An Effective Ergodic Theorem and A Few Applications (Extended Abstract)

Satyadev Nandakumar
Department of Computer Science
Iowa State University
Ames, Iowa 50011
U.S.A.

April 20, 2007

Abstract

In the theory of algorithmic randomness, we try to formulate probabilistic laws, i.e., laws which hold with probability 1, to laws which hold in their pointwise constructive form - i.e., laws which hold for every individual constructively random point. Since the set of constructively random points form a measure 1 set, this implies the classical result when the probability distribution is computable. It was conjectured that Birkhoff's Ergodic Theorem resists effectivization in this form. V'yugin [13] however, in a significant result, proved a constructive version of the Ergodic Theorem which holds when the probability space, the transformation and the random variable are computable. By the definition of a computable function over the reals, all computable random variables are continuous [15]. However, in the metric theory of numbers, many of the random variables used correspond to discontinuous functions over the reals. Braverman in a recent work [4] defined a notion called "graph-computability" which handles discontinuous functions. We prove that the class of graph computable functions can be appropriately restricted so that the constructive ergodic theorem holds. We then use this to give effective ergodic proofs of the effective versions of Lévy-Kuzmin and Khinchin Theorems relating to continued fractions.

1 Introduction

In the context of Kolmogorov's program to base the theory of probability on the theory of computing, an early achievement was Martin-Löf's work establishing that there is a unique smallest constructive measure 1 set whose objects are individual random sequences [7]. In this program, we formulate probabilistic laws, i.e., laws of the form

$$\text{Probability}[\{\omega : A(\omega) \text{ holds }\}] = 1$$

for some property A, in their effective form,

$$\text{If } \omega \text{ is random, then } A(\omega) \text{ holds.}$$

It is not known whether all such laws can be converted into this form: early work by Vovk on the Law of Iterated Logarithm [11] and van Lambalgen on the Strong Law of Large Numbers [10] were successes; but, it was conjectured that not all laws can be converted into the effective form. In particular, it was conjectured that two laws, the Ergodic Theorem of Birkhoff [2] and the Shannon-McMillan-Breiman Theorem [1] resist effectivization. Nevertheless, V'yugin in [13] converted a proof of a constructive version of the Ergodic Theorem by Bishop [3] to prove an effective version of the Ergodic Theorem.

The ergodic property is a weak form of independence obeyed by stochastic processes. If P is a finite measure, f is an integrable function and T is a transformation preserving the measure P, then Birkhoff's ergodic theorem states that the limit

$$\lim_{n\to\infty} \frac{1}{n} \left(f(\omega) + f(T\omega) + \cdots + f(T^{n-1}\omega) \right) \tag{1}$$

exists for almost all ω in the sample space (for instance, see [1]). Moreover, if T is ergodic (definitions in section 2), then the above said average is the same constant almost everywhere. V'yugin's version establishes that if P is a computable measure, f is an integrable computable function and T is a computable measure-preserving transformation, the limit exists for all individual random ω. If T is ergodic, then the average (1) is the same constant for all individual random points. The convergence to the constant need not be effective - a computable function may not be able to predict the rate of convergence. [13]

We wish to explore applications of the effective version of the Ergodic Theorem in this paper. The ergodic theory of continued fractions (see for example, Kraaikamp and Dajani [5]) provides some examples of non-trivial applications of the ergodic theorem to the metric theory of numbers. The celebrated theorems of Lévy-Kuzmin, and Khinchin, are examples. We would like to form effective versions of these theorems. These are known to hold effectively, and the proofs employ transfer operators [8].

Classically, the proofs of these properties fall out of the ergodic theorem. The lack of an effective ergodic theorem has hindered the proofs in their classical form being used to establish the theorems in their effective form. We find that V'yugin's version cannot be used for this purpose because of a technical limitation - computable functions are continuous, while most of the proofs employ functions which are discontinuous.

A recent work by Braverman [4] suggests a way of handling computability of discontinuous functions using a notion termed "graph-computability". Graph-Computability cannot be adopted without modifications to prove the ergodic theorem. Indeed, we exhibit a graph-computable function for which the Constructive Ergodic Theorem fails. However, with suitable restrictions on the class of graph-computable functions, we can prove a constructive version of the theorem. Our aim in this paper is twofold - to prove a version of the effective ergodic theorem applicable for the metric theory of numbers, and to give new proofs of some classical results in continued fractions in their effective form, using the above.

2 Preliminaries

As usual, $\mathbb{R}$ denotes the set of real numbers, $\mathbb{N}$ denotes the set of natural numbers, and $\mathbb{Q}$ denotes the set of rationals. We denote the positive part of set Q by Q^+ and the negative part by Q^-. The notation $0^{\mathbb{N}}$ represents the unary notation of natural numbers. Now, to define the sample space, we consider an alphabet. Let Σ be the alphabet, which, in this paper, could be assumed to be either the binary alphabet $\{0,1\}$, or the natural numbers together with ∞, $\mathbb{N} \cup \{\infty\}$. We consider finite words over the alphabet, denoted by Σ^*, and infinite sequences over the alphabet, denoted by Σ^∞. For positive integer i, the i^{th} position of sequence or word ω is denoted as ω_i. The substring $\omega_i \dots \omega_{j-1}$ is denoted $\omega[i \dots j-1]$. For a word ω, the length of $\omega = \omega_0\omega_1\omega_2 \dots \omega_{n-1}$ is denoted as

$|\omega|$ with value n. If x is a string and w is a word or a sequence, the symbol $x \sqsubseteq \omega$ denotes that x is a prefix of ω.

Ergodic Theory

Let $(\Omega, \mathcal{F}, P)$ denote a probability space, where $\Omega = \Sigma^\infty$ is the sample space, $\mathcal{F}$ denotes the Borel $\sigma-$ field generated by the cylinders $C_x = \{\omega \mid \omega \in \Omega, x \sqsubseteq \omega)$, and $P : \mathcal{F} \to [0, 1]$ is the probability measure.

We introduce some basic concepts from Ergodic Theory.

Let $T : \Omega \to \Omega$ be a transformation, *i.e.*, a measurable function from Ω to itself. In particular we consider the case when the transformation T is a measure-preserving transformation with respect to the probability space $(\Omega, \mathcal{F}, P)$. That is, for every measurable set A, we have $P[T^{-1}A] = P[A]$. A set A such that $T^{-1}A = A$ is called an invariant set. Such a measure-preserving transformation T is called an ergodic transformation if every invariant set, and only an invariant set, has measure either 0 or 1. A function $f : \Omega \to \mathbb{R}$ is called invariant if $f(T\omega) = f(\omega)$ almost surely. Successive applications of T are denoted as follows: $T^0\omega = \omega$ and, for all n, $T^{n+1}\omega = T(T^n\omega)$. Ergodic systems are weakly independent systems. For further details, see, for instance, Walters [14].

A *dynamical system* is a system $(\Omega, \mathcal{F}, P, T)$ where $(\Omega, \mathcal{F}, P)$ is the probability space, and $T : \Omega \to \Omega$ is the measure-preserving transformation (which need not necessarily be ergodic). Two examples of dynamical systems are given below.

1. The probability space $(\Omega, \mathcal{F}, P)$ where Ω is the set of binary sequences, $\mathcal{F}$ is the Borel σ-algebra generated by the cylinders $C_x = \{\omega \in \Omega \mid x \sqsubseteq \omega\}$, and $P : \mathcal{F} \to [0, 1]$ is the uniform probability measure, Lebesgue measure. The transformation $T : \Omega \to \Omega$ is the left-shift transformation, $T[\omega_1\omega_2\omega_3\omega_4 \ldots] = \omega_2\omega_3\omega_4 \ldots$, identified with the numerical function $T\omega = \{2\omega\}$, where $\{x\}$ denotes the fractional part of x; T is seen to be measure-preserving and ergodic with respect to the probability space.

2. The probability space $(\Omega, \mathcal{F}, P)$ where Ω is the set of integer sequences identified with the continued fraction mapping (see section 6), $\mathcal{F}$ is the Borel σ-algebra generated by the cylinders $C_x = \{\omega \in \Omega \mid x \sqsubseteq \omega\}$, and $P : \mathcal{F} \to [0, 1]$ is the Gauss measure, for any measurable set A, the probability $P[A] = \int_A \frac{1}{1+x} dx$. The transformation T is the left-shift transformation, as above, now identified with the numerical function $T\omega = \{\frac{1}{\omega}\}$ where $\{x\}$ denotes the fractional part of the number x. The Gauss measure is important, since T is measure-preserving with respect to the Gauss measure, but not with respect to the uniform measure. T is also ergodic with respect to the probability space. For details, see Billingsley [1].

Algorithms, Graph-Computability

We also consider algorithms which map finite objects to finite objects - for instance, of the type $(\mathbb{N} \to \mathbb{N})$, $(\mathbb{Q} \to \mathbb{Q})$ and $(\Sigma^* \to \mathbb{N})$. An element of $\mathbb{N}$, Q or Σ^* is a finite object. Any finite object is computable. A real number r is computable if there exists an algorithm $f : 0^{\mathbb{N}} \to \mathbb{Q}$ such that for any integer n presented in unary, $f(0^n)$ is a rational q such that $|r - q| \leq 2^{-n}$. For convenience, we fix an encoding of a finite object as an element of Σ^*. We assume a given universal algorithm $U : \mathbb{N} \times \Sigma^* \to \Sigma^*$ such that, for every algorithm $m : \Sigma^* \to \Sigma^*$, there exists an ('index') integer i such that for all words x, $U(i, x) = m(x)$. Further, we define the following notions in computable analysis(Weihrauch [15]).

Throughout this section, the symbol f is a function the type $\Omega \to \mathbb{R}$ or of the type $\Sigma^* \to \mathbb{R}$. A function $f : \Omega \to [-\infty, \infty]$ is said to be lower semicomputable if the set $G_l = \{(w, q) \mid w \in \Omega, q \in$

$\mathbb{Q}, q < f(\omega)\}$ is the union of a computably enumerable sequence of cylinders in the natural topology on $\Omega \times \mathbb{Q}$ or $\Sigma^* \times \mathbb{Q}$. The natural topology on $\Sigma^* \times \mathbb{Q}$ is the discrete topology. The natural topology on $\Omega \times \mathbb{Q}$ is the topology generated by the cylinders of the form (x, q) where $x \in \Sigma^*$ and $q \in \mathbb{Q}$. Analogously, the function f is said to be upper semicomputable if $-f$ is lower semicomputable. A function f is said to be computable if it is both lower and upper semicomputable. Equivalently, we can show that f is computable if and only if there is an algorithm computing a function $f' : \mathbb{Q} \to \mathbb{Q}$ such that for every natural number n, every real r and every rational q, if $|r - q| < 2^{-n}$, then $|f'(q) - f(r)| < 2^{-n}$. It follows that every computable function is necessarily continuous.

This introduces a problem: Many of the functions which come up in Ergodic proofs are not computable becausethey are not continuous. Graph-computability, introduced in Braverman [4], gives a framework for discussing computability of not necessarily continuous functions. A graph of a function $f : \mathbb{R} \to \mathbb{R}$ is said to be computable if there exists a computable function $f' : \mathbb{Q}^2 \times \mathbb{N} \to \{0, 1\}$ such that

$$f'((q_1, q_2), 0^n) \quad = \begin{cases} 0 & \text{if } B_d((q_1, q_2), 2^{-n}) \cap \Gamma_f = \emptyset \\ 1 & \text{if } B_d((q_1, q_2), 2^{-n}) \text{ intersects } \Gamma_f \\ 0 \text{ or } 1 & \text{otherwise.} \end{cases}$$

The neighborhood $B_d((q_1, q_2), r)$ represents the ball centered around (q_1, q_2) with radius r. A function is said to be graph-computable if the graph of f is computable. It follows that every computable function is graph-computable. In addition to this, some step functions, which were not computable according to the definition above, are now shown to be graph-computable. For example, the unit step function $\sqsubset f(x) =$ if $x > 0$then1, else $0 \sqsupset$ is graph-computable.

Now, we discuss computable transformations.

Transformations mapping Ω to itself are viewed as operating on the sequences themselves; the left-shift transformation is a simple case in point. Informally, a computable transformation is one which can be computed by an algorithm bit-by-bit, *i.e.*, reads part of the input and extends the output with more input read. Formally, a computable transformation $T : \Omega \to \Omega$ is called computable if there is an algorithm which enumerates the set $\mathcal{S}_T = \{(x, y) | x, y \in \Sigma^*$, such that

1. $(x, \emptyset) \in \mathcal{S}_T$, where $\emptyset$ is the empty string.
2. $(x, y) \in \mathcal{S}_T \Rightarrow \forall x' \sqsubseteq x, y' \sqsubseteq y, (x', y') \in \mathcal{S}_T$.
3. $(x, y)$and$(x, y') \in \mathcal{S}_T$ implies $y \sqsubseteq y'$ or $y' \sqsubseteq y$.

The transformation T is defined as:

$$T\omega \quad = \quad \sup\{y \mid (x, y) \text{ such that } x \sqsubseteq \omega\}.$$

A computable probability measure $P : 2^{\Omega} \to [0, 1]$ is one for which $P(x) = P(C_x)$ is computable.

3 Constructive Randomness

In this section, we the definition of a constructively random sequence, and introduce the notion of a measure of impossibility, which we use to prove that a computable measure-preserving transformation conserves the randomness of a sequence.

Let P be a computable probability measure defined on the Cantor Space(defined in section 3). For finite strings x, we consider cylinders C_x, the set of all infinite sequences with x as a prefix. A set S of sequences from the sample space of all sequences has P-measure zero if, for each $\varepsilon > 0$, there is a sequence of cylinders $C_{x_0}, C_{x_1}, \ldots, C_{x_i}, \ldots$ of cylinder sets such that

$$S \subseteq \cup_i C_{x_i} \text{ and } P(\cup_i C_{x_i}) < \varepsilon.$$

A set of sequences S has effective P-measure zero if there is a computable function $h(i, \varepsilon)$ such that $h(i, \varepsilon) = C_{x_i}$ for each i.

A probability measure $P : 2^{\Omega} \to [0, 1]$ is called computable if the probability measure $P : \Sigma^* \to [0, 1]$ is a computable function. The notatation $P(w)$ for a string $w \in \Sigma^*$ stands for $P(C_w)$, the probability of the cylinder C_w.

An equivalent tool to study randomness is the concept of a measure of impossibility [12].

Gacs in [6] extends the notion of Martin-Löf randomness to some non-compact spaces, one which he characterizes as spaces having recognizable boolean inclusions. We take the characterization and note that it works for Cantor Space, the space of infinite binary sequences, and Baire space, the space of infinite sequences of natural numbers.

Definition 1. *A function $p : \Omega \to \mathbb{R}^+ \cup \{\infty\}$ is called a* measure of impossibility *with respect to the probability space $(\Omega, \mathcal{F}, P)$ if the following hold:*

P1. p is lower semicomputable.

P2. $E_P\, p \leq 1$, where $E_P\, f$ is the expectation of the function f with respect to probability measure P.

A *measure of impossibility* p of ω with respect to the computable probability distribution P denotes whether ω is random with respect to the given probability distribution or not. In particular, we can see that $p(\omega) < \infty$ if and only if ω is random with respect to the probability distribution P [12], [6].

We now use this tool to give a proof of the fact that a computable, measure-preserving transformation conserves randomness. This extends, and gives a new proof of, Shen's [9] result on Cantor Space.

Lemma 2. *Let ω be a Martin-Löf random real in Baire space or Cantor Space. Then for any computable measure-preserving transformation T, $T\omega$ is also Martin-Löf random.*

Proof Sketch Let $T\omega$ be non-random. By assumption, there is a measure of impossibility p such that $p(T\omega) = \infty$. We define a new function $p' : \Omega \to \mathbb{R}^+ \cup \{\infty\}$ by $p'(\chi) = p(T\chi)$. p' is lower semicomputable by the lower semicomputability of p and the computability of T. Also, $\int p' dP = \int p dP$ by the measure conservation property of T. Thus p' is a measure of impossibility such that $p'(\omega) = \infty$. □

This lemma will be used in section 4.

4 Main Result

We would like to prove the following:

Ideal Theorem If $(\Omega, \mathcal{F}, P)$ is a probability space where $\Omega = \Sigma^{\infty}$, with Borel σ-field generated by C_x, $x \in \Sigma^*$ and $P : 2^{\Omega} \to [0, 1]$ is a computable probability measure, then for any function $f : \Omega \to \mathbb{R}$ which is graph-computable, $f \in L^1 P$, the ergodic average for any computable transformation $T : \Omega \to \Omega$ which is measure-preserving wrt the probability space converges.

However, there are graph-computable functions for which the ergodic average does not converge to the mean of the random function.

Example 1. We construct a function $f : \Omega \to [0, 1]$ which is graph-computable, but is such that the effective ergodic theorem fails to hold.

Consider f defined as follows.

Let ω be an arbitrary Martin-Löf random real, e.g., the halting probability in binary notation, and $Tx = \{2x\}$. Then ω is normal: For all $n \in \mathbb{N}$ and $x \in \{0, 1\}^n$,

$$\lim_{i \to \infty} \frac{|\{m : 0 < m + n < i \text{ and } \omega[m \dots m + n - 1] = x[0 \dots n - 1]\}|}{i} = 2^{-n}.$$

In particular, the set $\{T^j \omega : j \in \mathbb{N}\}$ is dense in the unit interval. Define, for all $j \in \mathbb{N}$, $f(T^j \omega) = 1$, and $f(x) = 0$ for all other x.

This function is graph computable because both the sets

$$\{x : f(x) = 0\}$$

and

$$\{x : f(x) = 1\}$$

are dense in $[0, 1]$. The function is graph computable with a witness $B((q_1, q_2), 2^{-n}) = 1$ if and only if $|q_2| < 2^{-n-1}$ or $|1 - q_2| < 2^{-n-1}$.

We notice

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

However, $\int f(x)dx = 0$, since $\{x : f(x) = 0\}$ is a measure 1 set, so the effective ergodic theorem fails to hold for ω. □

This example serves to prove that graph-computability needs restriction in order to serve our purpose. One of the problems of the above example is the presence of a dense set of discontinuities. We posit the following class of graph-computable functions.

Definition 3. *Let $\mathcal{G}$ be the class of graph-computable functions f with the property that f has atmost countable, nowhere dense, P-measure zero set of discontinuities, all on computable points.*

$\mathcal{G}$ is a superset of the class of computable functions. It is also large enough to subsume useful discontinuous functions used in some proofs of the metric theory of numbers.

We have the following.

Theorem 4. *Let $(\Omega, \mathcal{F}, P)$ be a probability space where $\Omega = \Sigma^\infty$, $\mathcal{F}$ is the Borel σ-field generated by the cylinders C_x, and P is a computable probability measure. Then, for any measure-preserving transformation $T : \Omega \to \Omega$ which is pointwise effectively invertible, and any graph-computable function $f : \Omega \to \Re$ in $\mathcal{G}$ with discontinuities on a nowhere dense measure 0 set of computable points, then for every individual random ω,*

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

Moreover, if T is ergodic, the abovementioned limit is a constant, for all individual random ω, $\tilde{f}(\omega) = \int f \, dP$.

The idea of the proof is as follows. For a graph-computable function f which obeys the restriction mentioned above, we prove that it is easy to obtain a function which is essentially an upper semicomputation except at points of discontinuity; moreover, by the same argument, there is also a lower semicomputation of the function f. Using these upper and lower semicomputations, an upcrossing function is defined, which behaves reasonably well at points of continuity - namely, it converges if and only if the ergodic sum at the point converges. This function is shown to be semicomputable from below. We bound the integral of the upcrossing function over the whole space (this is essentially due to V'yugin [2]), thus we prove that the upcrossing function we defined measure of impossibility. At every point of continuity, the measure of impossibility diverges if and only if the ergodic sum at that given point converges. Thus we have an upcrossing function which is integrable, lower semicomputable and which diverges only if the ergodic average diverges at the given point. This would imply that the ergodic sum converges at every individual random point. Details follow in subsequent sections.

5 Proof of Theorem 3

We first prove that for every function in $\mathcal{G}$, there exists an essential upper semicomputation and an essential lower semicomputation - i.e., functions which are semicomputations except at the points of discontinuity.

Lemma 5. *Let $f : \mathbb{R} \to \mathbb{R}$ be a graph-computable function in $\mathcal{G}$. Then there is a computable function $f^u : \mathbb{Q} \times \mathbb{N} \to \mathbb{Q}$ such that for every point of continuity $r \in \mathbb{R}$, every approximation witness $\hat{r}$ of r the following hold:*

1. *For all natural numbers n, $f^u(\hat{r}, 0^n) > f(r)$.*
2. $\lim_{n\to\infty} f^u(\hat{r}, 0^n) = f(r)$.

Proof Sketch. A function f in $\mathcal{G}$ can be seen as a function which is piecewise computable. We can use a standard "graph lookup" algorithm to come up with approximations to the value of $f(x)$. Using the standard proof that every computable function is upper semicomputable, we can construct a function f^u. □

Similarly, an analogous proof establishes that there is a lower pseudosemicomputation f^l of f, which at points of continuity, approximates f in the limit.

We make the following observation. V'yugin proves that for a computable function f, at every random point ω, f has to have a finite value at ω. This need be no longer true in the case of graph-computable functions. For example, let ε be a fixed algorithmically random point. It can be verified that f which is identically 0 everywhere except at ε, the value $f(\varepsilon)$ being ∞, is graph-computable according to Braverman's definition. We consider only bounded functions.

Now, we proceed to the proof of the main result.

5.1 Proof of Theorem 3

Let f be the graph computable function, as given, with upper pseudo-semicomputation f^u and lower pseudo-semicomputation f^l. Let ω be a random point. Define the following:

$$a(\omega, f, n) = \sum_{i=0}^{n-1} \left[f(T^i\omega) - \alpha\right] \tag{4}$$

$$b(\omega, f, n) = \sum_{i=0}^{n-1} \left[f(T^i\omega) - \beta\right] \tag{5}$$

for any rational numbers α and β, natural numbers n and f given above. The set of functions can be undefined if for some f, the value at a point is $\pm\infty$. We assume that the functions are bounded, so this is impossible. We define the sets A as follows:

$$\omega \in A(u, v) \equiv a(\omega, f, u) < b(\omega, v, f) \tag{6}$$

Let n be a non-negative integer. A sequence of integers $s = \{u_1, v_1, \ldots, u_n, v_n\}$ is called an n-admissible sequence if $-1 \leq u_1 < v_1 \leq u_2 < v_2 \leq \cdots \leq u_N < v_N \leq n$. We use $m_s = N$. We define $a(\omega, f, -1) = 0$. We define Bishop's upcrossing function.

$$\begin{aligned}\sigma'_n(\omega, \alpha, \beta) = \max\ (\{N | \omega \in \bigcap_{j=1}^{N} A(u_j, v_j) \cap \bigcap_{j=1}^{N-1} A(u_{j+1}, v_j) \\ \text{for some n-admissible } s = \{u_1, v_1, \ldots, u_N, v_N\}\} \cup \{0\}),\end{aligned} \tag{7}$$

We then define the function in its modified form:

$$\sigma_n(\omega, \alpha, \beta) \quad = \begin{cases} \sigma'_n(\omega, \alpha, \beta) & \text{if } \omega \in C_f \\ \infty & \text{otherwise,} \end{cases} \tag{8}$$

where C_f is the set of continuity points of f.

The upcrossing function σ_n is lower semicomputable, since we can use the upper pseudosemicomputation f^u and the lower pseudosemicomputation f^l to compute σ_n, and at points of continuity ω, $\lim_{n\to\infty} f^u(\omega, 0^n) = f(\omega) = \lim_{n\to\infty} f^l(\omega, 0^n)$. The points of discontinuity are all non-random points, and therefore there is a measure of impossibility p that attains ∞ on all of them. Define $\sigma' = \sup_n \sigma'_n$ and $\sigma = \sup_n \sigma$. Then σ is a lower semicomputable function.

We notice that $\sigma(\omega, \alpha, \beta)$ and $\sigma'(\omega, \alpha, \beta)$ differ only on a P-measure zero set. In particular, this means that

$$\int \sigma'(\omega, \alpha, \beta) dP = \int \sigma(\omega, \alpha, \beta) dP. \tag{9}$$

V'yugin proves that σ'_n is an integrable function. His proof assumes that f is computable and T is measure-preserving. We note that the proof that σ_n is integrable for graph-computable functions is identical to V'yugin's; it is reproduced here only to reflect the invariance in technical detail.

Since f is integrable, we have some constant M for which

$$\int (f(\omega) - \alpha)^+ dP \leq M + |\alpha|$$

holds.

Lemma 6. [*V'yugin [13]*] *Let T be a measure-preserving transformation and $f : \Omega \to \mathbb{R}$ be an integrable, computable function. Then*

$$\int (M + |\alpha|)^{-1}(\beta - \alpha)\sigma_n(\omega, \alpha, \beta) \leq 1. \tag{10}$$

We have thus bounded the upcrossing function.

Therefore we have defined a measure of impossibility. It follows that for all individual random ω, the ergodic average converges.

To see that for every individual random ω, the ergodic average converges to the same constant, we observe that every random ω is a point of continuity. Hence V'yugin's null cover can be used to prove that for every ω, if ω is random, then the ergodic average is the same constant.

6 Continued Fractions

We introduce some basic notions on the theory of continued fractions.

A real number $r \in (0, 1)$ is said to have continued fraction expansion $[a_1, a_2, \dots, a_n, \dots]$ if r can be expressed as

$$r = \cfrac{1}{a_1 + \cfrac{1}{a_2 + \ddots + \cfrac{1}{a_n + \ddots}}} \tag{11}$$

where $a_i \in \mathbb{N} \cup \infty$. If any of the a_i is ∞, the representation is finite, and r is a rational number. The set of finite representations constitute exactly the set of rational numbers. This is the basic difference of the continued fraction representation over a b-adic representation of the reals. We introduce some basic notations and inequalities.

Let $\frac{p_n}{q_n}$ be the representation of the rational $[a_1, \dots, a_n]$ obtained by truncating the representation of r to n places, written in lowest form (i.e., $\gcd(p_n,\, q_n) = 1$.). The fraction $\frac{p_n}{q_n}$ is called the n^{th} convergent of the real number r, the number p_n being the n^{th} partial quotient and q_n being the n^{th} partial denominator of the continued fraction. The following inequality is well-known (see, for example, Dajani and Kraikaamp [5].).

$$\forall n \; p_n q_{n+1} = -p_{n+1} q_n.$$

We prove the effective versions of two famous theorems in the ergodic theory of continued fractions, *viz.*, the Lévy-Kuzmin Thorem and Khinchin's Theorem.

7 Effective Notions in Ergodic Theory of Continued Fractions

We use Theorem One to prove the effective version of the Lévy-Kuzmin Theorem. The proof is an effective version of the classical theorem. We have to prove that the functions used in the proof conform to the conditions imposed in theorem 1. The classical proofs are adapted from the exposition in [5]. We just note that the appropriate functions are members of $\mathcal{G}$.

Theorem 7. *Let r be a real number in $(0, 1)$. If r is constructively random, the following hold:*

$$\lim_{n\to\infty} \frac{\log\, q_n(r)}{n} = \qquad \frac{\pi^2}{12\ \log\ 2} \lim_{n\to\infty}$$

Proof. We show that standard proofs using ergodic theory directly translate to the effective version. See [5] for the classical proof. It can be shown that

$$\frac{\log\, q_n(x)}{n} = \sum_{m=0}^{n-1} f(T^m(x)) + R(n,x)$$

for all x, where the absolute value of the error $|R(n,x)|$ is bounded. The function $\log(x)$ is computable, therefore, by V'yugin's ergodic theorem, for all individual random x,

$$\lim_{n\to\infty} \frac{\sum_{m=0}^{n-1} f(T^m(x))}{n} = \int \frac{\log x}{1+x} dx,$$

which proves the result. □

Now, we prove Khinchin's theorem [5] where the additional result over graph computability is useful.

Theorem 8. *For every random $\omega \in [0,1)$ with the standard continued fraction expansion,*

$$\lim_{n\to\infty} (a_1 a_2 \dots a_n)^{1/n} = \prod_{n=1}^{\infty} (1 + \frac{1}{k(k+1)})^{\frac{\log k}{\log 2}} = \mathcal{K} = 2.6854\dots.$$

Proof. It suffices to show that for every random ω, the ergodic average

$$\lim_{n\to\infty} \frac{\log a_1(\omega) + \dots + \log a_n(\omega)}{n} = \log \mathcal{K}.$$

The function $a_1(\omega) = \lfloor \frac{1}{\omega} \rfloor$ can be shown to be a function in $\mathcal{G}$, even though it is not computable in the usual sense. Since log is a computable function, it follows that $f(x) = \log(a_1(x)) \in \mathcal{G}$. Hence by the effective ergodic theorem for random variables in $\mathcal{G}$, we have

$$\begin{aligned} \lim_{n\to\infty} \frac{\log a_1(\omega) + \dots + \log a_n(\omega)}{n} &= \int \frac{\log a_1(\omega)}{(1+\omega)} d\omega \\ &= \sum_{k=1}^{\infty} \int_{\frac{1}{k+1}}^{\frac{1}{k}} \frac{\log a_1(\omega)}{1+\omega} d\omega \\ &= \sum_{k=1}^{\infty} \frac{\log k}{k(k+2)}, \end{aligned} \tag{12}$$

which is a convergent series with limit $\log \mathcal{K}$. □

8 Acknowledgments

The author wishes to thank his advisor Jack Lutz for helpful advice and discussions.

References

[1] P. Billingsley. *Ergodic theory and information.* John Wiley and Sons, New York, 1965.

[2] G. D. Birkhoff. Proof of the ergodic theorem. *Proceedings of the National Academy of Sciences, USA*, 17:656–660, 1931.

[3] E. Bishop. *Foundations of constructive analysis.* McGraw Hill, New York, 1967.

[4] M. Braverman. On the complexity of real functions. In *Proceedings of the forty sixth IEEE Annual Symposium on the Foundations of Computer Science*, pages 155–164, 2005.

[5] K. Dajani and C. Kraaikamp. *Ergodic Theory of Numbers.* Mathematical Association of America, 2004.

[6] P. Gács. Unifrom test of algorithmic randomness over a general space. *Theoretical Computer Science*, 341:91–137, 2005.

[7] P. Martin-Löf. The definition of random sequences. *Information and Control*, 9:602–619, 1966.

[8] J. Reimann. Hausdorff dimension, randomness and entropy. *Unpublished Lecture, Colloquium Logicum Heidelberg*, 2004.

[9] A. Shen. The frequency approach to finding the definition of a random sequence. *Manuscript.*

[10] M. van Lambalgen. *Random Sequences.* PhD thesis, Department of Mathematics, University of Amsterdam, 1987.

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

[12] V. G. Vovk and V. V. V'yugin. On the empirical validity of the bayesian method. *Journal of Royal Statistical Society B*, 55:253–266, 1993.

[13] V. V. V'yugin. Ergodic theorems for individual random sequences. *Theoretical Computer Science*, 207:343–361, 1998.

[14] P. Walters. *An introduction to ergodic theory.* Springer Verlag, New York, 1982.

[15] K. Weihrauch. *Computable Analysis.* Springer Verlag, Berlin, 2000.