibt, and Serge Troubetzkoy. Complexity and Randomness of Recursive Discretizations of Dynamical Systems. *Random and Computational Dynamics*, 2:1–19, January 1994.

[KT59] A.N. Kolmogorov and V.M. Tikhomirov. ϵ-entropy and ϵ-capacity. *Uspekhi Mat. Nauk*, 14:3–86, 1959. English transl. AMS Translations, 17:2 (1961), 277-364.

[KV61] A.N. Kolmogorov and Tikhomirov V.M. ϵ-entropy and ϵ-capacity of sets in functional spaces. *AMS Translations, Series 2*, 17:277–364, 1961.

[Lap52] Pierre Simon Laplace. *A philosophical Essay on Probabilities. (Originally published in 1819).* Dover publications, 1952.

[Lév25] Paul Lévy. *Calcul des probabilités.* Gauthiers-Villars, Paris, 1925.

[Lev84] Leonid A. Levin. Randomness conservation inequalities; information and independence in mathematical theories. *Inf. Control*, 61(1):15–37, 1984.

[Lut03] Jack H. Lutz. The dimensions of individual strings and sequences. *Information and Computation*, 187(1):49–79, 2003.

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

[Mañ87] Ricardo Mañe. *Ergodic theory and differentiable dynamics*. Springer-Verlag, New York, 1987.

[May01] Elvira Mayordomo. A Kolmogorov complexity characterization of constructive Hausdorff dimension. *Electronic Colloquium on Computational Complexity (ECCC)*, 8(059), 2001.

[Mil04] Joseph S. Miller. Degrees of unsolvability of continuous functions. *Journal of Symbolic Logic*, 69(2):555–584, 2004.

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

[ML71] Per Martin-Löf. Complexity oscillations in infinite binary sequences. *Z. Wahrscheinlichkeit*, 19:225–230, 1971.

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

[PER89] Marian B. Pour-El and J. Ian Richards. *Computability in Analysis and Physics*. Perspectives in Mathematical Logic. Springer, Berlin, 1989.

[Pes98] Yakov B. Pesin. *Dimension Theory in Dynamical Systems: Contemporary Views and Applications*. Chicago Lectures in Mathematics. The University of Chicago Press, 1998.

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

[Rei04] Jan Reimann. *Computability and fractal dimension*. PhD thesis, Ruprecht-Karls Universitat Heidelberg, 2004.

[Roj08] Cristóbal Rojas. Computability and Information in models of randomness and chaos. *Mathematical Structures in Computer Science*, 18:291–307, 2008.

[Sch71] Claus-Peter Schnorr. *Zufälligkeit und Wahrscheinlichkeit. Eine algorithmische Begründung der Wahrscheinlichkeitstheorie, Lecture Notes in Mathematics, Vol. 218.* Springer-Verlag, Berlin-New York, 1971.

[Sch72] Claus-Peter Schnorr. The process complexity and effective random tests. In *STOC*, pages 168–176, 1972.

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

[Sha48] Claude Elwood Shannon. A mathematical theory of communication. *The Bell System Technical Journal*, 27, 1948.

[Sie17] M.W. Sierpiński. Démonstration élémentaire du théorème de M. Borel sur les nombres absolument normaux et détermination effective d'un tel nombre. *Bulletin de la Société Mathématique de France*, 47:127–132, 1917.

[Sin59] Yakov G. Sinai. On the notion of entropy of dynamical systems. *Doklady Akademii Nauk*, 124(4):768–771, 1959. In Russian.

[Sta05] Ludwig Staiger. Constructive dimension equals Kolmogorov complexity. *Information Processing Letters*, 93(3):149–153, 2005.

[Str65] Volker Strassen. The existence of probability measures with given marginals. *Annals of Mathematical Statistics*, 36:423–439, 1965.

[Tur36] Alan Turing. On computable numbers, with an application to the entscheidungsproblem. *Proceedings of the London Mathematical Society*, 2, 42:230–265, 1936.

[Via97] Marcelo Viana. Stochastic dynamics of deterministic systems. *Lecture Notes XXI Braz. Math. Colloq. IMPA Rio de Janeiro*, 1997.

[Vil39] Jean Ville. *Etude critique de la notion de collectif.* Gauthiers-Villars, Paris, 1939.

[vL87] M. van Lambalgen. *Random sequences.* PhD thesis, Department of Mathematics, University of Amsterdam, 1987.

[Vov01] V. Vovk. Kolmogorov's complexity conception of probability, 2001.

[VV93] V.G. Vovk and V.V. Vyugin. On the empirical validity of the bayesian method. *Journal of the Royal Statistical Society*, B 55(1):253–266, 1993.

[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.

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

[V'y01] Vladimir V. V'yugin. Non-robustness property of the individual ergodic theorem. *Problems of Information Transmission*, 37(2):27–39, 2001.

[Wal82] Peter Walters. *An Introduction to Ergodic Theory*, volume 79 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 1982.

[Was69] L.N. Wasserstein. Markov processes over denumerable products of spaces describing large systems of automata. *Problems of Information Transmission*, 5:47–52, 1969.

[Wei93] Klaus Weihrauch. Computability on computable metric spaces. *Theoretical Computer Science*, 113:191–210, 1993. Fundamental Study.

[Wei99] Klaus Weihrauch. Computability on the probability measures on the borel sets of the unit interval. *Theoretical Computer Science*, 219:421–437, 1999.

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

[YMT99] Mariko Yasugi, Takakazu Mori, and Yoshiki Tsujii. Effective properties of sets and functions in metric spaces with computability stqucture. *Theoretical Computer Science*, 219(1-2):467–486, 1999.

[You02] Lai-Sang Young. What are srb measures, and which dynamical systems have them? *J. Stat. Phys.*, 108:733–754, 2002.

[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.