An Effective Ergodic Theorem and Some Applications

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

ABSTRACT

This work is a synthesis of recent advances in computable analysis with the theory of algorithmic randomness. In this theory, we try to strengthen probabilistic laws, i.e., laws which hold with probability 1, to laws which hold in their pointwise effective form - i.e., laws which hold for every individual constructively random point. In a tour-de-force, V'yugin [13] proved an effective version of the Ergodic Theorem which holds when the probability space, the transformation and the random variable are computable. However, V'yugin's Theorem cannot be directly applied to many examples, because all computable functions are continuous, and many applications use discontinuous functions.

We prove a stronger effective ergodic theorem to include a restriction of Braverman's "graph-computable functions". We then use this to give effective ergodic proofs of the effective versions of Lévy–Kuzmin and Khinchin Theorems relating to continued fractions.

Categories and Subject Descriptors

F.1.1 [**Theory of Computation**]: Models of Computation; G.3 [**Mathematics of Computation**]: Probability and Statistics

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 "Probability$[\{\omega : A(\omega) \text{ holds } \}] = 1$" for some property A, in their effective form, "If ω is random, then $A(\omega)$ 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, $\int f dP$, 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. Classically, the Strong Law of Large Numbers can be proved to be a special case of the Ergodic Theorem. Moreover, 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 no-

*Research supported in part by National Science Foundation grants 0344187, 0652569, and 0728806.

tion 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 threefold - to prove a version of the effective ergodic theorem which handles discontinuous functions, to prove the effective strong law of large numbers as a consequence of the effective ergodic theorem, 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 binary alphabet $\{0,1\}$. 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.

Basic Concepts in Ergodic Theory

Let $(\Omega, \mathcal{F}, P)$ denote a probability space, where $\Omega = \Sigma^\infty$ is the sample space, $\mathcal{F}$ denotes the Borel $\sigma-$ algebra 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 $P[T^{-1}A] = P[A]$. If, moreover, $T^{-1}A = A$ for only mesaure 0 and measure 1 sets, then T is called *ergodic*.

Ergodic systems are weakly independent systems. For further details, see, for instance, Walters [14].

Successive applications of T are denoted as follows: $T^0\omega = \omega$ and, for all n, $T^{n+1}\omega = T(T^n\omega)$. Customarily, the set $\{T^n\omega : n \in \mathbb{N}\}$ is called the *orbit* of ω under T.

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 be ergodic). Two examples of dynamical systems are given below.

1. The probability space $(\Omega, \mathcal{F}, \mu)$ 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 μ is the uniform probability measure, Lebesgue measure. The transformation $T : \Omega \to \Omega$ is the left-shift transformation,
$$T[\omega_1\omega_2\omega_3\omega_4 \dots] = \omega_2\omega_3\omega_4 \dots,$$
identified with the numerical function $T\omega = 2\omega \mod 1$; T is seen to be measure-preserving and ergodic with respect to the probability space.
2. The probability space $(\mathbb{N}^{+^\infty}, \mathcal{F}', \gamma)$ where $\mathbb{N}^{+^\infty}$ is the set of positive integer sequences identified with the continued fraction expansion of reals in $(0,1)$. (see section 6), $\mathcal{F}$ is the Borel σ-algebra generated by the cylinders $C_x = \{\omega \in \mathbb{N}^{+^\infty} \mid x \sqsubseteq \omega\}$, and γ is the Gauss measure, *i.e.* 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 $T\omega = \frac{1}{\omega} \mod 1$. 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 $\mathbb{N}$, $\mathbb{Q}$ 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 $|r - q| \le 2^{-n}$. The computable function f is called an *computability witness* for r. For convenience, we fix an encoding of a finite object as an element of Σ^*. Further, we define the following notions in computable analysis(Weihrauch [15]).

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 a real-valued f is computable if and only if there is a Turing machine M such that for every real r, if $\hat{r} : 0^N \to Q$ is a valid computability witness for r, then we have for all n, $|M^{\hat{r}}(0^n) - f(r)| < 2^{-n}$, where $M^{\hat{r}}$ is the machine M with oracle access to $\hat{r}$. It follows that every real-valued computable function on a bounded domain 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.

Definition 1 ([4]). *We say that a bounded subset S of $\mathbb{R}^n$ is bit-computable if there exists a computable function $f' : D^n \times \mathbb{N} \to \{0,1\}$ such that*
$$f'(d, 0^n) = \begin{cases} 0 & \text{if } B(d, 2.2^{-n}) \cap S = \emptyset \\ 1 & \text{if } B(d, 2^{-n}) \text{ intersects } S \\ 0 \text{ or } 1 & \text{otherwise,} \end{cases}$$
where the neighborhood $B(d,r)$ represents the ball centered around d with radius r. A bounded real function on a bounded domain $D \subseteq \mathbb{R}$ is said to be graph-computable *if the graph of $f = \{(x, f(x)) : x \in D\}$ is bit-computable as a set.*

Every computable function over a bounded domain 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 ($f(x) =$ if $x > 0$ then 1, else 0) is graph-computable.

Now, we discuss computable transformations. We follow the definition in [13].

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. Formally, a *computable transformation* $T : \Omega \to \Omega$ is a transformation for which there is an algorithm which enumerates the set $\mathcal{S}_T = \{(x, y) \mid x, y \in \Sigma^*\}$, such that

1. $(x, \lambda) \in \mathcal{S}_T$, where λ 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. $(x, y)\text{and}(x, y') \in \mathcal{S}_T$ implies $y \sqsubseteq y'$ or $y' \sqsubseteq y$.

The transformation T is defined as

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

For example, the mappings $T_1, T_2 : \Omega \to \Omega$ defined as follows. The transformation $T_1(\omega) = \lambda$ is one that maps every sequence to the empty string. The transformation $T_2(\omega_1\omega_2 \dots) = (\omega_2 \dots)$ is the left-shift transformation. Both are computable transformations, though only T_2 is measure-preserving with respect to the Lebesgue measure on $[0, 1)$.

A *computable probability measure* P is one for which for every string $x \in \Sigma^*$, $P(x) = P(C_x)$ is computable.

3. CONSTRUCTIVE RANDOMNESS

In this section, we define the notion 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 $\{0, 1\}^\infty$. 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 $C_{x_0}, C_{x_1}, \dots, C_{x_i}, \dots$ 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. Martin-Löf proved a universality property - that for every computable probability measure$P$, there is a unique largest effective P-measure zero set. The complement of this set is called the set of constructive random sequences wrt P.

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

Gács 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 2. *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 p is lower semicomputable and $\int pdP \leq 1$.*

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 ω is random with respect to the computable 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 that a measure-preserving transformation conserves individual randomness.

Lemma 3. *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. 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, $\int p'dP = \int pdP \leq 1$ by the measure conservation property of T. Thus p' is a measure of impossibility such that $p'(\omega) = \infty$. □

This lemma implies that no point in the orbit of a sequence random wrt a computable measure, will be computable. This 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-algebra generated by C_x, $x \in \Sigma^*$ and P is a computable probability measure, then for any function $f : \Omega \to \mathbb{R}$ which is graph-computable, $f \in L^1 P$, for any computable transformation T, and for any random ω wrt P, the ergodic average converges to $\int fdP$.*

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

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

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

In particular, the orbit of ω 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 $\lim_{n\to\infty} \frac{\sum_{m=0}^{n-1} f(T^m\omega)}{n} = 1$.

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. (End Example)

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 4. *Let $\mathcal{G}_P$ be the class of graph-computable functions f continuous almost everywhere wrt P, with the property that f has only simple discontinuities which form a nowhere dense (one-dimensional) bit-computable set.*

Note that a nowhere dense bit-computable set can contain only computable points, and hence there are at most countably many discontinuities. Hence by Lemma 3, for any constructively random ω, no point on its orbit $\{T^m\omega : m \in \mathbb{N}\}$ is a point of discontinuity. $\mathcal{G}_P$ 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.

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

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

We sketch the proof of this lemma in Section 5. We call any function which satisfies the above property as *essentially upper semicomputable*. A function f on a bounded domain is called *essentially lower semicomputable* if $-f$ is essentially upper semicomputable. A function f which has both properties is called *essentially computable*.

MAIN THEOREM 1. *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. If $T : \Omega \to \Omega$ is a computable measure-preserving transformation, then for every essentially computable $f : \Omega \to \mathbb{R}$, $f \in L^1[P] \cap \mathcal{G}_P$, there is an integrable function $\tilde{f}$ such that*

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

for every ω random wrt P, with $\tilde{f}(T\omega) = \tilde{f}(\omega)$ and $\int f dP = \int \tilde{f} dP$. 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 an essential upper semicomputation except at points of discontinuity; moreover, by the same argument, there is also an essential lower semicomputation of the function f. Using these essential 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 [13]), thus we prove that the upcrossing function we defined measure of impossibility.

The upcrossing function attains ∞ 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 the next sections.

5. PROOF OF MAIN THEOREM 1

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

Proof Sketch for Lemma 5. Let g be the computability witness for the set of discontinuities. If x is a continuity point and the discontinuities are nowhere dense, then there is a least n_1 such that $q + 2.2^{-n_1} < x - 2.2^{-n_1} < x + 2.2^{-n_1} < r - 2.2^{-n_1}$ for the discontinuities q, r closest to x, $r > q$. Thus, x can be detected to be a continuity point at precision $n_1 + 1$, using g as a witness. We define the upper semicomputation as follows.

For all $n > n_1$, if $B((d_1, d_2), 2^{-n})$ intersects the graph of the function and $|d_1 - x| < 2^{-n}$, then $d_2 + 2.2^{-n}$ is an upper approximation of $f(x)$. So, given 0^k, we see whether we can determine that x is a continuity point at precision k. If no, we output ∞. If yes, we enumerate balls of radius 2^{-k} until we find k distinct upper approximants to $f(x)$, and output their minimum.

It is routine to verify that this process is an essential upper semicomputation of f. □

Similarly, we establish that there is an essential lower semicomputation $\underline{f}$ of f, which at points of continuity, approximates f from below, and approaches f in the limit.

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

Proof of Main Theorem 1

Let f be the graph computable function, as given, with essential upper semicomputation $\overline{f}$ and essential lower semicomputation $\underline{f}$. Let ω be a random point. Define the following:

$$a(\omega, f, n) = \sum_{i=0}^{n-1} \left[\lim_{n\to\infty} \overline{f}(T^i\omega, 0^n) - \alpha \right] \tag{3}$$

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

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 and A' as follows:

$$\omega \in A'(u, v) \equiv \sum_{i=0}^{u-1} \left[f(T^i\omega) - \alpha \right] < \sum_{i=0}^{v-1} \left[f(T^i\omega) - \beta \right] \tag{4}$$

$$\omega \in A(u, v) \equiv a(\omega, f, n) < b(\omega, f, n). \tag{5}$$

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 $-1 \le u_1 < v_1 \le u_2 < v_2 \le \cdots \le u_N < v_N \le n$. We use $m_s = N$. We define $a(\omega, f, -1) = 0$. We define Bishop's upcrossing function.

$$\sigma'_n(\omega, \alpha, \beta) =$$

$$\max(\{N \mid \omega \in \bigcap_{j=1}^{N} A'(u_j, v_j) \cap \bigcap_{j=1}^{N-1} A'(u_{j+1}, v_j) \cap \mathcal{C}_f$$

$$\text{for some n-admissible } s = \{u_1, v_1, \ldots, u_N, v_N\}\} \cup \{0\}), \tag{6}$$

We then define the function in its modified form: $\sigma_n(\omega, \alpha, \beta)$ is defined analogous to σ'_n, with A replacing A' in the definition. It is easy to see that σ_n and σ'_n coincide on all points of continuity of f.

The function σ_n is lower semicomputable, since we can use the essential upper semicomputation $\overline{f}$ and the essential lower semicomputation $\underline{f}$ to compute σ_n, and at points of continuity ω, $\lim_{n\to\infty} \overline{f}(\omega, 0^n) = f(\omega) = \lim_{n\to\infty} \underline{f}(\omega, 0^n)$. The points of discontinuity are all non-random points. 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

$$\int \sigma'(\omega,\alpha,\beta)dP = \int \sigma(\omega,\alpha,\beta)dP. \tag{7}$$

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

$$\int (M+|\alpha|)^{-1}(\beta-\alpha)\sigma(\omega,\alpha,\beta) \le 1. \tag{8}$$

To define the measure of impossibility, we need to verify define an upcrossing function for every pair of rationals (q, r) where $q < r$. But we can enumerate all pairs of such rationals with a program, and let $\alpha(i)$ and $\beta(i)$ be the ith pair in this enumeration. The measure of impossibility is

$$p(\omega) = \frac{1}{2}\sum_{i=1}^{\infty} i^{-2}(M+|\alpha(i)|)^{-1}(\beta(i)-(\alpha(i))\sigma(\omega,\alpha(i),\beta(i)),$$

as in V'yugin [13]. It is easy to see that p is lower semicomputable and integrable.

This function attains ∞ only if the ergodic average diverges. It follows that for all individual random ω, the ergodic average converges.

Besides, $\tilde{f}(\omega) = \tilde{f}(T\omega)$ for any ω random wrt P, so $\tilde{f}$ is bounded almost everywhere P, and is therefore, integrable. Integrating (2) on both sides, by the measure preservation of T and the Dominated Convergence Theorem, we get $\int f dP = \int \tilde{f} dP$.

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 - if the ergodic average at a random $\omega = d \ne \int f dP = c$, and r_1 and r_2 are rationals that separate c and d, then the sets $\{x \in \Omega : r_1 < \frac{1}{n}\sum_{k=0}^{n-1} f(T^k x) < r_2\}$ can be used to define an effective null cover to contain ω [13]. □

The class of essentially computable functions include simple step functions like the indicator random variables of the cylinders C_w. These functions, being discontinuous are not computable. These are the functions that we use to observe that the Strong Law of Large Numbers in probability theory is a special case of the Ergodic Theorem. We now are able have a simple proof that the Effective Strong Law of Large Numbers, originally due to van Lambalgen [10] is a special case of the modified Effective Ergodic Theorem.

Corollary 7 (Effective SLLN [10]). *Let $(\Omega, \mathcal{F}, P)$ be such that P is generated by a computable coin-toss probability measure. Then $\lim_{n\to\infty} \sum_{i=0}^{n-1} \frac{\omega_i}{n} = P(1)$.*

Proof. The system $(\Omega, \mathcal{F}, P, T)$ is a computable dynamical system. The indicator function $f = I_1$ is an effectively computable, integrable function. Substituting $f = I_1$ in (2), we get that $\lim_{n\to\infty} \sum_{i=0}^{n-1} \frac{\omega_i}{n} = P(1)$. □

6. EFFECTIVE NOTIONS IN ERGODIC THEORY OF CONTINUED FRACTIONS

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

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

$$r = \cfrac{1}{a_1 + \ddots + \cfrac{1}{a_n + \ddots}} \tag{9}$$

where $a_i \in \mathbb{N}$. The set of finite representations constitute exactly the set of rational numbers. We introduce some basic notations and inequalities.

Let $\frac{p_n}{q_n}$ be the representation of the rational $[a_1, \ldots, a_n]$ 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 recurrence equality is well-known (see, for example, Dajani and Kraikaamp [5].), $\forall n p_n q_{n-1} - p_{n-1} q_n = (-1)^{n+1}$.

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.

We first note that the set of reals random wrt γ and μ are exactly the same, since the Radon-Nikodym derivative $\frac{d\gamma}{d\mu} = 1/x$ is a Lipshitz function, making every effective null-cover for μ, an effective null-cover for γ and conversely. We use Theorem 1 to prove the effective version of the Lévy-Kuzmin Theorem. The proof is an effective version of the classical proofs from the exposition in [5]. We only have to prove that the functions used in the proof are essentially computable.

Theorem 8. *Let r be a real number in $(0,1)$. If r is constructively random, then $\lim_{n\to\infty} \frac{\log q_n(r)}{n} = \frac{\pi^2}{12 \log 2}$.*

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 $\log q_n(x) = \sum_{m=0}^{n-1} \log(T^m(x)) + R(n, x)$ for all x, where the absolute value of the error $|R(n, x)|$ is bounded. The function $\log(x)$ is computable and monotone. The continued fraction map $g(x) = 1/x \mod 1$ on $(0,1)$ is a member of $\mathcal{G}_\mu$ and is a computable transformation.

Therefore, by the Effective Ergodic Theorem (Theorem 1), for all individual random x, we have

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

which proves the result. □

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

Theorem 9. *For every random $\omega \in [0,1)$ with the standard continued fraction expansion,*

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

Proof. It suffices to show that for every random ω, the ergodic average $\lim_{n\to\infty} \frac{\log a_1(\omega)+\cdots+\log a_n(\omega)}{n} = \log \mathcal{K}$. The function $a_1(\omega) = \lfloor \frac{1}{\omega} \rfloor$ can be shown to be essentially upper (lower) semicomputable, even though it is not computable in $\mathcal{G}_\mu$, by approximating it from above (below) by a computable sequence of computable functions. Since log is a computable function, it follows that $f(x) = \log(a_1(x))$ is essentially computable. Hence by the effective ergodic theorem for random variables in $\mathcal{G}$, we have

$$\lim_{n\to\infty} \frac{\log a_1(\omega) + \cdots + \log a_n(\omega)}{n} = \int \frac{\log a_1(\omega)}{(1+\omega)} d\omega \quad (10)$$

$$= \sum_{k=1}^{\infty} \int_{\frac{1}{k+1}}^{\frac{1}{k}} \frac{\log a_1(\omega)}{1+\omega} d\omega = \sum_{k=1}^{\infty} \frac{\log k}{k(k+2)},$$

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

7. ACKNOWLEDGMENTS

The author wishes to thank his advisor Jack Lutz for proposing the problem, for his patience with the ideas in gestation, and advice and discussions. We would like to acknowledge the help of Irina Tchoumanova who translated Shen's paper from Russian. The author wishes also to acknowledge Jan Reimann for insightful comments, and David Cyriac, Dave Doty, Xiaoyang Gu, María López-Valdés, Philippe Moser and two anonymous reviewers for very helpful comments on earlier drafts of this paper.

8. 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. Random functions. *Unpublished Lecture, Computability and Randomness, Cordoba*, 2004.

[9] A. Kh. Shen′. The frequency approach to the definition of a random sequence. *Semiotika i Informatika*, 18:14–42, 1982. (In Russian.).

[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.