Math. Struct. in Comp. Science (2008), *vol.* 18, *pp.* 291–307. © 2008 Cambridge University Press
doi:10.1017/S0960129508006646 Printed in the United Kingdom

Computability and information in models of randomness and chaos†

CRISTOBAL ROJAS

LIENS, CNRS – ENS, 45 rue d'Ulm, F-75230 Paris cedex 05, France and CREA, Ecole Polytechnique, 1 rue Descartes. 75005 Paris
Email: cristobal.rojas@ens.fr

Received 16 June 2006

In Memory of Sauro Tulipani

This paper presents a short survey of some recent approaches relating two different areas, *viz*. deterministic chaos and computability. Chaos in classical physics may be approached by dynamical (equationally determined) systems or stochastic ones (as random processes). However, randomness has also been effectively modelled using recursion theoretic tools by P. Martin-Löf. We recall its connections to Kolmogorov complexity and show some applications to dynamical systems. This allows us to introduce results that connect well-established notions of entropy and algorithmic information.

1. Introduction

Science begins with observations, quantitative ones when possible, that is to say, measurements. Usually, if a measurement is repeated, a different value is obtained. Thus, several measurements have to be made in order to obtain a sequence of values (a time series), which the scientist seeks to analyse in the hope of revealing some underlying structure in the process under investigation. Mathematics, although it could be considered a rather abstract subject that is independent of physical reality, has been enormously (and intriguingly) useful in modelling and analysing natural processes in order to make successful predictions about their behaviour.

Perhaps, almost every (if not all) physical process can be viewed as a combination of a deterministic part and a random one (for example, general ideas about the development of the physical universe assume that it was originally in a state of random movement of a hot plasma and then was transformed according to the equations of quantum mechanics, through which additional randomness may appear in the observable processes).

The deterministic part expresses a physical law governing the evolution of the system under investigation and, in classical physics, is usually modelled in the framework of (topological) dynamical systems. These systems are, at least in principle, predictable,

† This paper is part of the author's ongoing doctoral work under G. Longo's supervision. It is a considerable development of part of Longo's invited lecture at the Conference, a survey of the activity in Computability and Dynamics of the team CIM at LIENS, Ecole Normale Supérieure, and CREA, Ecole Polythecnique (see http://www.di.ens.fr/users/longo/). This work was partially supported by ANR grant number 05-2 45226-0ox.

https://doi.org/10.1017/S0960129508006646 Published online by Cambridge University Press

enabling the possibility of computing future states or measurements associated with such states. But it is known that this is not always the case, as we will see later.

The random part may appear, on the one hand, when the system shows no local regularities that can be expressed by deterministic rules, but, nevertheless, there may still be global regularities, but statistical ones. In this case the system is assumed to have a random behaviour and this situation is usually modelled within classical (that is, Kolmogorov's (Kolmogorov 1933) axiomatisation of) probability theory where the system evolution is represented by a stochastic process. A qualitative understanding can be achieved, but nothing can be said about a particular state of the system at a particular time, so no predictions about a specific measurement can be made. Instead, global (statistical) properties can be predicted, such as, for example, that the average of some repeated measurement will tend to stabilise around a certain value.

On the other hand, even if the system is governed by a physical law, that is, to have a deterministic model, random behaviour may still appear. This is a consequence of the famous sensitivity to initial conditions (chaotic behaviour) and the impossibility of making measurements with infinite precision (approximation): two initial conditions that are very close to each other (so that they are indistinguishable by physical measurement) may diverge in time, making the true evolution unpredictable. One associates this practical unpredictability with a random behaviour, and, consequently, these systems allow both deterministic and probabilistic descriptions, which do not exclude but, rather, complement each other. This is the subject of ergodic theory.

Whatever the cause, randomness is identified with unpredictability, which prevents the possibility of computing future states and the associated quantities (see Bailly and Longo (2007) for a recent discussion on randomness as unpredictability in both classical and quantum frames). It is here that computability plays its role, and in a sense that will be clarified as we proceed, randomness and 'chaoticity' will be closely associated with high orders of 'non-computability'.

In this article we give an overview of some basic concepts and facts of probability, dynamical systems and computability theories, first separately (in the rest of this section), and then we concentrate in showing how computability considerations, in particular, Kolmogorov complexity (Kolmogorov 1983) and Martin-Löf randomness (Martin-Löf 1966), may be useful for both probability and dynamical systems theory. For further background, see van Lambalgen (1987), Levin (1984) and Li and Vitanyi (1997).

1.1. *Computability theory*

We shall take as primitive the notion of an algorithm that transforms finite objects into finite objects. Integer and rational (but not all real) numbers, as well as finite sequences of finite object, are examples of such objects. We shall repeatedly use the space $2^{<\omega}$ of finite binary sequences. A finite binary sequence is also called a *word* or a *string*, and denoted by w. If $w \in 2^{<\omega}$ and $i \leqslant j$, then $w_{i:j}$ denotes the string $w_i w_{i+1} \dots w_j \in 2^{<\omega}$. It is understood that an algorithm need not terminate on every input. A *partial recursive function* is a function that can be computed by an algorithm. A set is called *recursively enumerable* (abbreviated to r.e. and denoted Σ_1) if it is the range of some partial recursive function. That is, if there exists an algorithm listing the set. A set is Π_1 if its complement is Σ_1. With

https://doi.org/10.1017/S0960129508006646 Published online by Cambridge University Press

this intuitive description it is more or less clear that there exists an effective procedure to enumerate the class of all partial recursive function, associating with each of them its *Gödel number*. For formal definitions, see, for example, Rogers (1967). The connection between the informal notion of an algorithm and the formal definition of partial recursive functions is provided by *Church's Thesis*, which states that every (intuitively formalisable) algorithm is a partial recursive function.

One usually does not formally verify that some apparently recursive function is indeed recursive; instead one exhibits an algorithm that computes the function, and then Church's Thesis is invoked to guarantee that the function is in fact recursive. We shall do likewise, but in constructing algorithms to prove their existence, we shall make free use of classical logic; as a consequence, proving the existence of a recursive function does not necessarily mean that it can be effectively constructed.

Since intrinsically non-algorithmic methods are rather rare in mathematics, recursive function theory allows us to construct analogues of many concepts of classical analysis by assuming the recursive enumerability of the countable sets considered. There is, however, an important exception: the existence of universal algorithms. For example, the set of all countable sets of integers is uncountable while the set of all recursively enumerable sets is recursively enumerable. This rather abstract difference opens new analytical possibilities that have no analogues in 'non-algorithmic' analysis.

We call a function $f : 2^{<\omega} \to \mathbb{R}$ lower semicomputable if its subgraph, that is, the set $\{(w,r) \in 2^{<\omega} \times \mathbb{Q} : r < f(w)\}$ is r.e.. We call f *upper semicomputable* if $-f$ is lower semicomputable. We call f *computable* if it is lower and upper semicomputable.

1.2. *Probability theory*

In this section we will give a brief presentation of the Kolmogorov axiomatisation of probability theory (Kolmogorov 1963). Let X be a set. A *σ-algebra* of subsets of X is a collection $\mathfrak{B}$ of subsets of X satisfying the following three conditions:

(i) $X \in \mathfrak{B}$.
(ii) if $B \in \mathfrak{B}$, then $B^c \in \mathfrak{B}$.
(iii) if $B_n \in \mathfrak{B}$ for $n \geqslant 1$, then $\cup_{n=1}^{\infty} B_n \in \mathfrak{B}$.

We call the pair $(X, \mathfrak{B})$ a *measurable space*. A *finite measure* on $(X, \mathfrak{B})$ is a function $\mu : \mathfrak{B} \to \mathbb{R}^+$ satisfying $\mu(\varnothing) = 0$ 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 $\mathfrak{B}$ that are pairwise disjoint subsets of X. A *finite measure space* is a triple $(X, \mathfrak{B}, \mu)$ where $(X, \mathfrak{B})$ is a measurable space and μ is a finite measure on $(X, \mathfrak{B})$. We shall usually only consider probability spaces, which are finite measure spaces $(X, \mathfrak{B}, \mu)$ with $\mu(X) = 1$.

In this paper we will often consider the measurable space $(2^{\omega}, \mathfrak{F})$ where 2^{ω} is the space of infinite binary sequences and $\mathfrak{F}$ is the Borel σ-algebra generated by the balls (or cylinders) of the type $[w] = \{x \in 2^{\omega} : w \subset x\}$ (here $w \in 2^{<\omega}$ and $w \subset x$ means that the sequence x is an extension of the sequence w). A subset $\{w_i\}_{i=1}^{\infty}$ of $2^{<\omega}$ is called prefix free if $\{[w_i]\}_{i=1}^{\infty}$ form a sequence of pairwise disjoint cylinders.

A finite measure on $(2^{\omega}, \mathfrak{F})$ is completely determined by its values on cylinders. Thus, every function $v : 2^{<\omega} \to \mathbb{R}$ that satisfies the additivity property on prefix-free sets can

https://doi.org/10.1017/S0960129508006646 Published online by Cambridge University Press

be extended to a finite measure on $(2^\omega, \mathfrak{F})$. We will say that a measure μ on $(2^\omega, \mathfrak{F})$ is computable if its restriction to cylinders (that is, as a function on $2^{<\omega}$) is computable. A *random variable* is a *measurable* function $f : X \to \mathbb{R}^*$, which means that for each interval I, the event that yields f to take values in I, namely $f^{-1}(I)$, is a measurable event, that is, is a member of $\mathfrak{B}$. For a random variable f, we use $\mathbf{E}_\mu(f) = \int f d\mu$ to denote its expectation value. We will drop the index μ when it is clear from the context.

The premise for the use of probabilistic methods is the assumption that the result x of a physical process arises randomly with probability distribution μ. This μ is discovered or hypothesised by, for example, analogy with other processes and statistical data about them, considerations of symmetry, and so on. Then, those properties of x that are satisfied with μ-probability 1 (or close to 1 in the finite case) are taken as probabilistic laws. For example, if ω is the sequence of outputs in a μ-distributed experiment that is independent and identically repeated (that is, $\mu(x_1, \ldots, x_n) = \mu'(x_1)\mu'(x_2)\ldots\mu'(x_n)$), then, for each property B, the law of large numbers tells us that, with μ-probability close to 1, the frequency of $B(\omega_i)$ realisation is close to the probability $\mu'(B)$. In any case, subjecting ω to such laws is predicted, that is, probabilistic laws are those properties that should be satisfied by random objects (with respect to some distribution).

The problem is that in this absolute sense, random objects do not exist: jointly the properties of probability 1 have probability 0! So, we cannot predict the realisation of all of them simultaneously. We should just choose a few of them (the standard ones). However, classical probability theory contains no principles that would allow the distinction of such standard laws from others. In Martin-Löf (1966), Martin Löf proposed a criteria for choosing the class of standard laws, namely those laws that can be proved in an effective way.

1.3. *Dynamical systems*

We will now say a few words about the standard way to model both deterministic and stochastic physical systems within the same framework, that of ergodic theory. The set of all possible *states* of the system will be denoted by X, which will usually be a metric space. A *point* $x \in X$ is supposed to give all possible information about the system; for example, absolutely perfect knowledge of the position and momentum of every particle constituting the system. Since such perfect information is unattainable, one only assumes one is able to know whether or not x is consistent with some *observable event* B that is a subset of X, that is, whether or not $x \in B$. To permit countable-infinite set-theoretic operations, one supposes 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 modelled by a random variable $f : X \to \mathbb{R}$.

Now we will describe dynamics. Development in time is modelled by a *measure preserving* transformation $T : X \to X$. This is a function that 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 the 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

https://doi.org/10.1017/S0960129508006646 Published online by Cambridge University Press

change in time. A typical picture is as follows: an initial point is chosen according to a 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 takes place, which can be quite complicated; the 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 the long term to quite different particular behaviours, though they are identical in a qualitative sense. An invariant measure is called *ergodic* if for all $B \in \mathfrak{B}$ such that $T^{-1}(B) = B$, we have $\mu(B) = 0$ or $\mu(B^c) = 0$.

Ergodic theory is primarily concerned with understanding 'essentially different' measure preserving transformations, where two transformations are considered to be essentially the same if they are *isomorphic* (see Petersen (1983) for a formal definition). The approaches to the isomorphism problem usually involves searching for isomorphism invariant quantities, which are useful in distinguishing and classifying transformations.

If we consider that access to the system is given by a finite measurable partition $\alpha = \{A_1, \ldots, A_n\}$ (a partition where the sets are measurable) that represents actually observable quantities, then we can associate with (X, μ, T) a *symbolic dynamical system* $(X_\alpha, \mu_\alpha, \sigma)$ (called the symbolic model of (X, μ, T)). The set X_α is a subset of $\{1, 2, \ldots, n\}^{\mathbb{N}}$. We associate with a point $x \in X$ an infinite sequence $\omega = (\omega_i)_{i \in \mathbb{N}} = \phi_\alpha(x)$ defined by

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

The transformation $\sigma : X_\alpha \to X_\alpha$ is *the shift* defined by $\sigma((\omega_i)_{i \in \mathbb{N}}) = (\omega_{i+1})_{i \in \mathbb{N}}$. Since α is a partition, the set $\phi_\alpha(x)$ will contain only one element, and this defines a function associating an infinite string with a point $x \in X$. The measure μ induces a measure μ_α in the associated symbolic dynamical system. Since the map ϕ_α is measurable, the measure μ_α can be defined by $\mu_\alpha(B) = \mu(\phi_\alpha^{-1}(B))$ for all measurable $B \subset X_\alpha$. Moreover, if (X, μ, T) is ergodic, then $(X_\alpha, \mu_\alpha, \sigma)$ is ergodic and $h_\mu(T, \alpha)$ equals $h_{\mu_\alpha}(\sigma)$ where h_v denotes metric entropy, which we now define.

Given a finite invariant dynamical system (X, T, μ) and a finite measurable partition $\alpha = \{A_1, \ldots, A_N\}$ of X, the *entropy* $h_\mu(\alpha)$ of the partition α is defined by

$$h_\mu(\alpha) = -\sum_{i=1}^{N} \mu(A_i) \log \mu(A_i)$$

with the assumption that $0 \log 0 = 0$. Now, $h_\mu(\alpha)$ is a measure of the finesse and homogeneity of the partition α with respect to μ.

Let $T^{-1}\alpha$ be the partition given by the counter images $T^{-1}A_i$. Then let

$$\alpha_n = \alpha \vee T^{-1}\alpha \vee T^{-2}\alpha \vee \ldots \vee T^{-n}\alpha$$

be the partition given by the sets of the form

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

varying A_{i_j} among all the sets of α. Knowing to which set of the partition α_n a point x belongs corresponds to knowing which sets of the partition α the orbit of x visits up to time $n-1$.

https://doi.org/10.1017/S0960129508006646 Published online by Cambridge University Press

The *metric entropy* $h_\mu(T, \alpha)$ *relative to the partition* α is defined as the limit

$$h_\mu(T, \alpha) = \lim_{n\to\infty} \frac{1}{n} h_\mu(\alpha_n).$$

Note that this limit is an infimum, since the sequence on the right-hand side is decreasing. Finally, we define the *metric entropy* $h_\mu(T)$ of (X, T, μ) by

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

A fundamental property of the metric entropy is its invariance under isomorphisms. We shall see more on this in Section 3.

2. Martin-Löf randomness

In Martin-Löf (1966), Martin-Löf proposed a canonical choice for the class of probabilistic laws: the class of probabilistic laws that can be proved effectively. To give a better explanation of this notion, we shall look at the proofs of probabilistic laws.

According to the usual interpretation, a probabilistic law with respect to a measure μ is a statement about some formula ϕ of the form $\mu\{x \in 2^\omega : \phi(x)\} = 1$, or, equivalently, $\mu\{x \in 2^\omega : \phi(x)\}^c = 0$. Typically, in order to prove such a statement, one constructs a sequence (O_n) of open sets such that each O_n contains the set $\{x \in 2^\omega : \phi(x)\}^c$ and $\mu(O_n) \leqslant 2^{-n}$ (or any other function of n that decreases to 0).

Roughly speaking, a probabilistic law is effective if the sequence (O_n) used in the proof can be constructed in an algorithmic way. Formally, we have the following definition.

Definition 2.1. Let μ be a computable measure on 2^ω. A subset N of 2^ω is *an effective μ-null set* if:

(i) N is of the form $\cap_n O_n$, where the O_n are recursively enumerable uniformly in n.
(ii) $O_{n+1} \subseteq O_n$.
(iii) $\mu(O_n) \leqslant 2^{-n}$.

When we say a sequence (O_n) is recursively enumerable uniformly in n, we mean that there exists an algorithm that on input n generates the set O_n. It is not difficult to see that for every recursively enumerable subset $A \subseteq 2^\omega$, one can effectively determine a recursively enumerable sequence of disjoint cylinders $[w_i]$ such that $A = \cup_i [w_i]$. So, to generate a set O_n for an algorithm means to enumerate the prefix-free sequence $(w_i)_{i\mathbb{N}} \subset 2^{<\omega}$.

Let $\mathscr{N}_\mu$ be the collection of all effective μ-null sets. For each $N \in \mathscr{N}_\mu$ we have $\mu(N) = 0$ by definition. Since $\mathscr{N}_\mu$ is countably infinite, we have

$$\mu\left(\bigcup_{N\in\mathscr{N}_\mu} N\right) = 0,$$

and we call the set $U_\mu = \cup_{N\in\mathscr{N}_\mu} N$ the *maximal μ-null set*. Moreover, we have the following central result.

Theorem 2.1 (Martin-Löf). Let μ be a computable measure. The collection $\mathscr{N}_\mu$ of all effective μ-null sets is recursively enumerable, and thus the set $U_\mu = \cup_{N\in\mathscr{N}_\mu} N$ is an effective μ-null set.

https://doi.org/10.1017/S0960129508006646 Published online by Cambridge University Press

We can now state the definition of a (Martin-Löf) random sequence.

Definition 2.2 (Martin Löf randomness). Let μ be a computable measure. $x \in 2^\omega$ is *random* with respect to μ (denoted $x \in R(\mu)$) if for all effective μ-null sets N, $x \notin N$, that is, $x \notin U_\mu$

It can be shown that most of the laws of probability (such as the law of large numbers or the law of the iterated logarithm) can indeed be proved by constructing μ-null sets covering the set of sequences that do not satisfy these laws. So, typical (random) sequences in the sense of Martin-Löf share the common laws of probability.

In particular, since the set of μ-random infinite sequences has μ-measure one, this definition allows us to replace the standard formulation of probability theory, 'ϕ **holds with probability one**', by the more effective 'ϕ **holds for each (Martin Löf) random infinite sequence** ω' for all effective predicates ϕ.

There is another way to express the typicality of an infinite sequence with respect to a given probability distribution. Suppose we have a function $t : 2^\omega \to \mathbb{R}^+ \cup \{\infty\}$ with finite expectation, say $\mathbf{E}_\mu(t) = \int t d\mu \leqslant 1$. If $x \in 2^\omega$ appears at random (with respect to the probability distribution μ), we must have $\mu\{x \in 2^\omega : t(x) = \infty\} = 0$ (characterising μ-null sets), and we may expect t to take larger values than $\mathbf{E}(t)$ with low probability only, that is, for all $\epsilon > 0$ there exists a sufficiently large m such that $\mu\{x \in 2^\omega : t(x) > m\} < \epsilon$. Hence, the function t may be seen as a test for the typicality of a sequence. To get a proper definition, we must add some effective structure to the test function t.

Definition 2.3. Let μ be a computable probability measure. A total function $t : 2^\omega \to \mathbb{N} \cup \{\infty\}$ of the form $t(x) = \sup_{n\in\mathbb{N}}\{\delta(x_{1:n})\}$ is an *effective sequential μ-test* if:

(i) $\delta : 2^{<\omega} \to \mathbb{N}$ is a lower semi-computable function; and

(ii) $\mu\{x \in 2^\omega : t(x) > m\} < 2^{-m}$, for each $m \geqslant 0$.

If x is such that $t(x) < \infty$, we will say that x *passes* the test t.

The relationship between effective sequential tests and effective null sets is given by the following lemma.

Lemma 2.1. Let μ be a computable probability measure. N is an effective μ-null set if and only if $N = \{x \in 2^\omega : t(x) = \infty\}$ for some effective sequential μ-test t.

Proof. Let $t(x) = \sup_{n\in\mathbb{N}}\{\delta(x_{1:n})\}$ be an effective sequential μ-test. Since δ is lower semi-computable, the set $\{x \in 2^\omega : t(x) > m\} = \{x \in 2^\omega : \exists n\ \delta(x_{1:n}) > m\}$ is a recursively enumerable one, and the μ-null set $\{x \in 2^\omega : t(x) = \infty\}$ may be written as

$$\{x \in 2^\omega : \forall m \exists n \delta(x_{1:n}) > m\} = \cap_{m\in\mathbb{N}}\{x \in 2^\omega : \exists n \delta(x_{1:n}) > m\},$$

so it is an effective μ-null set.

For the converse, given an effective μ-null set $N = \cap_{m\in\mathbb{N}} \cup_{j\in\mathbb{N}} [f(m, j)]$, we define the function δ by

$$\delta(w) = \begin{cases} m+1 & \text{if } m \text{ is the last integer for which } w = f(m, j) \text{ for some } j \\ 0 & \text{otherwise.} \end{cases}$$

Such a δ is lower semi-computable and $\{x \in 2^\omega : \exists n\ \delta(x_{1:n}) > m\} = \{x \in 2^\omega : \exists n\ \exists j\ x_{1:n} = f(m, j)\} = \cup_{j\in\mathbb{N}}[f(m, j)]$, thus $\mu\{x \in 2^\omega : t(x) > m\} < 2^{-m}$. □

https://doi.org/10.1017/S0960129508006646 Published online by Cambridge University Press

We now state the more classical definition of Martin-Löf random sequences as a theorem.

Theorem 2.2. An infinite sequence $\omega \in 2^\omega$ is (Martin-Löf) random with respect to a computable probability measure μ if and only if it passes all effective sequential μ-tests.

Proof. The statement follows directly from the lemma. □

The existence of a maximal effective μ-null set entails the existence of an *universal effective sequential test*, that is to say, an effective sequential test f such that for each effective sequential μ-test t_i there exists a constant c such that for all $x \in 2^\omega$ we have $f(x) \geqslant t_i(x) - c$. We denote this universal test f_μ. An infinite sequence ω is random with respect to μ if and only if $f_\mu(\omega) < \infty$.

2.1. *Application to ergodic theory*

Let $(\Omega, \mathscr{F}, \mu)$ be a probability space, T be a transformation preserving the measure μ, and f be an integrable real function.

As in Section 1.3, the time development of the system starting from state x at time 0 is given by the orbit $O(x) = \{x, Tx, T^2x, \ldots\}$. The scientist seeking to measure an observable quantity f obtains the sequence of numbers $\{f(x), f(Tx), f(T^2x), \ldots\}$ and may compute the sequence of averages

$$A_n f(x) = \frac{1}{n} \sum_{k=0}^{n-1} f(T^k x).$$

The existence of the limit of these averages as $n \to \infty$ is a question of fundamental importance for science. If the limit exists, it represents a mean or average value of the observable f that can carry some information about the underlying structure of the system. For example, if for a set $A \subset X$ we take the function $f_A(x) : X \to \{0, 1\}$ defined by $f_A(x) = 1 \Leftrightarrow x \in A$, then $\lim_{n\to\infty} A_n f_A(x)$ represents the fraction of time that the orbit of x spends on A.

Birkhoff's ergodic theorem, in its general form, is as follows.

Theorem 2.3 (Birkhoff). Let μ be a probability measure, T be a transformation preserving the measure μ, and f be a real function on Ω. Then there exists an integrable function $\bar{f}$ such that $\mathbf{E}(\bar{f}) = \mathbf{E}(f)$ and

$$\lim_{n\to\infty} A_n f(\omega) = \bar{f}(\omega) \tag{1}$$

μ-almost surely. In addition, $\bar{f}(T\omega) = \bar{f}(\omega)$. If the transformation T is ergodic, then $\bar{f}(\omega) = \mathbf{E}(f)$ μ-almost surely.

As mentioned earlier, many results in probability theory are of the form '$\phi(\omega)$ holds almost surely', were $\phi(\omega)$ is some property. So, in a sense, the ergodic theorem embodies many of them, all taking the form in (1). For example, if the transformation T is the shift

https://doi.org/10.1017/S0960129508006646 Published online by Cambridge University Press

over 2^ω (so the measure μ is stationary) and we set $f(\omega_1\omega_2\ldots) = \omega_1$, then we have

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} f(T^k\omega) = \lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} \omega_k \tag{2}$$

in (1), that is, the frequency of 1's in ω converges to $\mu([1])$ almost surely.

Unlike many other assertions of probability theory, which hold almost surely, the well-known proof of this theorem cannot effectively define the class of ω for which (1) holds. V'Yugin (1997) uses a more 'constructive' proof in order to prove its 'individual counterpart'. We must additionally assume the computability of the measure μ, transformation T and function f. Computability for the function f means that there is an algorithm that uses as inputs a rational ϵ and a sequence $\omega \in 2^\omega$ and, after a finite number of steps, this algorithm outputs a rational number r such that $|f(\omega) - r| < \epsilon$. Note that this algorithm uses only an initial finite part of the sequence ω. Computability for the transformation T means that there is an algorithm that takes a sequence ω sequentially bit by bit, processes it, and then produces an output sequence sequentially bit by bit also. The shift is the simplest example of a computable transformation. See V'Yugin (1997) for a more formal exposition.

Theorem 2.4 (V'Yugin). Let μ be a computable measure, T be a computable transformation preserving the measure μ, and f be a computable real function on Ω. Then there exists an integrable function $\bar{f}$ such that $\mathbf{E}(\bar{f}) = \mathbf{E}(f)$ and

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} f(T^k\omega) = \bar{f}(\omega) \tag{3}$$

for any typical (random in the sense of Martin-Löf) sequence ω with respect to the measure μ. In addition, $\bar{f}(T\omega) = \bar{f}(\omega)$. If the transformation T is ergodic, then $\bar{f}(\omega) = \mathbf{E}(f)$ for this sequence ω.

3. Kolmogorov complexity

There are various ways in which the intuition behind the definition of Kolmogorov complexity (Kolmogorov 1983) can be stated. Consider the following strings, which we suppose to be generated by a random process such as coin tossing:

$$\begin{aligned} x &= 000000000000000000000000 \\ y &= 001011001011001011001011 \\ z &= 100100110110001110110100 \end{aligned}$$

Under the hypothesis of random generation, the perfect regularity of x appears very extraordinary to us, and it may well not be accepted as a random outcome. A little less evident is the regularity of the second string y, which is also periodic. The third string z does not seem to have any obvious regularity, and it may well be accepted as a random outcome. However, probability theory assigns the same probability $P = 2^{-24}$ to all of them, and allows no distinction.

https://doi.org/10.1017/S0960129508006646 Published online by Cambridge University Press

In Gács (1993), Peter Gács makes the following remark:

'... 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.' (P.S. Laplace, *A philosophical Essay on Probabilities*, Dover, 1952. Originally published in 1819.)

But how can we formalise such a distinction between regular and irregular sequences? Laplace also makes a distinction 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.' (P.S. Laplace, *Ibid.*)

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.' (Lévy 1925)

So, we can expect regular sequences to be produced by causes other than chance. Being able to find such a cause, that is, to have a theory (or a model) explaining the regularities of large data, means being 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 $\mathscr{X}$ be a class of finite objects, and $\mathscr{C}$ be a class of finite words (the codes). A minimal description length function corresponds to a (surjective) decoding procedure $f : \mathscr{C} \to \mathscr{X}$ such that for all other decoding procedures $g : \mathscr{C} \to \mathscr{X}$ and for all object $x \in \mathscr{X}$, f satisfies an optimality property of the form

$$\min_{c \in \mathscr{C}} \{|c| : f(c) = x\} \leqslant \min_{c \in \mathscr{C}} \{|c| : g(c) = x\}.$$

If we let the class of decoding procedures be the class of all functions from $\mathscr{C}$ to $\mathscr{X}$, then there is no such a minimal procedure; the natural restriction to make here is to consider only algorithmic procedures.

Furthermore, the decoding procedure should be able to separate different codes that have been concatenated in a single word without using any other information besides 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

https://doi.org/10.1017/S0960129508006646 Published online by Cambridge University Press

contain all necessary information to reconstruct the coded finite object algorithmically, the Kolmogorov complexity is also called the *Algorithmic Information Content*. We prefer to use the latter term, since, as we will see later, random objects have a very high information content. But randomness is the absence of any 'structure' and 'organisation', so, in a sense, random objects should not be very complex.

Definition 3.1. A *prefix algorithm* is a partial recursive function $A : 2^{<\omega} \to 2^{<\omega}$ that has a prefix-free domain. Let $\mathscr{A}$ be the set of prefix algorithms.

Definition 3.2. Let $A : 2^{<\omega} \to 2^{<\omega} \in \mathscr{A}$ be a prefix algorithm. The *complexity* (or *Information Content*) $I_A(w)$ of $w \in 2^{<\omega}$ is defined to be

$$I_A(w) = \begin{cases} \infty & \text{if there is no } p \text{ such that } A(p) = w \\ |p| & \text{if } p \text{ is a shortest input such that } A(p) = w. \end{cases}$$

To get the promised minimality property, we require the existence of a universal prefix algorithm. At first sight it might seem that there could not be such an algorithm since the set of Gödel numbers of prefix algorithms is Π_1. But, despite this, a recursive enumeration of the set of prefix algorithms does exist.

Theorem 3.1. $\mathscr{A}$ is recursively enumerable.

Proof. We construct an algorithm P that transforms any number e into a Gödel number $P(e) = e^{A_{P(e)}}$ for a prefix algorithm $A_{P(e)}$ and such that every prefix algorithm has at least one Gödel number in the range of P. Given e, P generates the domain of the partial recursive function with Gödel number e, ϕ_e. A partial recursive function $A_{P(e)}$ with Gödel number $P(e)$ is determined as follows: $A_{P(e)}(q)$ equals $\phi_e(q)$ except for those $q \in dom\phi_e$ that 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 get the existence of a universal prefix algorithm.

Theorem 3.2. There exists a universal prefix algorithm that is asymptotically optimal, that is, a $U \in \mathscr{A}$ such that $\forall\ A \in \mathscr{A}\ \exists\ c_A \in \mathbb{N}$, which depends on A, such that $\forall\ p \in 2^{<\omega}\ \exists\ q \in 2^{<\omega}$ that satisfies $U(q) = A(p)$ 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 $p \in 2^{<\omega}$ is an input, then $q = 0^{e^A}1p$ is such that $U(q) = A(p)$ and $|q| = |p| + e^A + 1$. □

We fix an asymptotically optimal universal prefix algorithm U and let $I(w) = I_U(w)$. We can now state the *invariance theorem*, which gives the optimality property.

Theorem 3.3. $\forall A \in \mathscr{A}\ \exists\ c_A \in \mathbb{N}$ such that $\forall w \in 2^{<\omega}$ we have $I(w) \leqslant I_A(w) + c_A$.

Proof. Let p^* be a shortest input such that $A(p^*) = w$. Then $q = 0^{e^A}1p^*$ is such that $U(q) = w$, and thus $I(w) \leqslant I_A(w) + c_A$, where $c_A = e^A + 1$. □

https://doi.org/10.1017/S0960129508006646 Published online by Cambridge University Press

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, have 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 of a self-delimiting code that contains both the word w itself and the information about its own length. In fact, we have the following easy first upper bound.

Lemma 3.1. For some constant c and all w: $I(w) \leqslant |w| + I(|w|) + c$.

Note that to know the length of w, say n, is to know that w belongs to a specific set much smaller than $2^{<\omega}$, namely the set $S = \{w : |w| = n\}$ whose cardinality is 2^n. Then we can specify w with its position in an enumeration of this set, a number that can be represented in less than $\log(2^n) = n$ bits. In general, any information about a specific set to which w belongs can be used to code it in a shorter way. There is another well-known extra piece of information that can help us find short codes, *viz.* knowledge of an underlying probability distribution. The idea here is to assign shorter codes to words that have high probability. To illustrate this, consider a product probability distribution that assigns a high probability to 1 and a low probability to 0. Consequently, typical finite words should have a small number of 0's. If we enumerate these words in an increasing way with respect to the number of 0's, typical words should have a low index in this enumeration, so by specifying these indices we can give short codes. In fact, not knowing the underlying distribution corresponds to the case where the above technique is useless: the uniform case, where the probability of any n-length word is 2^{-n} and where we have $|w| = -\log(probability(w))$. More generally, we have the following theorem.

Theorem 3.4. Let μ be a computable measure on 2^ω. Then for some constant c and all w we have $I(w) \leqslant -\log \mu[w] + I(|w|) + c$

We now state the promised relation between randomness and incompressibility for infinite sequences, which is expressed as the incompressibility of their initial segments – see Chaitin (1987) for a proof.

Theorem 3.5 (Chaitin, Levin). Let μ be a computable measure. Then ω is a (Martin Löf) random sequence with respect to μ if and only if $\exists m\ \forall n\ I(\omega_{1:n}) > -\log \mu[\omega_{1:n}] - m$.

3.1. *Application to dynamical systems*

A sensitivity to initial conditions entails the unpredictability of a system. Invariant quantities capable of distinguishing its presence, such as Lyapunov exponents or entropy, are suitable for the identification and classification of these systems. These indicators can be seen in different ways:

— as a measure of the speed of separation of initially nearby orbits;
— as the average uncertainty removed from the system when observations are performed; or
— as the growing rate of precision (in the knowledge of the initial condition) needed to predict the evolution of the system in later and later times.

https://doi.org/10.1017/S0960129508006646 Published online by Cambridge University Press

We can also view both removed uncertainty and incremented precision as gained information about the corresponding state of the system. That is, to predict the system state at later and later times requires more and more information about the initial condition. Hence, the asymptotic behaviour of the relationship between the time of prediction and the information needed should give such an indicator. But how can we measure the information content of a single state in phase space, which is usually represented by a point in a metric space?

We have seen that Algorithmic information theory gives a rigorous and powerful tool for defining the amount of information contained in a binary string. This theory has been extended to more general spaces (Brudno 1983; Gács 1993; Galatolo 1999; Galatolo 2000). Following Galatolo, we will show how this theory can be translated to the dynamical systems framework, and be used to obtain a meaningful indicator of the information content of points in metric spaces, which is called *pointwise information*. Measuring the growth rate of the information needed to specify the system evolution to a certain time, we get another indicator of sensitivity, which is called *algorithmic entropy*.

3.1.1. *Computable structures.* For the definition of pointwise information, or, for short, *P-information*, we must choose an additional structure on the metric space, which is called a computable structure. Roughly speaking, a computable structure is a class I of dense immersions of the space of finite strings $2^{<\omega}$ in a metric space (X, d) such that the distance d restricted to $I(2^{<\omega})$ is a computable function. Computable structures allow us to interpret finite strings as points (or sequences of points) of the metric space. From now on we let (X, d) be a separable metric space. We begin with some definitions and results. The proofs can be found in Galatolo (2000; 1999).

Definition 3.3. An interpretation function on (X, d) is a function $I : 2^{<\omega} \to X$ such that $I(2^{<\omega})$ is dense in X and there exists a total recursive function $D : 2^{<\omega} \times 2^{<\omega} \times \mathbb{N} \to \mathbb{Q}$ such that $\forall s_1, s_2 \in 2^{<\omega}, n \in 2^{<\omega}$ we have

$$|d(I(s_1), I(s_2)) - D(s_1, s_2, n)| \leqslant \frac{1}{2^n}.$$

Definition 3.4. Let I_1 and I_2 be two interpretations in (X, d). We say that I_1 and I_2 are equivalent, denoted $I_1 \sim I_2$, if there is a total recursive function $D^* : 2^{<\omega} \times 2^{<\omega} \times \mathbb{N} \to \mathbb{Q}$ such that $\forall s_1, s_2 \in 2^{<\omega}, n \in \mathbb{N}$ we have

$$|d(I_1(s_1), I_2(s_2)) - D^*(s_1, s_2, n)| \leqslant \frac{1}{2^n}.$$

Lemma 3.2. The relation $\sim$ is an equivalence relation.

Definition 3.5. A computable structure $\mathscr{I}$ on X is an equivalence class $[I]_\sim$ of interpretations in X.

There are natural choices of computable structures for many concrete metric spaces used in analysis or geometry. For example, the space $\mathbb{R}^n$ has as a natural computable structure the equivalence class of interpretations containing some interpretation of $\mathbb{Q}^n$. In separable functions spaces there are structures containing polynomials with rational coefficients, and so on.

https://doi.org/10.1017/S0960129508006646 Published online by Cambridge University Press

Morphisms between metric spaces with computable structures must preserve computability. We have the following definition.

Definition 3.6. Let $(X, d, \mathscr{I})$ and $(Y, d', \mathscr{J})$ be spaces with computable structures. A function $\Psi : X \to Y$ is said to be a *morphism of computable structures* if Ψ is continuous and for each pair $I \in \mathscr{I}, J \in \mathscr{J}$ there exists a total recursive function $D^* : 2^{<\omega} \times 2^{<\omega} \times \mathbb{N} \to \mathbb{Q}$ such that $\forall s_1, s_2 \in 2^{<\omega}, n \in \mathbb{N}$ we have

$$|d'(\Psi(I(s_1), J(s_2)) - D^*(s_1, s_2, n)| \leqslant \frac{1}{2^n}.$$

Note that if Ψ is a morphism between X and Y, then, given a point $x \in I(2^{<\omega}) \subset X$, it is possible to find in an algorithmic way a point $y \in J(2^{<\omega}) \subset Y$ as near as we want to $\Psi(x)$. Note also that Ψ is not required to have dense image, so $\Psi \circ I$ is not necessarily an interpretation function equivalent to J.

3.1.2. *Pointwise information and algorithmic entropy.* The choice of a computable structure on X allows us to interpret the output of a computation as a point of X and to define the notion of the Information Content of a point of the space at a given accuracy.

We will use $K(s)$ to denote the Kolmogorov complexity of a word $s \in 2^{<\omega}$.

Definition 3.7. The *P-information of a point* $x \in X$ at accuracy $\lambda\ (> 0) \in \mathbb{Q}$ with respect to the interpretation $I \in \mathscr{I}$ is defined to be

$$\mathscr{P}^I_\lambda(x) = \min\{K(s) : s \in 2^{<\omega}, d(I(s), x) < 2^{-\lambda}\}.$$

To define the Information Content of a finite sequence in X, we need a way to interpret computations as a finite sequence. Let us consider a total recursive surjective function

$$\mathscr{Q} : 2^{<\omega} \to (2^{<\omega})^*$$

where $(2^{<\omega})^*$ is the set of finite sequences in $2^{<\omega}$. For each word $s \in 2^{<\omega}$, we define

$$U(s) = (U_1(s), U_2(s), \ldots U_{|U(s)|}(s)) = I(\mathscr{Q}(s)) \in X^*,$$

where X^* is the set of finite sequences in X (that is, $U_i(s) \in X$) and I has been extended in the obvious way.

Definition 3.8. The *P-information of a finite sequence* of points $(x_1, \ldots, x_n) \in X^*$ at accuracy $\lambda\ (> 0) \in \mathbb{Q}$ with respect to the interpretation $I \in \mathscr{I}$ is defined to be

$$\mathscr{P}^I_\lambda(x_1, \ldots, x_n) = \min\{K(s) : s \in 2^{<\omega}, \max_i(d(U_i(s), x_i)) < 2^{-\lambda}\}.$$

Now, consider

$$\underline{\mathscr{S}}_\lambda(T, X, I, x) = \liminf_{m\to\infty} \frac{\mathscr{P}^I_\lambda(x, T(x), T^2(x), \ldots, T^m(x))}{m}$$

$$\overline{\mathscr{S}}_\lambda(T, X, I, x) = \limsup_{m\to\infty} \frac{\mathscr{P}^I_\lambda(x, T(x), T^2(x), \ldots, T^m(x))}{m}.$$

The functions $\underline{\mathscr{S}}_\lambda, \overline{\mathscr{S}}_\lambda$ are non-decreasing in λ and, intuitively, they provide an estimate of the mean information needed to describe a step of the orbit of a point with accuracy

https://doi.org/10.1017/S0960129508006646 Published online by Cambridge University Press

$2^{-\lambda}$ and interpretation I. The value of $\underline{\mathscr{S}}_\lambda$ and $\overline{\mathscr{S}}_\lambda$ does not depend on the choice of the function $\mathscr{Q}$: if we change it, the values of $\mathscr{P}_\lambda$ remain unchanged up to a constant that vanishes in the limit. Since $\underline{\mathscr{S}}_\lambda$ and $\overline{\mathscr{S}}_\lambda$ are non-decreasing, the limits

$$\begin{aligned}\overline{\mathscr{S}}(T,X,I,x) &= \lim_{\lambda\to\infty} \overline{\mathscr{S}}_\lambda(T,X,I,x)\\ \underline{\mathscr{S}}(T,X,I,x) &= \lim_{\lambda\to\infty} \underline{\mathscr{S}}_\lambda(T,X,I,x)\end{aligned}$$

exist.

Lemma 3.3. If the interpretation function I_1 is equivalent o I_2, then $\forall x \in X$

$$\begin{aligned}\underline{\mathscr{S}}(T,X,I_1,x) &= \underline{\mathscr{S}}(T,X,I_2,x)\\ \overline{\mathscr{S}}(T,X,I_1,x) &= \overline{\mathscr{S}}(T,X,I_2,x).\end{aligned}$$

If a computable structure $\mathscr{I}$ is defined on the metric space X, this lemma allows us to define $\underline{\mathscr{S}}(T,X,\mathscr{I},x)$ and $\overline{\mathscr{S}}(T,X,\mathscr{I},x)$. That is the algorithmic entropy of the orbit of x with respect to $\mathscr{I}$.

Definition 3.9. If X is a metric space with a computable structure $\mathscr{I}$, $I \in \mathscr{I}$, then the algorithmic entropy of x with respect to $\mathscr{I}$ is defined by

$$\begin{aligned}\overline{\mathscr{S}}(T,X,\mathscr{I},x) &= \overline{\mathscr{S}}(T,X,I,x)\\ \underline{\mathscr{S}}(T,X,\mathscr{I},x) &= \underline{\mathscr{S}}(T,X,I,x).\end{aligned}$$

If X is a metric space with a computable structure, the algorithmic entropy is invariant under maps that are morphisms of the computable structures. So it may be useful for the classification problem. Note that this notion makes sense in every metric space, which need not necessarily be compact.

However, if X is compact, algorithmic entropy is equivalent to an older notion introduced by Brudno, which is called orbit complexity (Brudno 1983), and, as a consequence, the choice of a computable structure is not important: the next theorem shows that if a computable structure exists on a compact X, then $\mathscr{S}(T,X,\mathscr{I},x)$ is independent of the choice of $\mathscr{I}$. Moreover, it is equal to the metric entropy of the system for almost all points.

Theorem 3.6. Let (X,T) be a dynamical system. If X is compact and has a computable structure $\mathscr{I}$, and μ is an ergodic probability measure on (X,T), then

$$\overline{\mathscr{S}}(T,X,\mathscr{I},x) = \underline{\mathscr{S}}(T,X,\mathscr{I},x) = h_\mu(T)$$

for μ-almost each $x \in X$, where $h_\mu(T)$ is the metric entropy of (X,T) with respect to the invariant measure μ.

This result confirms, as we expected, that there is a deep connection between the amount of information as measured by Kolmogorov complexity and disorder degrees, for which a high value entails unpredictability.

https://doi.org/10.1017/S0960129508006646 Published online by Cambridge University Press

4. Conclusions

In this article we have given an overview of some basic but important concepts and results of computability, probability and dynamical systems theory.

In Section 2 we showed how to use computability theory to formalise the notion of an individual random sequence with respect to some measure μ. Namely, a sequence is μ-random if it satisfies all 'effective' (in an algorithmic sense) properties with μ-probability one.

Or, as we said in the introduction, randomness is usually identified with unpredictability, which prevents the possibility of computing future outcomes. Nevertheless, it is not true to say that μ-random sequences are always uncomputable. For instance, it is not dificult to see that every computable sequence ω is random with respect to a computable measure μ for which the sequence is an atom ($\mu(\{\omega\}) > 0$). On the other hand, for less rigid measures, the situation is radically different. For example, if a sequence is random with respect to a Bernoulli distribution, the sequence is highly uncomputable: it does not contain any computable subsequences.

In this sense, Martin-Löf randomness expresses algorithmic *typicality* or *genericity* with respect to a measure underlying some process rather than algorithmic *unpredictability*. Of course, for most measures, unpredictability is indeed a property that typical objects share.

In Section 3 we showed that Martin-Löf randomness can be characterised using Kolmogorov complexity, thereby formalising the intuitive relationship between randomness and irregularity. Following Galatolo, Kolmogorov complexity was extended and used to estimate the algorithmic information content of individual points in metric spaces. Following Theorem 3.5, it seems natural to extend the randomness condition to points in general metric spaces. We could say, for example, that a point $x \in X$ is random with respect to μ if there exits a constant c such that for all $n \in \mathbb{N}$ we have $\mathscr{P}_n^I(x) \geqslant -\log \mu(B(x, 2^{-n})) - c$.

In addition, effective null sets in general tological spaces can also be defined within the framework of computable analysis (Hertling and Weihrauch 1998). Here, a topological space is said to be effective if it has an enumerable basis given in an effective way. An open set A is recursively enumerable if the index set of basis elements included in A is recursively enumerable. In this framework, effective null sets (or tests) are countable intersections of uniformly recursively enumerable open sets whose measure tends to zero in an algorithmically controlled way. The existence of universal tests depends on the computability condition for μ, and general theorems for the conservation of randomness under computable transformations can be stated.

Many questions arise from all this. It is natural to look for possible rigorous links between randomness and the information content of points in general metric spaces and to decide whether μ-almost everywhere results hold for each individual random point. In particular, we conjecture a general individual ergodic theorem asserting the average convergence of observed quantities associated with a single random trajectory.

On the other hand, it would be interesting to investigate the concequences that algorithmic unpredictability could have on the dynamics. More precisely, if we suppose the access to a dynamical system (X, T) is given through a finite partition $\alpha = \{A_1, \ldots, A_n\}$ of phase space X whose elements represent physically observable (distinguishable) sets,

https://doi.org/10.1017/S0960129508006646 Published online by Cambridge University Press

and if we construct the symbolic model (X_α, σ) of the system (X, T) as described in the introduction (without fixing any particular measure), then for each orbit of the system there corresponds an infinite sequence in X_α to which we would attribute some (un)predictability condition. This sequence represents a series of measurements performed on the system, and we could ask, for example, for it to be random with respect to some 'natural' measure, and then see which properties of the dynamics can be deduced. We guess, for instance, that if the sequence generated by a *typical* initial condition is random with respect to some generalised Bernoulli measure, then there exits some invariant measure in X with respect to which the metric entropy of the entire system is positive.

References

Bailly, F. and Longo, G. (2007) Randomness and Determination in the interplay between the Continuum and the Discrete (to appear). (Downloadable as revised English version of the appendix to the book: Bailly, F. and Longo, G. (2006) *Mathématiques et sciences de la nature. La singularité physique du vivant*, Hermann, Paris.)

Birkhoff, G. D. (1931) Proof of the ergodic theorem. *Proc. Nat. Acad. Sci.* **17** 656–660.

Brudno, A. A. (1983) Entropy and the complexity of the trajectories of a dynamical system. *Trans. Mosc. Math. Soc.* **44** 127–151.

Chaitin, G. (1987) *Algorithmic Information Theory*, Cambridge University Press.

Kolmogorov, A. N. (1933) Grundbegriffe der Wahrsceinlichkeitsrechung. *Ergebnisse der Mathematik und ihrer Grenzgebiete.*

Kolmogorov, A. N. (1963) The theory of probability. *Transactions of mathematical monographs* **1**, AMS.

Kolmogorov, A. N. (1983) Combinatorial basis of information theory and probability theory. *Russ. Math. Surveys* **38** 29–40

Gács, P. (1993) Lecture notes on descriptional complexity and randomness, Boston University 1–67.

Galatolo, S. (1999) Pointwise information entropy for metric spaces. *Nonlinearity* **12** 1289–1298.

Galatolo, S. (2000) Orbit complexity by computable structures. *Nonlinearity* **13** 1531–1546.

Hertling, P. and Weihrauch, K. (1998) Randomness spaces. *Springer-Verlag Lecture Notes in Computer Science* **1443** 796–807.

van Lambalgen, M. (1987) *Random sequence*, Academish Proefschrift, Amsterdam.

Levin, L. A. and Zvonkin, A. K. (1970) The complexity of finite objects and the algorithmic foundations of the notions of information and randomness. *Russ. Math. Surveys* **25**.

Levin, L. A. (1984) Randomness conservation inequalities: information and independence in mathematical theories. *Inf. and Control* **61** 15–37.

Lévy, P. (1925) *Calcul des probabilités*, Gauthiers-Villars.

Li, M. and Vitanyi, P. (1997) *An introduction to Kolmogorov complexity and its applications*, Springer-Verlag.

Martin Löf, P. (1966) The definition of random sequence. *Inf. and Control* **9** 602–619.

Petersen, K. (1983) *Ergodic theory*, Cambridge University Press.

Rogers, H. (1967) *Theory of recursive functions and effective computability*, McGraw Hill.

V'Yugin, V. V. (1997) Effective convergence in probability and an ergodic theorem for individual random sequences. *Theory Probab. Appl.* **42** (1) 39–50.

https://doi.org/10.1017/S0960129508006646 Published online by Cambridge University Press