Inductive Success of Semimeasures on Random Sequences

Damini Kusum

July 2025

Department of Philosophy
Carnegie Mellon University
Pittsburgh, PA 15213

Thesis Committee:

Prof. Francesca Zaffora Blando (Chair)
Prof. Krzysztof Mierzewski

Submitted in fulfillment for the requirement of
Master of Science in Logic, Computation and Methodology.

Abstract

For sequential prediction tasks, Solomonoff's [30] celebrated result guarantees the inductive success of the "simplicity" prior by showing that the predictive probabilities induced by any optimal c.e. semimeasure converge to the posterior of the true probability almost surely. Hutter and Muchnik [13] have shown that there exists an optimal c.e. semimeasure for which one can construct a Martin-Löf sequence along which this optimal semimeasure fails to converge to the uniform measure. Since the study of algorithmic randomness can be used to characterize "almost-sure" convergence theorems, one can ask whether stronger notions of randomness suffice for Solomonoff convergence. We show that the minimal strength of randomness required for off-sequence convergence is at most 2-computable randomness; this establishes an upper bound on the weakest notion of randomness that ensures the success of Solomonoff induction. Furthermore, due to its rejection of Turing complete reals as random, difference randomness prima facie appears to be the weakest notion that could guarantee convergence. Contrary to this intuition, we prove a negative result with respect to the uniform measure. The machinery of this proof can be extended to show that any notion of randomness that does not guarantee convergence of all c.e. supermartingales will always yield a negative result for Solomonoff convergence.

Contents

1 Introduction

1.1 Outline

When faced with competing explanations of a phenomenon, the principle of Occam's razor recommends the simplest explanation as the correct one. This preference for simplicity has long been a recurring theme in empirical inquiry, at the heart of which lies the task of making inferences from observed data. This thesis focuses on a specific form of such inference: sequential prediction. Suppose a sequence α is generated from a finite alphabet by an unknown probabilistic source. We have the following problem at hand: given the first n observations of the sequence, what is the value of the next entry?

Adopting what is often described as an Occam-style simplicity bias, Ray Solomonoff [30] proposed an idealized solution to this problem. His framework — now known as Solomonoff induction —- offers a rigorous method for sequential predictions. Solomonoff induction, while not implementable in practice, has an important place in computer science, machine learning, and information theory. It quintessentially formalizes the idea of using compressibility as a guiding principle in learning-theoretic settings—an idea that is of practical consequence. As Ilya Sutskever, Chief Scientist at OpenAI, aptly put it [29]: "compression equals generalization". The key idea is that the more succinctly a model can describe the observed data, the better it can be expected to generalize to unseen instances [[25], [33]].

Solomonoff demonstrated the reliability of his method through the following much-celebrated theorem: in the long run, simpler extrapolations of the past data *generally* lead to truth. However, the guarantee holds with probability-one relative to the true probability — and it is this qualifier that prompts further scrutiny. The aim of this thesis is to investigate whether the tools of algorithmic randomness can sharpen Solomonoff's result by eliminating this qualifier, precisely identifying the sequences on which the convergence is guaranteed.

The thesis is organized as follows. The remainder of this section reviews the necessary measure-theoretic and computability-theoretic preliminaries. In Section 2, we introduce the notion of a Solomonoff prior, whose conditionalization induces predictions along an observed sequence. Section 2 also relates these predictors to a class of functions known as semimeasures. In Section 3, we provide a brief introduction to algorithmic randomness and define several notions that will be used later in the thesis. In Section 4, we draw the connection between Solomonoff induction and algorithmic randomness, thereby motivating the central research question of this thesis. In particular, we are interested in determining the degree of randomness required for successful learning within Solomonoff's framework. Section 5 discusses a key negative result due to Hutter and Muchnik [13],

which shows that the canonical notion of randomness—Martin-Löf randomness—is insufficient for Solomonoff induction to yield asymptotically correct predictions. In Section 6, we examine whether stronger notions of randomness can support Solomonoff's inductive program, and we show that both 2-computable randomness and strong density randomness (SDR) suffice for this purpose. Section 7 establishes a negative result for difference randomness, which, despite being stronger than Martin-Löf randomness, still fails to guarantee success, thereby further narrowing the class of candidate notions. Finally, in Section 8, we discuss several open questions that, if resolved, would further clarify the results of this thesis, and we outline possible directions for future work.

1.2 Preliminaries

Throughout this thesis, we restrict our attention to Cantor space. This means that the data relevant to our prediction task consists of infinite binary sequences. Let $\{0,1\}^\infty$ and $\{0,1\}^*$ be the sets of all infinite binary sequences and finite binary strings, respectively. Given a sequence $\alpha \in \{0,1\}^\infty$, we use α_n to denote the n-th bit of α, and $\alpha_{1:n}$ to denote its first n bits. We say a bit $\alpha'_n \in \{0,1\}$ is off-sequence if $\alpha'_n \neq \alpha_n$, and on-sequence otherwise. The empty string is denoted by ϵ.

Cantor space is topologized by the collection of basic open sets of the form $[\sigma]$, where $\sigma \in \{0,1\}^*$ is a binary string and

$$[\sigma] = \{\alpha \in \{0,1\}^\infty : \sigma \prec \alpha\}$$

Here, $\sigma \prec \alpha$ means that σ is a prefix of α, or equivalently, α is an extension of σ. Let $A \subseteq \{0,1\}^*$. Then,

$$[A] = \bigcup_{\sigma \in A} [\sigma]$$

A is also called the generator set of $[A]$.

In our inductive setting, we also assume the existence of a true data-generating process, which we identify with a probability measure on $\{0,1\}^\infty$. Let this measure be denoted by μ. The probability space we will work in is given by the triple $(\{0,1\}^\infty, \mathcal{B}(\{0,1\}^\infty), \mu)$, where $\mathcal{B}(\{0,1\}^\infty)$ is the Borel σ-algebra on Cantor space — that is, the smallest σ-algebra containing all basic open sets. The measure μ assigns probabilities to sets in $\mathcal{B}(\{0,1\}^\infty)$, and therefore, is a Borel measure. Any function $f : \{0,1\}^* \to [0,1]$ for which $f(\epsilon) = 1$ and $f(\sigma) = f(\sigma 0) + f(\sigma 1)$ for all $\sigma \in \{0,1\}^*$ can be extended to a Borel measure on $\{0,1\}^\infty$ by Carathéodory's Extension Theorem. Any Borel measure μ is uniquely determined by the value it assigns to the basic open sets, $\mu([\sigma])$. For convenience, we write $\mu(\sigma)$ to mean $\mu([\sigma])$.

Definition 1.2.1. *The canonical Borel measure on $\{0,1\}^\infty$ is the uniform or Lebesgue measure,*

which is given by

$$\lambda(\sigma) = 2^{-|\sigma|}$$

for all $\sigma \in \{0,1\}^*$, where $|\sigma|$ is the length of the string.

In this thesis — and as is often the convention — the measure λ will refer to the uniform measure.

In a Bayesian framework, inductive learning proceeds via conditionalization. Within the purview of our inductive task, we will be looking at sequences $\alpha \in \{0,1\}^\infty$ that are in the support of μ, so the conditional probabilities $\mu(\alpha_n|\alpha_{1:n-1})$ are always well defined. However, for an off-sequence bit α'_n, it is possible that $\mu(\alpha'_n \mid \alpha_{1:n-1}) = 0$. Another important foundational assumption in this thesis is that we are operating within a computable environment. In our setting, this translates to assuming that the probability measure from which the data is sampled is computable. Given this, we introduce a few computability-theoretic notions that are useful for defining a computable measure and related notions.

Definition 1.2.2. *A real $r \in \mathbb{R}$ is said to be computable if it can be computably approximated within any arbitrary degrees of precision. Equivalently, a real is computable if there is a computable sequence of rationals $\{q_n\}_{n\in\mathbb{N}}$ such that $|q_n - r| < 1/n$ for all $n \in \mathbb{N}$.*

Definition 1.2.3. *A real $r \in \mathbb{R}$ is said to be left c.e. if there is a monotonically increasing computable sequence of rationals $\{q_n\}_{n\in\mathbb{N}}$ such that $lim_{n\to\infty} q_n = r$. A real $r \in \mathbb{R}$ is said to be right c.e. if there is a monotonically decreasing computable sequence of rationals $\{q_n\}_{n\in\mathbb{N}}$ such that $lim_{n\to\infty} q_n = r$.*

Definition 1.2.4. *A limit computable real is the limit of a computable sequence of rational numbers. Such a real admits computable approximations, but there is no computable modulus of convergence.*

Remark: Note that Cantor space and the interval [0,1] in $\mathbb{R}$ are near-homeomorphic. Specifically, any sequence α can be translated to a real $0.\alpha \in [0,1]$. In most cases, this correspondence allows us to move fluidly between reals and sequences. A sequence $\alpha \in \{0,1\}^\infty$ is c.e. if there are computable, increasing approximations α^t_n for the n-th bit of α. Suppose α is a c.e. sequence. Then we can approximate $0.\alpha$ with $q_t = 0.(\alpha^t_{1:t})$. Therefore, $0.\alpha$ is a left c.e. real.

Definition 1.2.5. *A real-valued (partial) function $f : D \to \mathbb{R}$, defined on domain D, is (partial) computable if the values the output assumes are computable, uniformly in the input. Formally, if f is computable, then there exists a function $g : D \times \mathbb{N} \to \mathbb{R}$ such that $|g(x,n) - f(x)| < 1/n$ for all $n \in \mathbb{N}$.*

A real-valued (partial) function $f : D \to \mathbb{R}$, defined on domain D, is (partial) c.e. if the values are left c.e reals, uniformly in the input.

We can obtain the definition of co c.e. functions in a similar way by replacing "left c.e." with "right c.e.". Now, we can define what it means for μ to be computable. A measure on $\{0,1\}^\infty$ is said to be computable when the measure of each string is a computable real, uniformly in the string.

Computable relations, on the other hand, generate what is known as the arithmetic hierarchy:

Definition 1.2.6. *A set $A \subseteq \{0,1\}^\infty$ is a Σ^0_n class if there is a computable relation R such that*

$$A = \{\alpha \in \{0,1\}^\infty : \exists q_1 \forall q_2 \dots Q q_n R(\alpha_{1:q_1}, \alpha_{1:q_2}, \dots, \alpha_{1:q_n}\}$$

where Q is $\forall$ if n is even, Q is $\exists$ if n is odd. The complement of Σ^0_n class is a Π^0_n class. If a class is both Σ^0_n and Π^0_n, then it is a Δ^0_n class.

Limit computable reals, as defined in 1.2.4, are exactly those that can be expressed at the level of Δ^0_2 in the arithmetic hierarchy. There also exist functions whose values fall strictly between the levels of Δ^0_1 and Δ^0_2. One such example is an ω-c.e. function.

Definition 1.2.7. *A function $f : \mathbb{N} \to \mathbb{N}$ is called ω-c.e. if there exists a computable approximation $\phi(x,s)$ such that $\lim_{s\to\infty} \phi(x,s) = f(x)$ for all $x \in \mathbb{N}$, and the number of changes in the value of $\phi(x,s)$ as s increases is bounded by a computable function.*

Lastly, to conclude our discussion of computability-theoretic preliminaries, we state one of the most fundamental theorems in computability theory, known as Post's Theorem [26]. The Turing jump of a decision problem X — denoted by X' — is equivalent to the halting problem relative to an oracle machine that solves X. Subsequently, the Turing jump of the empty set, $\varnothing'$, encodes the halting problem. Post's theorem establishes a relationship between the Turing jump operator and the arithmetic hierarchy in the following way:

Theorem 1.2.8. ***(Post's Theorem)*** *For $n \geq 0$:*

1. *A set A is $\Sigma^0_n \Leftrightarrow A$ is c.e. in $\varnothing^{(n-1)}$*
2. *The set $\varnothing^{(n-1)}$ is $\Sigma^0_{(n-1)}$-complete.*
3. *A set A is $\Delta^0_n \Leftrightarrow S$ is computable from $\varnothing^{(n-1)}$*

Namely, this tells us how much computational power is required for an oracle machine to compute functions of certain complexity in the arithmetic hierarchy. This will become particularly helpful in defining relativized notions randomness later in the thesis.

2 Solomonoff's Theory of Inductive Inference

2.1 Solomonoff Prior

If one takes the Bayesian approach to sequential predictions, the first task one is confronted with is choosing the correct prior over the universe of hypotheses. This prior would induce the probability of observing the first n bits of any sequence α. Let's denote this probability as $M(\alpha_{1:n})$. To compute $M(\alpha_{1:n})$, we consider all possible hypotheses that are compatible with the observed data. Each hypothesis is a program that is encoded as a binary string. To formally capture the simplicity of these hypotheses, we will employ the notion of Kolmogorov complexity.

We first turn our attention to Turing machines, which become central in our computability-theoretic setting. The Turing machine we are interested is a monotone universal Turing machine [[21], [32]]. Let's call this machine Q. Universal Turing machines are those that can simulate any computable function that is calculable by a standard Turing machine. Moreover, the input tape of a monotone Turing machine is unidirectional, which ensures that nothing more than the bits of the input program are read. The output tape is also unidirectional, which prevents it from overwriting any previous output. This ensures that if $p \preceq q$, then $Q(p) \preceq Q(q)$. Then, Kolomogorov complexity is defined by

$$K_Q(\sigma) = \min_p\{|p| : Q(p) = y \wedge \sigma \prec y\}$$

In our setting of probabilistic prediction of infinite binary sequences, the hypotheses we consider can be identified with the class $\mathcal{P}$ of minimal programs, defined with respect to K, that generate infinite binary sequences beginning with the initial segment $\alpha_{1:n}$ [[21], [12]]. In computability-theoretic terms, the task of determining $M(\alpha_{1:n})$ naturally translates to the following question: given a universal Turing machine Q, what is the probability that a randomly chosen program from $\mathcal{P}$ appears on the input tape? Thus, we can interpret $M(\alpha_{1:n})$ as the sum of the probabilities of all programs for which Q produces an output that starts with $\alpha_{1:n}$.

Since we are interested in determining the probability of seeing an arbitrary program on the input tape, our prediction task is best modeled with respect to universal probabilistic Turing machines that are equipped with an additional read-only tape. This additional tape consists of a sequence of independent random bits. Let $\mathcal{P}$ be the set of minimal programs for σ. Given access to independent random bits, the probability of seeing a program p in $\mathcal{P}$ on the input tape is $2^{-|p|}$, where $|p|$ is the length of the program. The construction, so far, can be formalized as follows:

$$M(\sigma) = \sum_{p \in \mathcal{P} : Q(p) \in [\sigma]} 2^{-|p|}$$

If the length of the program is taken to be a reasonable gauge of the simplicity of the hypothesis,

then the above formulation is clearly informed by the intuition that a simpler hypothesis should receive a greater probability. It is easy to see that the above construction of M implicitly represents a distribution over $\{0,1\}^\infty$. As the observer gains more information, $M(\alpha_{1:n})$ is updated in line with Bayes' rule. Given a binary string $\alpha_{1:n}$, the prior M induces a predictive probability over the value of $\alpha_{n+1} \in \{0,1\}$.

The monotonicity of the machine Q — and the corresponding minimality of the programs in $\mathcal{P}$ — produces two crucial features of M. First, for any σ, since the set of the minimal programs in trivially prefix-free, we are guaranteed that $M(\sigma) = \lambda\left(\bigcup_{p\in\mathcal{P}:Q(p)\in[\sigma]}[p]\right) \leq 1$. Second, M is superadditive: that is, $M(\sigma) \geq M(\sigma 0) + M(\sigma 1)$. To see this, let $\mathcal{J}$ and $\mathcal{F}$ be the sets of minimal programs that generate outputs beginning with $\sigma 0$ and $\sigma 1$, respectively. Observe that if j is a minimal program in $\mathcal{J}$, then either j is in $\mathcal{P}$ or a prefix of j is in $\mathcal{P}$. If it is the case that j is in $\mathcal{P}$, then it certainly cannot be in $\mathcal{F}$. Otherwise, if a prefix $\nu \prec j$ belongs to $\mathcal{P}$, then the probability of observing ν on the input tape is at least twice that of observing j on the input tape. From this, we can conclude:

$$\sum_{p\in\mathcal{P}:Q(p)\in[\sigma]} 2^{-|p|} \geq \sum_{j\in\mathcal{J}:Q(j)\in[\sigma 0]} 2^{-|j|} + \sum_{f\in\mathcal{F}:Q(f)\in[\sigma 1]} 2^{-|f|}$$

Now, is the prior M constructed so far computable? It is not. The justification of its incomputability is akin to that of the halting problem. For a given string σ and an arbitrary program g that has produced a prefix of σ on the output tape, we cannot establish that $g \notin \mathcal{P}$; if we could, we would be able to solve the halting problem. However, if a program p produces σ on the output tape, it can be enumerated into $\mathcal{P}$. This means $M(\sigma)$ can be approximated from below; M is a c.e. function in the sense of Definition 1.2.5.

2.2 Optimal Semimeasures

In the last subsection, we introduced Solomonoff's prior M as a foundational tool for sequential prediction within a Bayesian framework. This prior captures a weighted distribution over the universe of hypotheses, with a bias toward simpler programs, and is grounded in the architecture of universal monotone machines. The superadditivity of the Solomonoff prior corresponds to a general class of functions, known as semimeasures. Semimeasures are powerful predictive tools that can be defined in the following way:

Definition 2.2.1. *A semimeasure is a function $\rho : \{0,1\}^* \to [0,1]$ that satisfies the following properties:*

1. $\rho(\epsilon) \leq 1$ *where ϵ is the empty string*
2. $\rho(\sigma) \geq \rho(\sigma 0) + \rho(\sigma 1)$

Observe that replacing the instances of "$\geq$" in Definition 2.2.1 with "=" yields probability measures. Thus, one can think of semimeasures as 'defective' probability measures. An important property of certain semimeasures is their optimality.

Definition 2.2.2. *A c.e. semimeasure ρ is optimal if for all c.e. semimeasures ρ', there exists a constant $c > 0$ such that for all $\sigma \in \{0,1\}^*$, we have $\rho(\sigma) \geq c\rho'(\sigma)$.*

Zvonkin and Levin [35] showed that the class of all c.e. semimeasures is computably enumerable. Let $\mathcal{O} = \{\upsilon_1, \upsilon_2, \upsilon_3, ...\}$ be the class of all c.e. semimeasures. Additionally, let g_i be a c.e. sequence of weights such that $\sum_{i\in\mathbb{N}} g_i \leq 1$. Then, an optimal c.e. semimeasure, say ξ, can be represented as a Bayesian mixture of semimeasures in $\mathcal{O}$, with weights[1] assigned based on the complexity of the semimeasure. We get:

$$\xi(\sigma) = \sum_{i\in\mathbb{N}} g_i \upsilon_i(\sigma)$$

It turns out that M, the prior defined for Solomonoff induction, always multiplicatively dominates ξ, and consequently, is also an optimal c.e semimeasure [12]; M can be represented as a Bayesian mixture of c.e. semimeasures.

Optimal semimeasures are of particular interest within Solomonoff's inductive program as the following convergence result holds for all optimal semimeasures.

Theorem 2.2.3. ***(Solomonoff [30])*** *Suppose μ is a computable measure and M an optimal c.e. semimeasure. Then, for μ-almost every $\alpha \in \{0,1\}^\infty$:*

$$\sum_{b\in\{0,1\}} (M(b|\alpha_{1:n-1}) - \mu(b|\alpha_{1:n-1}))^2 = 0 \text{ and } \frac{M(\alpha_n|\alpha_{1:n-1})}{\mu(\alpha_n|\alpha_{1:n-1})} \longrightarrow 1$$

Note that there is a subtle difference between the two results. The result in terms of ratios, established by Gács [21], is only concerned with on-sequence predictions since $\frac{M(\alpha_n|\alpha_{1:n-1})}{\mu(\alpha_n|\alpha_{1:n-1})}$ may be undefined off-sequence. Regardless, these are powerful results as they guarantee the success of Solomonoff's program in the long run. Specifically, the theorem states that if α is sampled from a computable measure μ, the predictive probabilities induced by M converge to the true probability almost surely. This yields that the set of sequences on which the convergence result holds is a μ-measure-one set.

However, this leaves open the question of what happens for individual sequences. For an arbitrary data stream drawn from the underlying probability space, convergence — as in Theorem 2.2.3 —

[1] These weights will be discussed in more detail in Section 4.1

can no longer be assured. This is precisely where algorithmic randomness comes into the picture: it allows us to capture convergence properties of individual sequences.

3 Algorithmic Randomness

3.1 Introduction: Randomness as Typicality

The field of algorithmic randomness, at its core, is concerned with formalizing what it means for individual objects in infinite product spaces to be random [10]. The objects of our study, as is customary for the field, are infinite binary sequences. "Randomness" is certainly a colloquial term; we appeal to randomness in our daily lives. Within algorithmic randomness, these pre-theoretic ideas are made rigorous using tools of computability theory and measure theory. Crucially, different pre-theoretic notions motivate three main paradigms within which randomness is defined: typicality, unpredictability, and incompressibility. We discuss both the incompressibility and unpredictability paradigms in greater detail in the later sections. For now, we focus on the typicality paradigm, within which Martin-Löf randomness — widely considered to be the standard notion of randomness — emerges.

Associating randomness with typicality, random sequences are those that exhibit global regularities with respect to a given probability measure. This underscores an important feature of defining randomness: the choice of measure relative to which randomness is defined becomes fundamental to whether or not a sequence is considered random. Typicality is the complement of rarity, implying that a random sequence must not satisfy any rare property. This can be explicated in measure-theoretic terms as the avoidance of measure-zero sets. However, for many commonly-studied probability measures such as the uniform measure, every sequence belongs to at least one measure-zero set – the singleton set that only contains itself. To circumvent a vacuous definition of randomness, we can view randomness as the absence of every effectively specifiable property of measure-zero. This is tantamount to saying that random sequences are those that satisfy every measure-one property that can be algorithmically described. This approach was pioneered by Per Martin-Löf who formalized these algorithmic descriptions by associating them with every performable statistical test whose components are defined at different levels of the arithmetic hierarchy. Random sequences, then, are those that pass all such tests. The precise definition is as follows:

Definition 3.1.1. ***(Martin-Löf [22])*** *Let μ be a computable measure. A μ-Martin-Löf test is a sequence $\{\mathcal{U}_i\}_{i\in\mathbb{N}}$ of uniformly Σ^0_1 classes such that $\mu(\mathcal{U}_i) \leq 2^{-i}$. A sequence $\alpha \in \{0,1\}^\infty$ is μ-Martin-Löf random if $\alpha \notin \bigcap_{i\in\mathbb{N}} \mathcal{U}_i$ for all μ-Martin-Löf tests.*

Given a measure μ, there are only countably many μ-Martin-Löf tests. It follows that the set of all μ-Martin-Löf random sequences has μ-measure one. To obtain uniformly Σ^0_1 classes, in the background, we fix an enumeration, $\mathcal{W} = \{\mathcal{W}_n\}_{n\in\mathbb{N}}$, of c.e. sets of strings. In Definition 3.1.1, each component of the test is uniformly Σ^0_1 if there exists a computable function $g : \mathbb{N} \to \mathbb{N}$ that

picks out the index of $\mathcal{W}$ such that $\mathcal{U}_i = [\mathcal{W}_{g(i)}]$. Further, a μ-Martin-Löf random α does not belong to $\bigcap_{i\in\mathbb{N}} \mathcal{U}_i$[2], equating Martin-Löf randomness with membership in all Σ^0_2 classes of effective μ-measure one. This implies that, whenever μ is a Bernoulli measure, μ-Martin-Löf random sequences obey many quintessential measure-one statistical laws that can be expressed as Σ^0_2 classes, such as the law of large numbers and iterated logarithm.

When it is clear from the context, we will omit specifying the measure with respect to which randomness is defined.

3.2 Randomness Hierarchy

Since the descriptional complexity of the measure-one properties discussed in the last section can vary, there is no singular notion of randomness. In effect, there is an infinite hierarchy of randomness notions that correspond to different effectivity constraints. In this section, we will explore a few examples of randomness notions that are stronger than Martin-Löf randomness.

Looking at Definition 3.1.1, there are three primary axes on which the definition can be altered to yield other notions of randomness.

- Effectiveness condition: The effectivity constraint imposed on the way the test sequence $\{\mathcal{U}_i\}_{i\in\mathbb{N}}$ is defined can be varied. Most commonly the sequence consists of uniformly Σ^0_1 classes, as is the case for the definition of Martin-Löf randomness.

- Rate of Convergence: The speed of convergence of $\mu(\mathcal{U}_i)$ to zero can be calibrated differently. For Martin-Löf tests, the measures of the components of the test converge to zero at a computable rate.

- Passing condition: A sequence passes a Martin-Löf randomness test given by $\{\mathcal{U}_i\}_{i\in\mathbb{N}}$ as long as it does not belong to at least one of the test components. In contast, the Solovay passing condition is significantly stricter: a sequence Solovay-passes a randomness test if it belongs to only finitely many of the $\mathcal{U}_i$. It worth mentioning that when the tests are nested[3], the two passing conditions are equivalent.

To obtain some familiarity with the randomness hierarchy, we will look at some instances of randomness notions that can be derived by modifying the definition of Martin-Löf randomness. Within this hierarchy, Martin-Löf random sequences are also known as 1-random sequences. There is a natural way to obtain stronger notions of randomness — defined at higher levels in the arithmetic hierarchy

[2]The intersection of uniformly Σ^0_1 classes is a Π^0_2 class.

[3]That is, $\mathcal{U}_{i+1} \subseteq \mathcal{U}_i$.

— by relativizing 1-randomness. Consider a Turing machine that produces a Martin-Löf test. If we augment this Turing machine with a halting oracle, the tests produced by this machine yield the notion of 2-randomness. Specifically, the components of this test are uniformly $\Sigma_1^{0,\varnothing'}$ classes. More generally, n-randomness can be defined as 1-randomness relativized to $\varnothing^{n-1}$, where $\varnothing^{n-1}$ is the $(n-1)$-st Turing jump of the empty set, in the same manner.

Definition 3.2.1. ***(Kurtz [17])*** *Let μ be a computable measure. A μ-Σ_n^0 test is a sequence $\{\mathcal{U}_i\}_{i\in\mathbb{N}}$ of uniformly Σ_n^0 classes such that $\mu(\mathcal{U}_i) \leq 2^{-i}$. A sequence $\alpha \in \{0,1\}^\infty$ is μ n-random if $\alpha \notin \bigcap_{i\in\mathbb{N}} \mathcal{U}_i$ for all μ-Σ_n^0 tests.*

Next, we can also obtain the notion of weak 2-randomness by taking the Definition 3.1.1 and modifying it to require only that the measures of the $\mathcal{U}_i$'s go to zero as i approaches infinity [17]. Weak 2-randomness interpolates 2-randomness and Martin-Löf randomness, and is defined in the following way:

Definition 3.2.2. ***(Kautz [14])*** *Let μ be a computable measure. A μ-weak 2-random test is a sequence $\{\mathcal{U}_i\}_{i\in\mathbb{N}}$ of uniformly Σ_1^0 classes such that $\lim_i \mu(\mathcal{U}_i) = 0$. A sequence $\alpha \in \{0,1\}^\infty$ is μ-weak 2-random if $\alpha \notin \bigcap_{i\in\mathbb{N}} \mathcal{U}_i$ for all μ-weak 2 tests. In other words, μ-weak 2-random sequences are those that belong to every Σ_2^0 class of measure-one.*

Another notion of randomness, which is famously incomparable with weak 2-randomness, comes from Osvald Demuth's line of research. Demuth introduced randomness tests with moving components, permitting a computably bounded number of changes to each test component.

Definition 3.2.3. ***(Demuth [7])*** *Let μ be a computable measure. A Demuth test is a sequence $\{\mathcal{U}_i\}_{i\in\mathbb{N}}$ of Σ_1^0 classes such that $\mu(\mathcal{U}_i) \leq 2^{-i}$ and there is a ω-c.e. function g for which $\mathcal{U}_i = [\mathcal{W}_{g(i)}]$. A sequence $\alpha \in \{0,1\}^\infty$ is μ-Demuth random if for almost every $\mathcal{U}_i$, $\alpha \notin \mathcal{U}_i$ for all μ-Demuth tests.*

Note that the definition of Demuth tests closely resembles that of Martin-Löf tests, except that the indexing function g is ω-c.e. rather than computable. We get a slightly weaker notion of randomness if we relax the passing condition to require the sequence to pass the tests in the Martin-Löf sense.

Definition 3.2.4. ***(Kučera and Nies [16])*** *Let μ be a computable measure. A Demuth test is a sequence $\{\mathcal{U}_i\}_{i\in\mathbb{N}}$ of Σ_1^0 classes such that $\mu(\mathcal{U}_i) \leq 2^{-i}$ and there is a ω-c.e. function g for which $\mathcal{U}_i = [\mathcal{W}_{g(i)}]$. A sequence $\alpha \in \{0,1\}^\infty$ is μ-weak Demuth random if $\alpha \notin \bigcap_{i\in\mathbb{N}} \mathcal{U}_i$ for all μ-Demuth tests.*

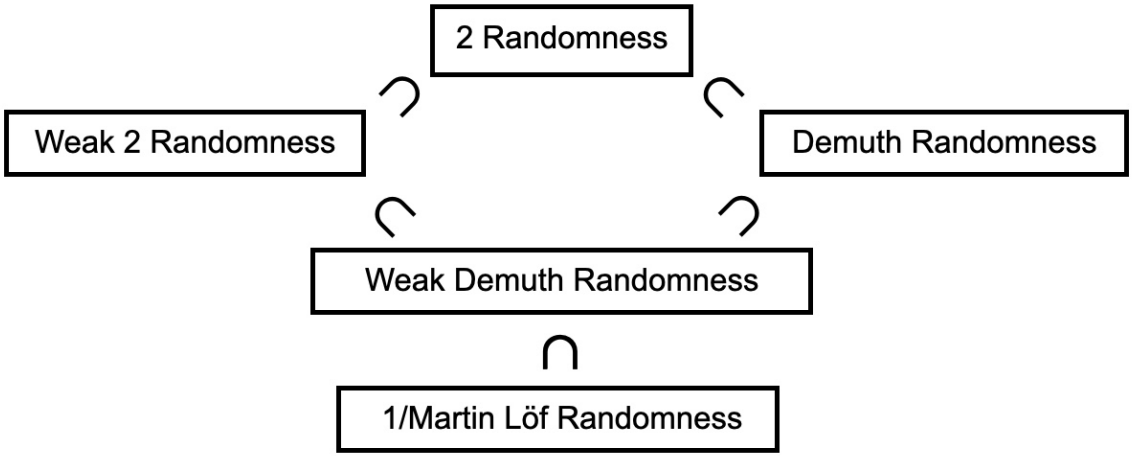

Figure 1: A Small Segment of the Hierarchy

3.3 Randomness as Unpredictability

Another approach, known as the unpredictability paradigm, arguably bears a more intuitive connection to the inductive framework discussed thus far. As the name suggests, this paradigm characterizes random sequences as those for which predictions of the next bit cannot consistently (in the long run) be accurate. From a gambler's perspective, this implies that no betting strategy that makes bets on each successive bit of a random sequence should be able to make unbounded gains. The accumulated capital of the betting strategy is tracked by functions known as (super)martingales.

Definition 3.3.1. *Given a computable measure μ, a function $d : \{0,1\}^* \to \mathbb{R}^{\geq 0}$ is a μ-martingale if it satisfies the following conditions:*

- *(Impossibility Condition) Whenever $\mu(\sigma) > 0$, $d(\sigma)$ is defined.*
- *(Fairness condition)*

$$d(\sigma)\mu(\sigma) = d(\sigma 1)\mu(\sigma 1) + d(\sigma 0)\mu(\sigma 0) \tag{1}$$

If $d(\sigma)$ is undefined, $\mu(\sigma) = 0$. In that case, we take $d(\sigma)\mu(\sigma)$ to be 0.

If we replace "=" with "$\geq$" in the fairness condition for martingales, we get a supermartingale. Observe that the condition ensures that if the sequence is sampled from a probability space with measure μ, the expected value of the capital made by betting on an additional bit does not exceed the current capital the bettor possesses. For supermartingales, relaxing the fairness condition in (1) is practically the same as allowing the bettor to discard some amount of his capital at each round.

The gambler starts with the initial capital of $d(\epsilon)$. Betting on each consecutive bit of the sequence, the gambler's betting strategy must not succeed on a random sequence. Formally, a supermartingale is said to succeed on a sequence $\alpha \in \{0,1\}^\infty$ if $\limsup_n d(\alpha_{1:n}) = \infty$. We get the following characterization of Martin-Löf randomness in terms of supermartingales.

Theorem 3.3.2. ***(Schnorr [27])****: Let μ be a computable probability measure. $\alpha \in \{0,1\}^\infty$ is μ-Martin-Löf random if and only if no c.e. μ-supermartingale succeeds on it.*

Note that one get an equivalent characterization, as in Theorem 3.3.2, in terms of martingales.

However, under the unpredictability paradigm, computable martingales are the most obvious instances of effectivized martingales. Subsequently, defining randomness in terms of computable martingales yields the notion of computable randomness. Schnorr [27] defined computable randomness in the following way:

Definition 3.3.3. *A sequence β is μ-computably random if there is no computable μ-martingale that succeeds on it.*

It is easy to see that computable randomness is a weaker notion of randomness than Martin-Löf randomness. Perhaps less immediate is that computable randomness not only ensures that a gambler cannot accumulate an infinite amount of capital through betting, but also that the profits stabilize in the long run. In other words, computable randomness guarantees the convergence of computable martingales.

Theorem 3.3.4. *α is μ-computably random if and only if $\lim_n d(\alpha_{1:n})$ exists and is finite for all computable μ-martingales d.*

With this, we end our brief introduction to algorithmic randomness. Now that we have understood the theories of algorithmic randomness and Solomonoff induction, we can, in the next section, examine points of contact between the two.

4 Connection Between Algorithmic Randomness and Solomonoff Induction

4.1 Some History: Compressibility and Kolmogorov Complexity

The notion of compressibility serves as a prefatory conduit between the theory of algorithmic randomness and Solomonoff's inductive program. The incompressibility paradigm of randomness encodes a coder's approach to defining what a random sequence is. The (plain) Kolmogorov complexity $C(\sigma)$ of a string σ is defined as the length of the shortest computer program that outputs σ. A string σ is said to be incompressible if its complexity $C(\sigma)$ is close to the length of σ itself. Intuitively, this means that the most efficient way to describe σ is simply to reproduce it in its entirety. However, to derive a definition of randomness that is analogous to Martin-Löf randomness within the incompressibility paradigm, we need to transition from C to prefix-free Kolmogorov complexity K. K is defined using self-delimiting machines, enabling measure-theoretic tools for analyzing the complexity of strings. A string σ is said to be K-incompressible if $K(\sigma) \geq c + |\sigma|$ for some constant c. Schnorr [28] established that an infinite binary sequence α is Martin-Löf random if and only if all of its initial segments are K-incompressible.

Although Kolmogorov complexity now bears Kolmogorov's name, the notion of Kolmogorov complexity predates Kolmogorov; it first appeared in the context of Solomonoff's theory of induction [31]. As Vitányi and Li somewhat peculiarly state, "inductive reasoning was the midwife that stood at the cradle of Kolmogorov complexity" [20]. For Solomonoff, compressibility is a formal measure of simplicity. As we have already seen, Kolmogorov complexity is used to capture the set of minimal programs corresponding to a given sequence of data. In fact, because the shortest programs dominate the Solomonoff prior, the prior defined in Section 2 can be expressed in terms of Kolmogorov complexity [12].

$$M(\sigma) = 2^{-K(\sigma)+O(K(|\sigma|))}$$

Note that this formulation also makes explicit why M is incomputable. The function K is well known to be a co-c.e. [Theorem 2.3.2, [21]], and since $M(\sigma)$ is inversely proportional to $K(\sigma)$, the resulting M is a c.e. semimeasure.

Whereas Kolmogorov associates randomness with incompressibility, Solomonoff exploits departures from randomness to construct a prior that yields asymptotically accurate predictions. Having observed a historical connection between Kolmogorov's and Solomonoff's work, the next subsection will develop the crux of the research question that we seek to address in this thesis.

4.2 Motivation: Algorithmic Randomness and Almost-sure Convergence

An actively expanding line of research in algorithmic randomness investigates the degree of randomness required for effective versions of classical "almost everywhere" or "almost surely" theorems in analysis and ergodic theory [11]. A theorem holds almost everywhere or almost surely on a measure space if it holds at all points except possibly on a set of measure zero. As we have already seen, random sequences satisfy measure one properties that can be effectively described. While descriptional complexity may vary, such effectivization of measure one sets — and equivalently, effectivization of null sets — is suggestive of the idea that algorithmic randomness can be equated with almost surely convergence.

For example, a well-known result in classical analysis states that any non-decreasing function $f : [0,1] \to \mathbb{R}$ is differentiable almost everywhere in [0,1]. Demuth [8] was the first to ask the following question: what amount of randomness must a real exhibit to make effective versions of such functions differentiable at that point? For computable functions, computable randomness has been shown to be the appropriate level of randomness for the theorem; all non-decreasing computable functions $f : [0,1] \to \mathbb{R}$ are differentiable on computably random reals [5]. Observe that in asking such a question, one must specify the effectivity constraints imposed on the mathematical objects to which the classical theorem applies. In our example, instead of restricting attention to computable functions, one might consider c.e. functions. Different choices of effectivity constraints yield different randomness notions under which the effectivized theorem holds. Conversely, one may also approach the problem in reverse: by selecting a particular effective version of a classical theorem, it is possible to tailor a corresponding randomness notion for which the random sequences satisfy the conclusion of the theorem. Consequently, this approach enriches the hierarchy of randomness notions, as we will see throughout the thesis.

The goal of this thesis is to extend this line of research into the learning-theoretic setting of Solomonoff induction. Accordingly, the question arises: what level of randomness, relative to the true probability measure, suffices to guarantee the almost sure convergence stated in Theorem 2.2.3. In other words, we aim to identify a notion of randomness such that the set of sequences deemed random under this definition forms a subset of the set on which Solomonoff convergence holds.

4.3 Semimeasures and Supermartingales

Just as compressibility is a conduit between randomness and induction, the notion of predictability is the conduit between betting strategies and sequential predictors. If we think of semimeasures as predictors, it should not come as a surprise that supermartingales and semimeasures are inherently related. In particular, semimeasures can be represented as supermartingales and vice versa. The

following is a useful fact[4] about supermartingales:

Fact 4.3.1. *Given a computable measure μ, let d be a μ-supermartingale such that $d(\epsilon) \leq 1$. Then, $d(\sigma)\mu(\sigma)$ is a semimeasure. Similarly, let ξ be a semimeasure. Then, $\frac{\xi(\sigma)}{\mu(\sigma)}$ defines a μ-supermartingale d such that $d(\epsilon) \leq 1$.*

From this, we obtain a natural correspondence between supermartingales and semimeasures, and by extension, between c.e. supermartingales and left c.e. semimeasures. In order to reformulate the definition of Martin-Löf randomness in terms of semimeasures, we require a notion of optimality for supermartingales.

Definition 4.3.2. *A c.e. supermartingale d is optimal if for each c.e. supermartingale m, there exists a constant c such that $cd(\sigma) \geq m(\sigma)$ for all $\sigma \in \{0,1\}^*$.*

In other words, just like semimeasures, an optimal c.e. supermartingale multiplicatively dominates other c.e. supermartingales. Since it is possible to construct a computable enumeration of c.e. supermartingales, we have the following theorem due to Schnorr:

Theorem 4.3.3. ***(Schnorr [27])*** *There exists an optimal c.e. supermartingale.*

It can be inferred from the multiplicative dominance of the optimal c.e. supermartingales that, given a sequence, boundedness of an optimal c.e. supermartingale entails boundedness of every other c.e. supermartingale on that sequence. Conversely, if there exists any c.e. supermartingale that succeeds on a sequence, then all optimal c.e. supermartingales must also succeed on that sequence. We already know that Martin-Löf randomness guarantees boundedness. Given Fact 4.3.1 and Theorem 4.3.3, we can now restate Martin-Löf randomness in the following way:

Theorem 4.3.4. ***(Levin [1])*** *Let M be an optimal c.e. semimeasure and μ a computable measure. A sequence α is μ-Martin-Löf random if and only if there exists a c such that for all prefix x of α, the ratio $M(x)/\mu(x) \leq c$.*

Since there is a formulation of μ-Martin-Löf random sequences in the form of Theorem 4.3.4, it could reasonably be conjectured Martin-Löf randomness many be enough to induce convergence in the Solomonoff sense. As we will see in the next section, this is not the case.

[4] A slight variation of this fact is proved in [34].

5 Non-convergence in the Martin-Löf Sense

5.1 A Negative Result

Since Martin-Löf randomness is widely regarded as the paradigmatic notion of randomness, it is an obvious starting point in investigating which randomness notions suffice for the convergence result in Solomonoff's inductive program. For any given measure μ, we know that the set of all μ-Martin-Löf random sequences has μ-measure one. Further, Theorem 2.2.3 guarantees that the failure of convergence could at most happen for sets that have μ-measure zero. We are interested in asking the following question: is the former a subset of the set on which Theorem 2.2.3 holds? In light of this question, the following is a notable result:

Theorem 5.1.1. ***(Hutter and Muchnik [13])****: There exists an optimal c.e. semimeasure M and a sequence $\beta \in \{0,1\}^\infty$ such that β is λ-Martin-Löf random and*

$$M(\beta_n|\beta_{1:n-1}) \not\to \lambda(\beta_n|\beta_{1:n-1})$$

Theorem 5.1.1 was later strengthened by Hutter and Lattimore [18]; the negative result is valid for all optimal c.e. Bayesian mixtures of the semimeasures. The sequence β can be inductively constructed in the following way: $\beta_n = 0$ if $M(\beta_{1:n-1}0) < 2^{-n}$, and $\beta_n = 1$ otherwise. Hutter and Muchnik show that for all n, $M(\beta_{1:n-1}1) \leq 2^{-n}$, and since $\lambda(\beta_{1:n}) = 2^{-n}$, it follows that β is λ-Martin-Löf random by Theorem 4.3.4. Since M is a c.e. semimeasure, we know that a monotonically increasing rational-valued function g_M approximates M. Let $M^t(\sigma) = g_M(\sigma, t)$. Now, we can inductively define the approximations β^t from M^t as follows: $\beta^t_n = 0$ if $M^t(\beta_{1:n-1}0) < 2^{-n}$, and $\beta^t_n = 1$ otherwise. Since $M^t \nearrow M$, $\beta^t \nearrow \beta$. Observe that, by construction, whenever $\beta^t_n = 1$ and $\beta^{t+1}_n = 0$, it is the case that β^{t+1}_n switches its value from 0 to 1 for some $m < n$. Sequences with this property are known as almost c.e. sequences.

The rest of the proof of Theorem 5.1.1 can be readily modified to show that the theorem can be extended to all almost c.e. Martin-Löf random sequences.

Theorem 5.1.2. *Let M be an optimal c.e. semimeasure M and β a almost c.e. Martin-Löf random sequence. There exists an optimal c.e. semimeasure M' such that*

$$M'(\beta_n|\beta_{1:n-1}) \not\to \lambda(\beta_n|\beta_{1:n-1})$$

Proof. Since β is almost c.e, we have lexicographically increasing approximations β^t such that $\beta = \lim_t \beta^t$.

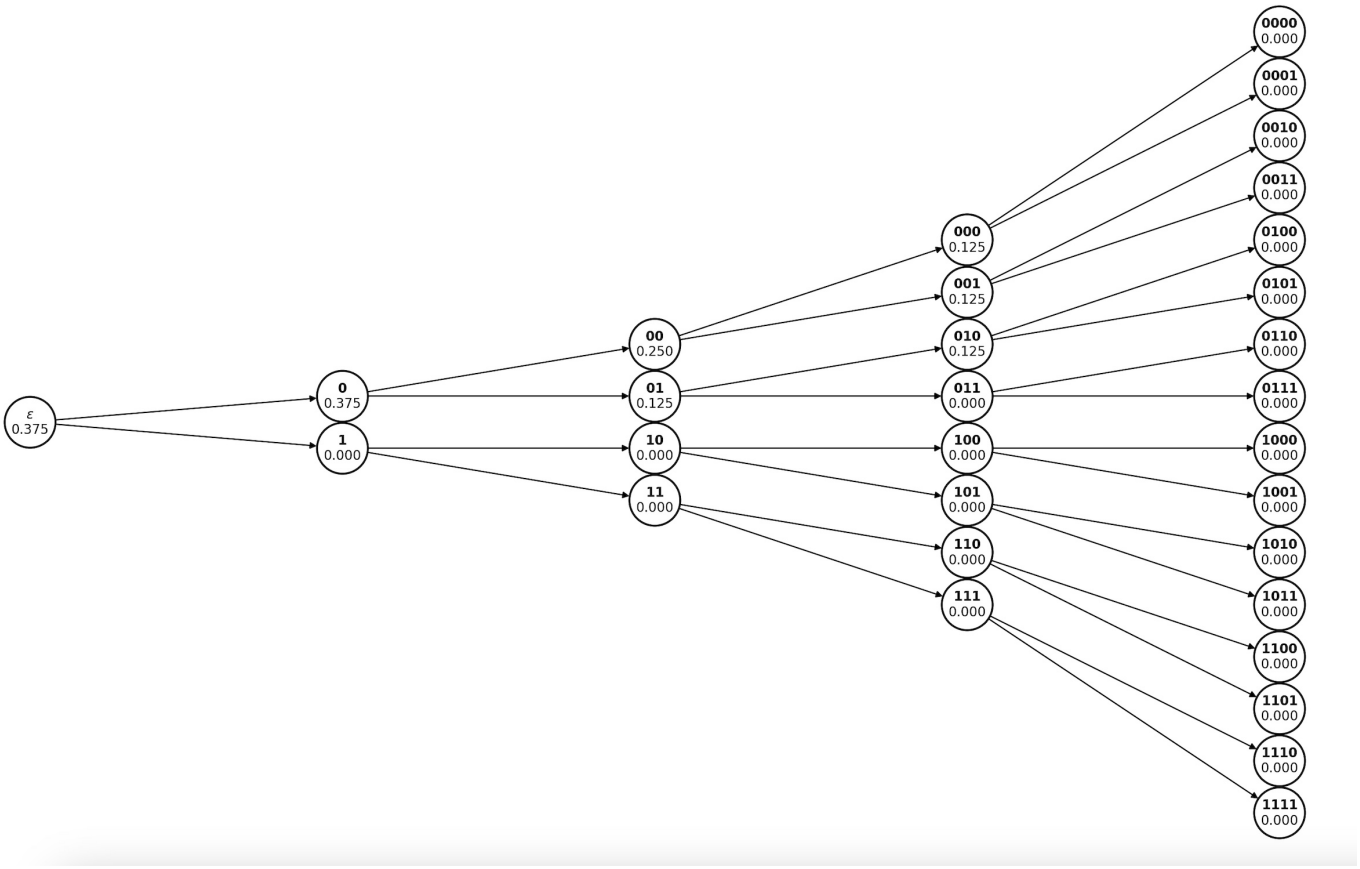

Figure 2: ν^t when $t = 3$ and $\beta^t = 011...$

We define a function ν as follows:

$$\nu^t(\sigma) := \begin{cases} 2^{-t} & \text{if } |\sigma| = t \text{ and } \sigma < \beta^t_{1:t} \\ \nu^t(\sigma 0) + \nu^t(\sigma 1) & \text{if } |\sigma| < t \\ 0 & \text{otherwise} \end{cases}$$

where $<$ is the lexicographical ordering on sequences. Figure 2 illustrates an example of the values ν^t takes for a given t and β^t. Given t and a sequence β^t, the recursive clause ensures that for all σ with $|\sigma| < t$, ν^t is additive. For $|\sigma| > t$, the descendant nodes are assigned the value of 0. ν^t satisfies condition (ii) in Definition 2.2.1. It remains to verify that the value assigned to the empty string is less than or equal to 1. By definition, $\nu^t(\epsilon)$ is the sum of 2^{-t} over all strings of length t that are lexicographically less than $\beta^t_{1:t}$. The number of such strings is at most 2^t, and each contributes 2^{-t}, so we have $\nu^t(\epsilon) \leq 1$. Therefore, ν^t is a semimeasure. Since β^t is computable, so is ν^t. Moreover, because β^t is lexicographically increasing in t, ν^t is monotonically increasing in t, and hence $\nu := \lim_{t\to\infty} \nu^t$ is a c.e. semimeasure.

Inductively assume that $\nu^t(\sigma 0) = 2^{-|\sigma|-1}$ and $\nu^t(\sigma 0) = 2^{-|\sigma|-1}$, where $|\sigma| < t$ and $\sigma x < \beta^t_{1:|\sigma|+1}$ for $x \in \{0,1\}$. Then, $\sigma < \beta^t_{1:|\sigma|}$ and by definition, $\nu^t(\sigma) = \nu^t(\sigma 0) + \nu^t(\sigma 1) = 2^{-|\sigma|}$. Thus, for any σ such that $\sigma < \beta^t_{1:|\sigma|}$ and $|\sigma| \leq t$, we have $\nu^t(\sigma) = 2^{-|\sigma|}$. Similarly, it is straightforward to check that for all σ such that σ is lexicographically after $\beta^t_{1:|\sigma|}$, $\nu^t(\sigma) = 0$.

Since sequence β is a λ-Martin-Löf random sequence, 01 occurs in the sequence infinitely many times. In other words, it is infinitely often the case that $\beta_n = 0$ and $\beta_{n+1} = 1$. We fix one such n. For $t \geq n$, we have $\nu^t(\beta_{1:n-1}) = \nu^t(\beta_{1:n-1}0) + \nu^t(\beta_{1:n-1}1)$ by definition. Since $\beta_n = 0$ and β is lexicographically increasing, $\beta_{1:n-1}1 > \beta_{1:n} \geq \beta^t_{1:n}$. We get $\nu^t(\beta_{1:n-1}1) = 0$ from the definition of ν^t, implying $\nu^t(\beta_{1:n-1}) = \nu^t(\beta_{1:n})$. It follows that $\nu(\beta_{1:n-1}) = \nu(\beta_{1:n})$ infinitely often.

Keeping the n fixed, take a sufficiently large t such that $\beta^t_{1:n+1} = \beta_{1:n+1}$. Obviously, then $\nu^t(\beta_{1:n}) = \nu^t(\beta^t_{1:n})$. Then, since $\beta^t_{1:n+1} > \beta^t_{1:n}0$, $\nu^t(\beta^t_{1:n}0) = 2^{-(n+1)}$. Using the superadditivity of semimeasures, $\nu^t(\beta^t_{1:n}) \geq \nu^t(\beta^t_{1:n}0)$. This implies that $\nu^t(\beta_{1:n}) \geq 2^{-(n+1)}$. ν^t is monotonically increasing in t, so $\nu(\beta_{1:n}) \geq 2^{-(n+1)}$.

Since M is an optimal c.e. semimeasure and β is λ-Martin-Löf random, we know that $\dfrac{M(\sigma)}{\lambda(\sigma)}$ is bounded along β. Then, there exists a $c \geq 1$ such that $\dfrac{M(\beta_{1:n})}{\lambda(\beta_{1:n})} \leq c$ for all n. Pick a γ such that $0 < \gamma < \dfrac{1}{1+4c}$. Define $M'(x) := (1-\gamma)\nu(x) + \gamma M(x)$ for all x. M' is also an optimal semimeasure as it multiplicatively dominates all c.e. semimeasures. Since $M(\beta_{1:n-1}) \leq 2^{-n+1}.c$ and $M(\beta_{1:n}) > 0$, we get

$$M'(\beta_n \mid \beta_{1:n-1}) = \frac{(1-\gamma)\nu(\beta_{1:n}) + \gamma M(\beta_{1:n})}{(1-\gamma)\nu(\beta_{1:n-1}) + \gamma M(\beta_{1:n-1})} \geq \frac{(1-\gamma)\nu(\beta_{1:n})}{(1-\gamma)\nu(\beta_{1:n-1}) + \gamma c 2^{-n+1}}$$

Just as before, choose an n such that $\beta_n = 0$ and $\beta_{n+1} = 1$. We know that $\nu(\beta_{1:n}) \geq 2^{-(n+1)}$ and $\nu(\beta_{1:n-1}) = \nu(\beta_{1:n})$. Therefore, we have

$$M'(\beta_n \mid \beta_{1:n-1}) \geq \frac{(1-\gamma)\nu(\beta_{1:n})}{(1-\gamma)\nu(\beta_{1:n}) + \gamma c 2^{-n+1}} \geq \frac{(1-\gamma)2^{-n-1}}{(1-\gamma)2^{-n-1} + 4\gamma c} \geq \frac{1-\gamma}{1-\gamma+4\gamma c}$$

$$\frac{1}{1+4c} > \gamma \implies \frac{1-\gamma}{1-\gamma+4\gamma c} > \frac{1}{2}$$

We already know that $\lambda(\beta_n \mid \beta_{1:n-1}) = \dfrac{1}{2}$ for all n. Since 01 occurs infinitely many times along β, $M'(\beta_n \mid \beta_{1:n-1}) \neq \lambda(\beta_n \mid \beta_{1:n-1})$ as many times. $\square$

5.2 Understanding the Failure of Convergence

This section futher examines why it is not possible to substitute "μ-almost surely" with "μ-Martin Löf random" in Theorem 2.2.3. The following proof is an adaptation of Miyabe's construction in [4] in our setting of Cantor Space. Let M be an optimal c.e. semimeasure, and μ be a computable probability measure. Given some initial segment $\beta_{1:n-1}$ of $\beta \in \{0,1\}^\infty$ in the support of μ, the Hellinger distance between the two measures is given by the following equation:

$$h_n = \sum_{x\in\{0,1\}} (\sqrt{M(x|\beta_{1:n-1})} - \sqrt{\mu(x|\beta_{1:n-1})})^2 \tag{2}$$

$$\text{Let } N_n = \sum_{x\in\{0,1\}} \sqrt{M(x|\beta_{1:n-1})\mu(x|\beta_{1:n-1})}$$

$$\text{Then, } h_n = [\sum_{x\in\{0,1\}} (M(x|\beta_{1:n-1}) + \mu(x|\beta_{1:n-1}))] - 2N_n$$

Since $\sum_{x\in\{0,1\}} M(x|\beta_{1:n-1}) \leq 1$ and $\sum_{x\in\{0,1\}} \mu(x|\beta_{1:n-1}) \leq 1$, $h_n \leq 2 - 2N_n$. Since $(1-x) \leq e^{-x}$, we have $2 - 2(e^{\frac{-h_n}{2}}) \leq h_n$. We get the following inequality: $2 - 2(e^{\frac{-h_n}{2}}) \leq 2 - 2N_n$. Thus, $1 \geq N_n e^{\frac{h_n}{2}}$.

Define function $S : \{0,1\}^* \to \mathbb{R}^{\geq 0}$ in the following way:

$$S(\sigma) = \sqrt{\frac{M(\sigma)}{\mu(\sigma)}} e^{\sum_{i=1}^{|\sigma|} \frac{h_i}{2}} \tag{3}$$

$$\frac{\sum_{x\in\{0,1\}} S(\sigma x)\mu(\sigma x)}{S(\sigma)\mu(\sigma)} = \sum_{x\in\{0,1\}} \sqrt{\frac{M(\sigma x)}{M(\sigma)} \cdot \frac{\mu(\sigma x)}{\mu(\sigma)}} e^{\frac{h_{|\sigma|+1}}{2}}$$

$$\frac{\sum_{x\in\{0,1\}} S(\sigma x)\mu(\sigma x)}{S(\sigma)\mu(\sigma)} = N_{|\sigma|+1} e^{\frac{h_{|\sigma|+1}}{2}} \leq 1$$

Therefore, S is a μ-supermartingale.

Note: Similar to Definition 3.3.1, when S is undefined, $\mu(\sigma) = 0$. But $S(\sigma)\mu(\sigma)$ is taken to be 0.

Proposition 5.2.1. *If lim sup$_n S(\beta_{1:n}) < \infty$, then lim$_n h_n = 0$.*

Proof. Let lim sup$_n S(\beta_{1:n}) < \infty$. Suppose, towards contradiction, lim$_n \sum_{i\leq n} h_i = \infty$. It follows that lim$_n e^{\sum_{i\leq n} \frac{h_i}{2}} = \infty$. Given the definition of S, sup$_n S(\beta_{1:n}) < \infty$ and lim$_n e^{\sum_{i\leq n} \frac{h_i}{2}} = \infty$ imply that $\sqrt{\frac{M(\beta_{1:n})}{\mu(\beta_{1:n})}} \longrightarrow 0$ as n tends to ∞. Since every computable measure is a c.e. semimeasure, M dominates μ, which means that there exists a $c > 0$ such that for all binary strings σ, $M(\sigma) \geq c\mu(\sigma)$.

Then, $\sqrt{\frac{M(\beta_{1:n})}{\mu(\beta_{1:n})}} \geq \sqrt{\frac{c\mu(\beta_{1:n})}{\mu(\beta_{1:n})}}$. $\sqrt{\frac{M(\beta_{1:n})}{\mu(\beta_{1:n})}} \longrightarrow 0$, but $\sqrt{\frac{c\mu(\beta_{1:n})}{\mu(\beta_{1:n})}} \not\longrightarrow 0$ as n goes to infinity. We arrive at a contradiction. As a consequence, $\lim_n \sum_{i \leq n} h_i = \infty$, implying that $\lim_n h_n = 0$. □

Suppose β is μ-Martin-Löf random. By Theorem 3.3.2, if S is a c.e. μ-supermartingale, then $\limsup_n S(\beta_{1:n}) < \infty$. By Proposition 5.2.1, it would follow that $\lim_n h_n(\beta_{1:n}) = 0$. In effect, if S is a c.e. μ-supermartingale, it would guarantee $M(\beta_{1:n}|\beta) \longrightarrow \mu(\beta_{1:n}|\beta)$ for any arbitrary sequence β that is μ-Martin-Löf random. By Theorem 2.2.1, we know that this is not the case, hence S is not necessarily a c.e. function.

The state of art on characterizing Solomonoff's convergence theorem using randomness is as follows. We have a negative result for Martin-Löf randomness. As a consequence of Proposition 5.2.1, Miyabe suggests, in [23], that stronger notions of randomness such as 2-randomness and difference randomness may be sufficient. In the next section, we will examine some stronger notions of randomness for which the theorem could potentially hold.

6 Searching for the Weakest Notion for Convergence

6.1 2-Computable Randomness

As mentioned earlier, randomness is not a singular notion, and the flexibility of effectivization allows us to define an infinite number of randomness notions. Since Martin-Löf randomness does not admit a characterization in terms of almost sure convergence in the Solomonoff case, in searching through stronger notions we are asking: what is the weakest notion of randomness for which the Solomonoff convergence result holds?

In this section, we use the construction in the previous subsection to prove that 2-computable randomness guarantees Solomonoff convergence in terms of the Hellinger distance. First, we show that:

Proposition 6.1.1. *S in (3) is a limit computable μ-supermartingale.*

Proof. Whenever the μ-supermartingale $S(\sigma)$ is defined, we know that $M(\sigma) > 0$ and $\mu(\sigma) > 0$. Since M is a c.e. semimeasure, there exists a monotonically increasing rational-valued approximating function $g_M : \{0,1\}^* \times \mathbb{N} \to \mathbb{Q}$ such that $\lim_y g_M(\sigma, y) = M(\sigma)$ and $g_M(\sigma, y+1) \geq g_M(\sigma, y)$. Also, because μ is computable, there exists a monotonically decreasing rational-valued approximating function $g_\mu : \{0,1\}^* \times \mathbb{N} \to \mathbb{Q}$ such that $\lim_y g_\mu(\sigma, y) = \mu(\sigma)$ and $g_\mu(\sigma, y+1) \leq g_\mu(\sigma, y)$.

Then, there is a rational-valued approximating function $\dfrac{g_M}{g_\mu}$ such that

$$\lim_y \frac{g_M}{g_\mu}(\sigma, y) = \frac{M}{\mu}(\sigma)$$

$$\frac{g_M}{g_\mu}(\sigma, y+1) \geq \frac{g_M}{g_\mu}(\sigma, y)$$

.

Therefore, $\dfrac{M(\sigma)}{\mu(\sigma)}$ is a c.e. function. Along the same line, since $q \to \sqrt{q}$ is monotone increasing and continuous in [0, ∞], it follows that $\sqrt{\dfrac{M(\sigma)}{\mu(\sigma)}}$ is also a c.e. function.

However, for $1 \leq i \leq |\sigma|$, the Hellinger distance h_i is generally not a c.e. function. Complex c.e. semimeasures such as Solomonoff prior are not closed under conditioning [19]; $M(x|\sigma)$ where $x \in \{0,1\}$ is only limit computable. Because the class of limit computable functions is closed under addition, $e^{\sum_{i=1}^{|\sigma|} \frac{h_i}{2}}$ is a limit computable function. Thus, the output of S must also be limit computable, uniformly in $\sigma \in \{0,1\}^*$. □

Although the hierarchy of computable randomness has not received much attention in the literature, it is natural to define μ-2-computable randomness by relativizing the notion of computable randomness to the halting set:

Definition 6.1.2. *Given a computable measure μ, a sequence α is μ-2-computably random if no martingale that is computable from $\varnothing'$ succeeds on it.*

Proposition 6.1.3. *Given an optimal c.e. semimeasure M and a computable probability measure μ, the following holds for all $\alpha \in \{0,1\}^\infty$ such that α is μ-2-computably random:*

$$\lim_n \sum_{b \in \{0,1\}} (M(b|\alpha_{1:n-1}) - \mu(b|\alpha_{1:n-1}))^2 = 0$$

Proof. Consider the construction given for S in (3). From Proposition 6.1.1, we know that S is a limit computable μ-supermartingale. By Post's theorem, limit computable reals are computable from $\varnothing'$, hence S is a supermartingale that is computable from $\varnothing'$. By Definition 6.1.2, no martingale that is computable from $\varnothing'$ succeeds on α. Intuitively, because supermartingales are obtained by allowing bettors to discard a portion of their capital, no supermartingale should be able to succeed on α either. Formally, we can always scale the supermartingale S with an appropriate scaling factor to obtain a martingale Q that succeeds on α, and since limit computable reals are closed under arithmetic operations Leike and Hutter [19], the resulting martingale is also limit computable. We proceed as follows. Let:

$$\begin{aligned} Q(\epsilon) &= S(\epsilon) \\ Q(\sigma 0) &= S(\sigma 0) \\ Q(\sigma 1) &= 2S(\sigma) - S(\sigma 0) \end{aligned}$$

It is easy to verify that Q dominates S in the sense that $Q(\sigma) \geq S(\sigma)$ for all σ. Q is also limit computable. Suppose, towards contradiction, that $\limsup_n S(\alpha_{1:n}) = \infty$. Then, $\limsup_n Q(\alpha_{1:n}) = \infty$, implying that α is not μ-2-computably random. Therefore, $\limsup_n S(\alpha_{1:n}) < \infty$. The desired result follows immediately from Proposition 5.2.1. □

Since the construction in (3) is based on the Hellinger distance, we established that the stronger version of Solomonoff's convergence theorem holds for 2-computable randomness: the predictive probabilities of the semimeasure converge off-sequence to the true conditional probabilities for all 2-computably random sequences. In the sections that follow, we turn our attention to the weaker version of Solomonoff's theorem — concerning on-sequence convergence — and investigate whether it holds under randomness notions that are strictly weaker than 2-computable randomness.

6.2 Martingale Convergence

Given μ as the true computable probability measure and a sequence α sampled from μ, for on-sequence convergence, we can take for granted that $\mu(\alpha'_n|\alpha_{1:n-1}) \neq 0$ for $\alpha'_n \in \{0,1\}$. In other words, we are only looking at predictions along α. The following fact will be useful in discussing weaker notions of randomness for which convergence in terms of ratios can be realized.

Fact 6.2.1. *Let M be an optimal c.e. semimeasure, μ a computable measure, and α a μ-Martin-Löf random sequence. If the supermartingale defined as $R(\sigma) = \dfrac{M(\sigma)}{\mu(\sigma)}$ converges along α, then $\dfrac{M(\alpha_{1:n}|\alpha_{1:n-1})}{\mu(\alpha_{1:n}|\alpha_{1:n-1})}$ converges to 1.*

Proof. First, we argue that R is a bounded supermartingale along α. Theorem 4.3.4 ensures that R is bounded from above along α. We conclude that $0 < l < R(\alpha_{1:n}) < u < \infty$ for all n. If R converges, we know that $\dfrac{R(\alpha_{1:n})}{R(\alpha_{1:n-1})}$ converges, implying that $\dfrac{M(\alpha_{1:n}|\alpha_{1:n-1})}{\mu(\alpha_{1:n}|\alpha_{1:n-1})}$ also converges. Suppose $\dfrac{M(\alpha_{1:n}|\alpha_{1:n-1})}{\mu(\alpha_{1:n}|\alpha_{1:n-1})}$ converges to η. If $\eta > 1$, then there exists a $c \in (0,1)$ for which $\exists N \forall n \dfrac{R(\alpha_{1:n})}{R(\alpha_{1:n-1})} > 1 + c$, implying that $\lim_n R(\alpha_{1:n}) = \infty$. Similarly, if $\eta < 1$, then there exists a $c \in (0,1)$ for which $\exists N \forall n \dfrac{R(\alpha_{1:n})}{R(\alpha_{1:n-1})} < 1 - c$, implying that R tends to 0 along α. In both the cases, as we have already argued, we reach a contradiction. Therefore, η must be 1 if R converges along α. □

The above fact about bounded supermartingales glaringly hints at the idea that a notion of randomness that suffices for converges of all supermartingales would be sufficient for convergence in the Solomonoff sense. Doob's martingale convergence theorem naturally motivates asking for such a notion. Recall that the theorem says that all μ-martingales (or μ-supermartingales) converge on almost every in $\{0,1\}^\infty$.

Definition 6.2.2. *Let μ be a computable measure. A sequence α is μ-SD random if and only if all c.e. μ-supermartingales converge along α.*

The acronym SD in the above definition stands for "strong density". The naming of this new notion of randomness will become clear in later sections. Using Fact 6.2.1, it is easy to see that Theorem 2.2.3 holds for all SD random sequences. Since every c.e. supermartingale is limit computable, it immediately follows that SD randomness is at least as weak as 2-computable randomness. In the next section, we will argue that SD randomness may be a significantly weaker notion than 2-computable randomness.

6.3 Oberwolfach Randomness

We are seeking to address the question of where SD randomness fits within the randomness hierarchy. To this end, we begin by discussing a notion known as Oberwolfach randomness, which may provide a potential resolution. The test characterization of Oberwolfach randomness closely resembles that of Martin-Löf randomness in that the test components are uniformly Σ^0_1 classes [2]. The crucial distinction, however, lies in that the rate at which the measures of these components converges to zero is no longer computable: instead, the rate is now tracked by a left c.e. real.

Definition 6.3.1. *A μ-left c.e. bounded test is a nested sequence $\{\mathcal{U}_i\}_{i\in\mathbb{N}}$ of Σ^0_1 classes such that $\mu(\mathcal{U}_i) \leq \nu - \nu_i$, where $\{\nu_i\}_{i\in\mathbb{N}}$ are approximations of a left c.e. real ν. A sequence α is μ-Oberwolfach random if $\alpha \notin \bigcap_{i\in\mathbb{N}} \mathcal{U}_i$ for all μ-left c.e. bounded tests.*

Since weak 2-randomness, as defined in 6.1.2, stipulates no modulus of convergence for the rate at which the measures of the components approach zero, Oberwolfach randomness is at least as weak as weak-2 randomness. Bienvenu et al. also provide an equivalent characterization of Oberwolfach randomness in terms of interval tests, relative to the uniform measure.

A rational open ball in $\{0,1\}^\infty$ is a basic clopen subset of the form $[\sigma]$ for some $\sigma \in \{0,1\}^*$, and in $[0,1]$ is an open interval with rational endpoints (including $[0,a)$ and $(b,1]$).

Definition 6.3.2. *An interval array is a computable function G that maps the collection of rational balls in $[0,1]$ to Σ^0_1 classes in $\{0,1\}^\infty$ such that:*

(i) For all I, $\lambda(G(I)) \leq \lambda(I)$

(ii) If $I \subseteq J$ then $G(I) \subseteq G(J)$.

An interval test consists of an interval array G and a left c.e. real $k \in [0,1]$. The set of sequences in $\{0,1\}^\infty$ which are captured by the test (G,k) is

$$\bigcap_{k\in I} G(I).$$

Using interval tests, Bienvenu et al. [Theorem 5.6, [2]] have shown that if α is an Oberwolfach random sequence, then every c.e. martingale converges on α. The proof of this result essentially relies on three facts. To show that Oberwolfach randomness entials SD randomnes, we have to ensure that these three facts hold for c.e. supermartingales. Firstly, if M is a left c.e. martingale then there exists an increasing uniformly computable sequence of rational-valued martingales $\langle M_t\rangle$ which converge pointwise to M. Secondly, if $\mathcal{O}_n(M,a,b)$ denotes the set of sequences in $\{0,1\}^\infty$ such that M oscillates below a and above b at least n times then $\lambda(\mathcal{O}_n(a,b)) \leq (a/b)^n$. This latter

fact is a general property of supermartingales which follows from the upcrossing inequality and so holds even if M is a c.e. supermartingale. Thirdly, given a c such that $a < c < b$, we can bound the probability that sufficiently close computable approximations of our c.e. martingale will switch from oscillating between a and b at least n times to suddenly oscillating between c and b fewer than n times.

We will now show that the first fact also holds if $\mathcal{M}$ is a c.e. supermartingale.

Lemma 6.3.3. *If $\mathcal{M}$ is a c.e. supermartingale, then there exists an increasing uniformly computable sequence of rational-valued supermartingales $\{\mathcal{M}_t\}$ which converge pointwise to $\mathcal{M}$.*

Proof. Since $\mathcal{M}$ is c.e., there exists an approximating rational-valued function M such that $\lim_n M(\sigma, n) = \mathcal{M}(\sigma)$ and $M(\sigma, n) \leq M(\sigma, n+1)$ for all σ.

We now define $\mathcal{M}_0$ as follows: first, set $\mathcal{M}_0(\epsilon) = M(\epsilon, 0)$. Next, given $\mathcal{M}_0(\sigma)$, we define $\mathcal{M}_0(\sigma 0)$ and $\mathcal{M}_0(\sigma 1)$ by checking whether or not $\mathcal{M}_0(\sigma) \geq \frac{M(\sigma 0, 0) + M(\sigma 1, 0)}{2}$. If so, then we set $\mathcal{M}_0(\sigma 0) = M(\sigma 0, 0)$ and $\mathcal{M}_0(\sigma 1) = M(\sigma 1, 0)$, satisfying the supermartingale condition. If not, we find non-negative rationals $m \leq M(\sigma 0, 0)$ and $n \leq M(\sigma 1, 0)$ so that $\mathcal{M}_0(\sigma) \geq \frac{m+n}{2}$. We then set $\mathcal{M}_0(\sigma 0) = m$ and $\mathcal{M}_0(\sigma 1) = n$, satisfying the supermartingale condition.

Further, given $\mathcal{M}_{t-1}$, we define $\mathcal{M}_t \geq \mathcal{M}_{t-1}$ as follows. First, set $\mathcal{M}_t(\epsilon) = M(\epsilon, t)$. Next, given we have defined $\mathcal{M}_t(\sigma) \geq \mathcal{M}_{t-1}(\sigma)$, we then define $\mathcal{M}_t(\sigma 0)$ and $\mathcal{M}_t(\sigma 1)$ by finding the largest natural $k \in [0, t]$ for which $\mathcal{M}_t(\sigma) \geq \frac{M(\sigma 0, k) + M(\sigma 1, k)}{2}$. If such a k exists then set $\mathcal{M}_t(\sigma 0) = M(\sigma 0, k)$ and $\mathcal{M}_t(\sigma 1) = M(\sigma 1, k)$, satisfying the supermartingale condition as well as ensuring that $\mathcal{M}_t(\sigma 0) \geq \mathcal{M}_{t-1}(\sigma 0)$ and $\mathcal{M}_t(\sigma 1) \geq \mathcal{M}_{t-1}(\sigma 1)$ since $\mathcal{M}_t(\sigma) \geq \mathcal{M}_{t-1}(\sigma)$. If no such k exists then set $\mathcal{M}_t(\sigma 0) = \mathcal{M}_0(\sigma 0)$ and $\mathcal{M}_t(\sigma 1) = \mathcal{M}_0(\sigma 1)$, satisfying the supermartingale condition since $\mathcal{M}_t(\sigma) \geq \mathcal{M}_0(\sigma) \geq \frac{\mathcal{M}_0(\sigma 0) + \mathcal{M}_0(\sigma 1)}{2}$ as well as ensuring that $\mathcal{M}_t(\sigma 0) = \mathcal{M}_{t-1}(\sigma 0)$ and $\mathcal{M}_t(\sigma 1) = \mathcal{M}_{t-1}(\sigma 1)$ since $\mathcal{M}_t(\sigma) \geq \mathcal{M}_{t-1}(\sigma)$.

Finally, for each string σ and natural k, there exists a natural t so that $\mathcal{M}_t(\sigma) \geq M(\sigma, k)$ by construction. Hence $\{\mathcal{M}_t\}$ converges pointwise to $\mathcal{M}$. □

Given Lemma 6.3.3 and the upcrossing inequality, we can now attempt to capture the oscillations of a c.e. supermatingale using interval tests. Let $\mathcal{M}$ be a c.e. supermartingale which oscillates on α. Without loss of generality, we can take $\mathcal{M}(\epsilon) \in (0, 1)$ by rescaling and find rationals $a < b$ so that $\liminf_n \mathcal{M}(\alpha_{1:n}) < a$ and $b < \limsup_n \mathcal{M}(\alpha_{1:n})$. There exists an increasing uniformly computable sequence of rational-valued supermartingales $\{\mathcal{M}_t\}$ which converges pointwise to $\mathcal{M}$.

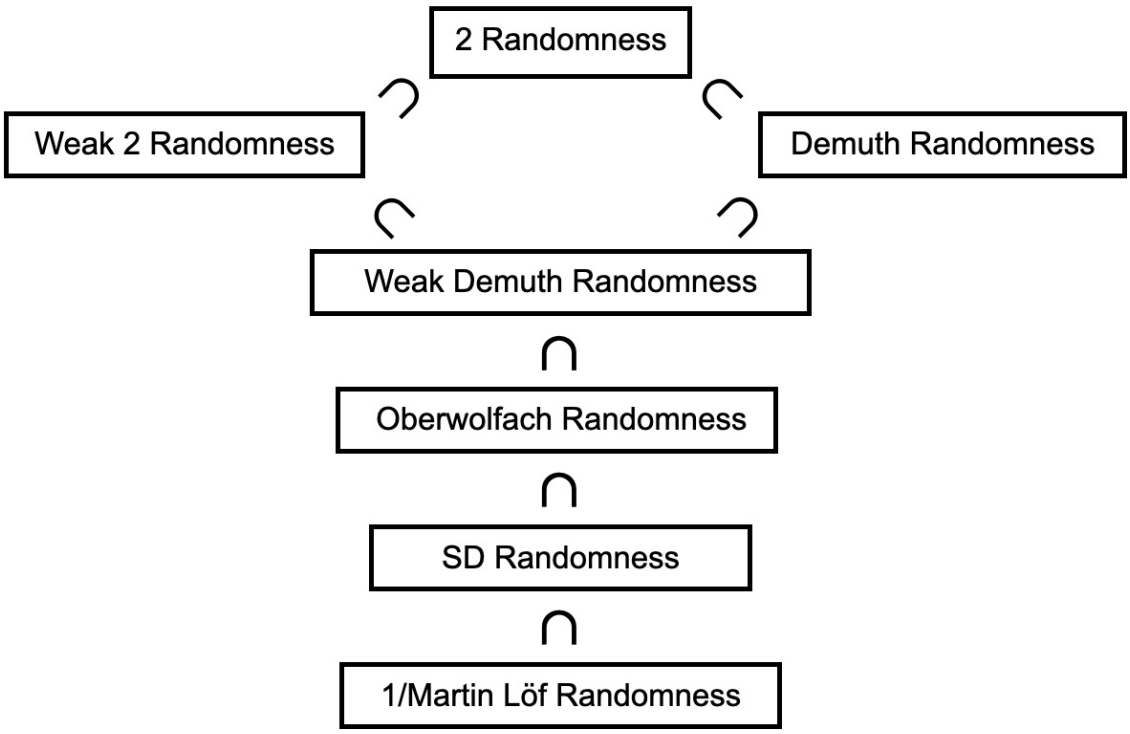

Figure 3: A potential hierarchy?

For each interval $I \subseteq [0,1]$ we can find n_I so that:

$$G(I) = \bigcup_{t:\mathcal{M}_t(\varepsilon)\in I} \mathcal{O}_{n_I}(\mathcal{M}_t, a, b)$$

is an interval array.

However, we can't yet establish the third fact required for extending Theorem 5.6 in [2] to c.e. supermartingales.

It is worth mentioning that Bienvenu et al. have shown that every Demuth test is also an Oberwolfach test. As a consequence, weak Demuth randomness implies Oberwolfach randomness. IIf one could further establish that Oberwolfach randomness implies SD randomness, this would allow us to situate SD randomness in the hierarchy as indicated in Figure 3.

7 Density and Difference Randomness

7.1 Difference Randomness: The Possibility of a Positive Result

In this section, we consider an intermediate notion of randomness that lies strictly between SD randomness and Martin-Löf randomness. Specifically, we motivate why one might think difference randomness suffices for Solomonoff convergence. For any c.e. sequence α, we can always construct a Martin-Löf test in the following way. Take $\mathcal{U}_i = [\alpha_{1:i}]$. Then, $\alpha \in \bigcap_i \mathcal{U}_i$, and hence, no c.e. sequence can be Martin-Löf random. Interestingly, the sequence β, defined in the proof of Theorem 5.1.2—for which convergence fails—is close to being computably enumerable.

β is a lexicographically first λ-Martin-Löf random sequence; its left cut forms a computably enumerable set of finite strings. This implies that $0.\beta$ is a left c.e. real. Kučera and Slaman [15] have shown that every c.e. Martin-Löf random real is the Chaitin's Ω of some universal Turing machine. Chaitin's Ω, or the halting probability, refers to a real number that represents the probability that a universal Turing machine will halt for an arbitrarily constructed program. The realized value of Ω is dependent on the choice of the universal Turing machine. Further, Ω is Turing equivalent to the halting problem. Turing equivalence between X and Y asserts that X can be computed from Y and Y can be computed from X. Subsequently, a Turing machine equipped with an oracle tape containing β can compute the halting problem.

Incompatible with putative intuitions about randomness, Martin-Löf random sequences like Chaitin's constant exhibit remarkable computational power. It has been contended that it is desirable for random sequences to carry little to no useful information when accessed by oracle machines. From this perspective, it is natural to search for a randomness notion that successfully captures the trade-off between permissiveness and computational power. The significance of difference randomness lies in the fact that it identifies the subclass of Martin-Löf random sequences that are computationally too weak to compute the halting problem.

Franklin and Ng [9] define difference random sequences in the following way:

Definition 7.1.1. *Let μ be a computable measure. A μ-d.c.e. test is a sequence $\{\mathcal{U}_i \backslash \mathcal{V}_i\}_{i \in \mathbb{N}}$ where $\mathcal{U}_i$ and $\mathcal{V}_i$ are uniformly Σ^0_1 classes such that $\mu(\{\mathcal{U}_i \backslash \mathcal{V}_i\}) \leq 2^{-i}$ for all i. A sequence β is difference random (with respect to μ) if it passes all μ-d.c.e tests.*

Franklin and Ng have shown that the class of Martin-Löf random sequences can be partitioned into exactly two classes: those that Turing compute $\varnothing'$ and those that don't.

Theorem 7.1.2. ***(Franklin and Ng [9])*** *A sequence is difference random if and only if it is Martin-*

Löf random and does not compute the halting problem.

The theorem above establishes that sequences such as β, which are used to demonstrate the negative result for Martin-Löf randomness, are not classified as random under the definition of difference randomness. This observation may, in part, motivate Miyabe's suggestion [23] that a positive result for difference randomness could be attainable. The following subsections show that this is not the case.

7.2 Dyadic Density and Martingales

In order to prove a negative result for difference randomness, we will first proceed with a discussion on dyadic densities in Cantor space.

Definition 7.2.1. *Let $\mathcal{C} \subseteq \{0,1\}^\infty$ and $\alpha \in \{0,1\}^\infty$. Then, we can define lower and upper dyadic densities — ρ^u and ρ^l, respectively — of $\mathcal{C}$ at α with respect to λ in the following way:*

$$\rho^l(\mathcal{C}|\alpha) = \textit{lim inf}_n \frac{\lambda(\mathcal{C} \cap [\alpha|n])}{\lambda([\alpha|n])}$$

$$\rho^u(\mathcal{C}|\alpha) = \textit{lim sup}_n \frac{\lambda(\mathcal{C} \cap [\alpha|n])}{\lambda([\alpha|n])}$$

In the definition above, if $\lim_n \frac{\lambda(\mathcal{C} \cap [\alpha|n])}{\lambda([\alpha|n])}$ exists, upper and lower densities coincide. The common value is known as dyadic density of $\mathcal{C}$ at α.

Recall that the Lebesgue density theorem states that for any measurable set $\mathcal{C} \subseteq \{0,1\}^\infty$, $\rho^u(\mathcal{C}|\alpha) = \rho^l(\mathcal{C}|\alpha) = 1$ for almost every α in $\mathcal{C}$. As has been the recurring theme throughout this thesis, we ask what level of randomness of a sequence guarantees that the dyadic density at that point is 1. It is easily seen that the dyadic density of measurable sets that are Σ^0_1 in Cantor space is 1 at every point in the set. As a result, the effective instance of the Lebesgue density theorem that we are concerned with equates measurable sets with Π^0_1 classes.

A natural starting point is to examine whether Martin-Löf randomness suffices. It has been shown by Bienvenu et al. [Proposition 5.4, [2]] that for any Π^0_1 class $\mathcal{P}$ and a Martin-Löf random sequence X in $\mathcal{P}$, $\rho^u(\mathcal{P}|X) = 1$. The same cannot be said for lower dyadic densities. As a result, a new notion of randomness emerges out of the project of effectivizing the Lebesgue density theorem. The random sequences on which this effective version of the theorem can be validated are classified under the definition of density randomness.

Definition 7.2.2. ***(Miyabe, Nies, and Zhang [24])*** *A sequence α is λ-density random if and only if it is λ-Martin-Löf random and a dyadic density-one point. To be dyadic density-one point, $\rho^l(\mathcal{P}|\alpha) = 1$ for all Π^0_1 class $\mathcal{P}$ containing α.*

As it turns out, difference randomness can be characterized in terms of dyadic densities.

Theorem 7.2.3. ***(Bienvenu et al. [3])*** *A sequence α is λ-difference random if and only if it is λ-Martin-Löf random and a dyadic positive-density point. To be dyadic positive-density point, $\rho^l(\mathcal{P}|\alpha) > 0$ for all Π^0_1 class $\mathcal{P}$ containing α.*

Clearly, density randomness entails difference randomness. The reverse, however, does not hold. In particular, Day and Miller [6] proved a separation result between density and difference randomness; there exists a difference random sequences that is not density random.

Theorem 7.2.4. ***(Day and Miller [6])*** *There exists a Δ^0_2 sequence z that satisfies the following desiderata:*

1. *z is λ-Martin-Löf random;*
2. *z is not a density-one point with respect to λ;*
3. *z is a positive-density point with respect to λ*

The proof of Theorem 7.2.4 uses a notion of forcing that separates density-one points from positive-density ones on Martin-Löf random sequences. If z is sufficiently generic for this notion of forcing, then z is shown to be in compliance with properties (1), (2), and (3).

Now, we answer the question as to why dyadic densities warrant special interest. There is a natural translation between dyadic densities and martingales. Let $\mathcal{C} \subseteq \{0,1\}^\infty$ be a measurable set. Define a function $\mathcal{M}(\sigma) = \lambda(\mathcal{C} \mid \sigma)$. Observe that:

$$\begin{aligned}
\mathcal{M}(\sigma).\lambda(\sigma) &= \lambda(\mathcal{C}|[\sigma]).\lambda(\sigma) \\
&= \lambda(\mathcal{C} \cap [\sigma]) \\
&= \lambda\left(\mathcal{C} \cap [\sigma 0]\right) + \lambda\left(\mathcal{C} \cap [\sigma 1]\right) \\
&= \lambda(\mathcal{C}|[\sigma 0]).\lambda(\sigma 0) + \lambda(\mathcal{C}|[\sigma 1]).\lambda(\sigma 1) \\
&= \mathcal{M}(\sigma 0).\lambda(\sigma 0) + \mathcal{M}(\sigma 1).\lambda(\sigma 1)
\end{aligned}$$

Thus, $\mathcal{M}$ is a λ-martingale. It follows that the dyadic density of $\mathcal{C}$ at X exists if and only if $\mathcal{M}$ converges along X. Moreover, for any given Π^0_1 class $\mathcal{P}$, the function $\mathcal{M}(\sigma) = \lambda(\mathcal{P} \mid \sigma)$ is a co-c.e. λ-martingale. Subsequently, the function $\sigma \mapsto (1 - \mathcal{M}(\sigma))$ defines a c.e. λ-martingale.

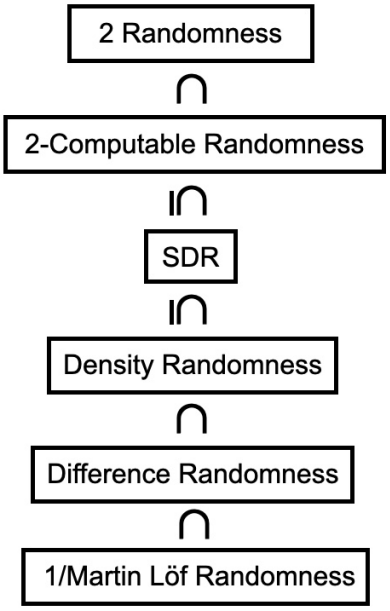

Figure 4: Placing difference and density randomness on the hierarchy

Fact 7.2.5. *There exists a c.e. λ-martingale $\mathcal{M}$ that oscillates on a λ-difference random sequence X.*

Proof. By Theorem 7.2.4, we get a λ-difference random sequence that is not a density-one point with respect to λ. Let's call it X. Every λ-difference random sequence is a λ-Martin-Löf random, and hence, the upper dyadic density at X is always 1 for all Π^0_1 classes to which X belongs. But X is not a density-one point. This implies that there exists a Π^0_1 class $\mathcal{H}$ that contains X such that dyadic density of $\mathcal{H}$ at X does not exist. Let $\mathcal{M}(\sigma) = 1 - \lambda(\mathcal{H}|\sigma)$ be a c.e. λ-martingale. It follows that $\mathcal{M}$ diverges along X.

We now have all the ingredients necessary to prove a negative result.

7.3 A Negative Result for Difference Randomness

Building on the last subsection, we finally provide a negative result for difference randomness; we can find an appropriate optimal c.e. semimeasure D and λ-difference random X such that Solomonoff convergence does not hold. First, we prove the following auxiliary lemma.

Lemma 7.3.1. *Let $\mathcal{M}$ be the bounded c.e. martingale that oscillates on α. There exists a c.e. supermartingale $\mathcal{L}$ that oscillates on consecutive bits of α, infinitely many times.*

Proof. By Lemma 6.3.3, there exists a sequence $\{\mathcal{M}_s\}_{s\in\mathbb{N}}$ that witnesses $\mathcal{M}$. Moreover, if $\lim_n \mathcal{M}(\alpha_{1:n})$ does not exist and $\limsup_n \mathcal{M}(\alpha_{1:n}) < \infty$, we can find rationals $b < a$ such that there are infinitely

many n with $\mathcal{M}(\alpha_{1:n}) \geq a$ and infinitely many n with $\mathcal{M}(\alpha_{1:n}) \leq b$. We want to construct a c.e. supermartingale $\mathcal{L}$ such that $\mathcal{L}(\alpha_{1:n-1}) \geq a$ and $\mathcal{L}(\alpha_{1:n}) \leq b$ for infinitely many n.

We construct a witnessing sequence $\{\mathcal{L}_s\}_{s\in\mathbb{N}}$ as follows:

$$\mathcal{L}_s(\epsilon) = \begin{cases} \mathcal{M}_s(\epsilon) & \text{if } \mathcal{M}_s(\epsilon) < a \\ a & \text{otherwise} \end{cases}$$

For all σ such that $|\sigma| \geq 1$,

$$\mathcal{L}_s(\sigma) = \begin{cases} \mathcal{M}_s(\sigma) & \text{if } \mathcal{M}_s(\sigma) \leq b \\ \mathcal{M}_s(\sigma) & \text{if } b < \mathcal{M}_s(\sigma) < a \text{ and } \mathcal{L}_s(\sigma^-) < a \\ a & \text{otherwise} \end{cases}$$

To show that $\mathcal{L}_s$ is monotonically increasing in s, we proceed by induction along an arbitrary sequence α. By construction and using the fact that $\mathcal{M}_s(\sigma) \leq \mathcal{M}_{s+1}(\sigma)$, we are guaranteed that $\mathcal{L}_s(\epsilon) \leq \mathcal{L}_{s+1}(\epsilon)$ for all s. Inductively, assuming that $\mathcal{L}_s(\alpha_{1:n}) \leq \mathcal{L}_{s+1}(\alpha_{1:n})$ for all s, we have the following cases:

- If $\mathcal{M}_s(\alpha_{1:n+1}) \geq a$, then $\mathcal{M}_{s+1}(\alpha_{1:n+1}) \geq a$. This implies that $\mathcal{L}_s(\alpha_{1:n+1}) = a = \mathcal{L}_{s+1}(\alpha_{1:n+1})$ so we are done.
- Otherwise, if $\mathcal{M}_s(\alpha_{1:n+1}) < a$ and $\mathcal{L}_s(\alpha_{1:n+1}) = \mathcal{M}_s(\alpha_{1:n+1})$, then since $\mathcal{M}_s(\alpha_{1:n+1}) \leq \mathcal{M}_{s+1}(\alpha_{1:n+1})$ and we know $\mathcal{L}_{s+1}(\alpha_{1:n+1}) = a$ or $\mathcal{L}(\alpha_{1:n+1})_{s+1} = \mathcal{M}_{s+1}(\alpha_{1:n+1})$, it must be the case that $\mathcal{L}_s(\alpha_{1:n+1}) \leq \mathcal{L}_{s+1}(\alpha_{1:n+1})$.
- Finally, if $\mathcal{M}_s(\alpha_{1:n+1}) < a$ and $\mathcal{L}_s(\alpha_{1:n+1}) \neq \mathcal{M}_s(\alpha_{1:n+1})$, then it must be the case that $b < \mathcal{M}_s(\alpha_{1:n}) < a$ and $\mathcal{L}_s(\alpha_{1:n}) \geq a$. Namely, since $\mathcal{M}_s(\alpha_{1:n+1}) \leq \mathcal{M}_{s+1}(\alpha_{1:n+1})$ and $\mathcal{L}_s(\alpha_{1:n}) \leq \mathcal{L}_{s+1}(\alpha_{1:n})$, it follows that $b < \mathcal{M}_{s+1}(\alpha_{1:n+1})$ and $\mathcal{L}_{s+1}(\alpha_{1:n}) \geq a$. Thus $\mathcal{L}_{s+1}(\alpha_{1:n+1}) = a$ and so $\mathcal{L}_s(\alpha_{1:n+1}) \leq \mathcal{L}_{s+1}(\alpha_{1:n+1})$.

Since α was arbitrary, we conclude that $\mathcal{L}_s(\sigma) \leq \mathcal{L}_{s+1}(\sigma)$ for all s and σ. Now, we show that each $\mathcal{L}_s$ is a supermartingale. There are two cases to consider:

- If $\mathcal{L}_s(\sigma) = a$, then by the above construction of $\mathcal{L}_s$, we are guaranteed that $\mathcal{L}_s(\sigma 0) \leq a$ and $\mathcal{L}_s(\sigma 1) \leq a$. Thus, the supermartingale property is trivially satisfied.
- If $\mathcal{L}_s(\sigma) < a$, then it must be the case that $\mathcal{M}_s(\sigma) = \mathcal{L}_s(\sigma)$. Since $\mathcal{M}_s$ satisfies the supermartingale property, we know that $\mathcal{M}_s(\sigma 0) + \mathcal{M}_s(\sigma 1) < 2a$. If both $\mathcal{M}_s(\sigma 0) < a$ and

$\mathcal{M}_s(\sigma 1) < a$, then $\mathcal{L}_s(\sigma x) = \mathcal{M}_s(\sigma x)$ for $x \in \{0,1\}$. In this case, $\mathcal{L}_s(\sigma 0) + \mathcal{L}_s(\sigma 1) = 2\mathcal{L}_s(\sigma)$. On the other hand, if $\mathcal{M}_s(\sigma x) \geq a$, then $\mathcal{M}_s(\sigma(1-x)) < a$. Here, $\mathcal{L}_s(\sigma x) = a$ and $\mathcal{L}_s(\sigma(1-x)) = \mathcal{M}_s(\sigma(1-x)) < a$ by construction. It follows that $\mathcal{L}_s(\sigma x) + \mathcal{L}_s(\sigma(1-x)) \leq 2\mathcal{M}_s(\sigma) = 2\mathcal{L}_s(\sigma)$.

For all σ, let $\mathcal{L}(\sigma) = \lim_s \mathcal{L}_s(\sigma)$. Since $\{\mathcal{L}_s\}_{s\in\mathbb{N}}$ is uniformly computable in s, $\mathcal{L}$ is left c.e. uniformly in σ. Furthermore, since each $\mathcal{L}_s$ is a supermartingale, $\mathcal{L}$ is also a supermartingale.

Along α, whenever $\mathcal{M}_s$ starts betting a or more, $\mathcal{L}_s$ bets a until we find an n such that $\mathcal{M}_s(\alpha_{1:n})$ is below or at b, after which, $\mathcal{L}_s$ copies $\mathcal{M}_s$'s bets. In other words, if we can find k and n such that $\mathcal{M}_s(\alpha_{1:k}) \geq a$ and $\mathcal{M}_s(\alpha_{1:l}) \leq b$, then $\mathcal{L}_s(\alpha_{1:n-1}) = a$ and $\mathcal{L}_s(\alpha_{1:n}) \leq b$. Because $\mathcal{M}$ cuts across the interval $[a,b]$ infinitely many times along α, $\mathcal{L}$ oscillates on consecutive bits as many times. □

Theorem 7.3.2. *There exists a sequence $X \in \{0,1\}^\infty$ and an optimal c.e. semimeasure D such that X is λ-difference random and*

$$D(X_n|X_{1:n-1}) \not\to \lambda(X_n|X_{1:n-1})$$

Proof. Let M be an optimal c.e. semimeasure. By Fact 7.2.5 and Lemma 7.3.1, we know that there exists a c.e. λ-supermartingale r and a λ-difference random sequence X such that $r(X_{1:n-1}) \geq a$ and $r(X_{1:n}) \leq b$ for infinitely many n for some a, b with $0 < b < a$. Let $R(\sigma) = \dfrac{M(\sigma)}{\lambda(\sigma)}$. Since every difference random sequence is also Martin-Löf random, R is an optimal supermartingale that is bounded on X, that is, $0 < l < R(X_{1:n}) < u < \infty$ for all n. Now, if $\dfrac{R(X_{1:n})}{R(X_{1:n-1})} \not\to 1$, we are done. If not, for sufficiently large n_0, we have $|R(X_{1:n}) - R(X_{1:n-1})| < \varepsilon < \gamma(a-b)$ for all $n \geq n_0$ and some sufficiently small $\varepsilon > 0$. The goal is to contaminate the values of R with the values of r along X. Let M' be a c.e. semimeasure that corresponds to r in the sense that $r(\sigma) = \dfrac{M'(\sigma)}{\lambda(\sigma)}$. We define a new λ-supermartingale $R'(\sigma) = \dfrac{R(\sigma) + \gamma r(\sigma)}{1+\gamma}$[5]. Then, for (the infinitely many) $n \geq n_0$ for which $r(X_{1:n-1}) = a$ and $r(X_{1:n}) \leq b$, we have

$$1 - \frac{R'(X_{1:n})}{R'(X_{1:n-1})} = \frac{R(X_{1:n-1}) - R(X_{1:n}) + \gamma r(X_{1:n-1}) - \gamma r(X_{1:n})}{R(X_{1:n-1}) + \gamma r(X_{1:n-1})} \geq \frac{-\varepsilon + \gamma(a-b)}{u+a} > 0$$

This implies that $\dfrac{R'(X_{1:n})}{R'(X_{1:n-1})} \not\to 1$. Let $D = \dfrac{M + \gamma M'}{1+\gamma}$. Since R' is an optimal λ-supermartingale, D is an optimal semimeasure. It follows that

$$\frac{R'(X_{1:n})}{R'(X_{1:n-1})} \not\to 1$$

[5]Without loss of generality, we can assume $r(\epsilon) \leq 1$.

$$\frac{(R(X_{1:n}) + \gamma r(X_{1:n}))/(1+\gamma)}{(R(X_{1:n-1}) + \gamma r(X_{1:n-1}))/(1+\gamma)} \not\to 1$$

$$\frac{(M(X_{1:n}) + \gamma M'(X_{1:n}))/(1+\gamma)}{(M(X_{1:n-1}) + \gamma M'(X_{1:n-1}))/(1+\gamma)} \cdot \frac{\lambda(X_{1:n-1})}{\lambda(X_{1:n})} \not\to 1$$

$$\frac{D(X_n|X_{1:n-1})}{\lambda(X_n|X_{1:n-1})} \not\to 1$$

We conclude that convergence fails on-sequence, implying an off-sequence failure of convergence.

□

8 Conclusion

8.1 Main Contributions

The success of Solomonoff induction can be made precise in terms of off-sequence convergence and on-sequence convergence, with the former constituting the stronger notion of success. Using Miyabe's construction, we show that 2-computable randomness is sufficient for off-sequence convergence of Solomonoff induction. By contrast, Hutter and Muchnik show that there exists an optimal c.e. semimeasure and a λ-Martin-Löf random sequence α such that Solomonoff induction fails to achieve on-sequence convergence. We strengthen this negative result by showing that such failure occurs for all Martin-Löf random sequences α that are left c.e. reals. Miyabe had previously suggested that this negative phenomenon could be avoided by restricting attention to difference random sequences. However, we demonstrate that there exists an optimal c.e. semimeasure and a λ-difference random sequence α such that Solomonoff induction again fails on-sequence convergence.

Moreover, our proof shows that in order to guarantee on-sequence convergence for all semimeasures and probability measures, one must require a notion of randomness ensuring that no c.e. supermartingale diverges along the sequence. The seemingly weaker requirement that no optimal c.e. supermartingale merely oscillates turns out to be equivalent. So long as there is a c.e. supermartingale which oscillates, one can contaminate any optimal c.e. supermartingale with it to obtain an optimal c.e. supermartingale that diverges. Finally, we argue that this notion of randomness is potentially weaker than well-known notions such as weak-2 randomness and Demuth randomness.

8.2 Open Questions

The problem of isolating a notion of randomness that is necessary and sufficient for the convergence of all c.e. supermartingales has long been elusive in the algorithmic randomness literature. While several candidates, such as density randomness and Oberwolfach randomness, have been proposed, no existing notion has yet been shown to fully capture this convergence behavior. In this thesis, we show that 2-computable randomness suffices for on-sequence convergence of Solomonoff induction. However, we leave open whether Oberwolfach randomness also suffices. Namely, does Oberwolfach randomness entail SD randomness, i.e. does it guarantee the convergence of every c.e. supermartingale?

A second, closely related issue concerns density-based notions. It has been claimed in [24] that density randomness entails convergence of every c.e. supermartingale. This claim is erroneous,[6]

[6]This error has been identified by Simon Hutteger, Sean Walsh, and Francesca Zaffora Blando in a currently unpub-

and motivates distinguishing strong density (SD) randomness from density randomness. In this thesis, SD randomness is defined precisely to capture convergence of all c.e. supermartingales. It remains open whether density randomness nevertheless implies SD randomness. These questions interact with another open problem posed in [24], namely whether density randomness implies Oberwolfach randomness.

Answering both questions in the affirmative would imply equivalence between these notions. More precisely, if density randomness *does* entail Oberwolfach randomness, and if Oberwolfach randomness *also* implies SD randomness, then the three notions collapse: density randomness, Oberwolfach randomness, and SD randomness coincide. This follows immediately from the fact that Oberwolfach randomness suffices to ensure that a sequence is a dyadic density-one point (Theorem 1.3 of [2]), and hence already implies density randomness.

Of course, there is also a more direct route to locating SD randomness within the randomness hierarchy, namely by establishing an explicit separation result. Concretely, such a separation would amount to constructing a sequence on which every c.e. supermartingale converges, while at least one limit computable supermartingale oscillates. A sequence with this property is SD random by definition, yet fails to satisfy 2-computable randomness, since 2-computable randomness precisely prohibits oscillation of limit computable supermartingales. Establishing the existence of such a sequence would, therefore, demonstrate that SD randomness is strictly weaker than 2-computable randomness.

8.3 Future Direction

The preceding discussion concerns the fine structure of randomness notions as they arise from convergence properties of martingales and semimeasures. However, these distinctions between randomness notions also admit an interpretation in terms of the inductive assumptions built into different predictive procedures. As Zaffora Blando argues in [34],

> Algorithmic randomness notions may be taken to encode a specific type of inductive assumptions—or commitments (either explicit or implicit)—that result from the subjective prior with respect to which randomness is defined.

Which inductive assumptions are, then, implicit in Solomonoff's predictors? Sterkenburg [32] has shown that Solomonoff's algorithmic predictors can be characterized as precisely those Bayesian predictors that operate under the inductive assumption of effectiveness—that is, the assumption that the true probability measure is at least a c.e. function. Exploring how randomness notions could

lished manuscript.

encode such an inductive assumption necessitates establishing which sequences are deemed random with respect to a given semimeasure, and how this classification interacts with randomness defined relative to an underlying computable measure. One natural direction for further investigation is to focus on sequences that lie in the intersection of these randomness classes: data that is agreed to be random by both the predictor and the true probability measure. Restricting attention to such sequences may allow one to isolate those negative results concerning Solomonoff induction that stem from mismatches between the predictor's inductive assumptions and the true data-generating process.

Some preliminary progress in this direction has been made by Bienvenu et al. [4], who develop notions of randomness relative to semimeasures. However, many of the randomness notions central to this thesis—such as density randomness and difference randomness—currently lack semimeasure-relative formulations. Extending these notions to the semimeasure setting thus represents a promising avenue for future work, both for refining the theory of algorithmic randomness and for clarifying the inductive commitments underlying algorithmic predictors.

References

[1] LEVIN L. A. "On the notion of a random sequence". In: *Soviet. Math. Dokl.* 14.5 (1973), pp. 1413–1416 (cit. on p. 19).

[2] Laurent Bienvenu et al. "Coherent randomness tests and computing the K-trivial sets". In: *Journal of the European Mathematical Society* 18.4 (Mar. 8, 2016), pp. 773–812. ISSN: 1435-9855, 1435-9863 (cit. on pp. 28, 30, 32, 39).

[3] Laurent Bienvenu et al. *Denjoy, Demuth, and Density*. Feb. 8, 2014. arXiv: 1308.6402[math] (cit. on p. 33).

[4] Laurent Bienvenu et al. "Randomness and Semimeasures". In: *Notre Dame Journal of Formal Logic* 58.3 (Jan. 1, 2017). ISSN: 0029-4527 (cit. on p. 40).

[5] Vasco Brattka, Joseph S. Miller, and Andr Nies. "Randomness and differentiability". In: *Transactions of the American Mathematical Society* 368.1 (2016). Publisher: American Mathematical Society, pp. 581–605. ISSN: 0002-9947 (cit. on p. 18).

[6] Adam R. Day and Joseph S. Miller. *Density, forcing, and the covering problem*. Apr. 9, 2013. arXiv: 1304.2789[math] (cit. on p. 33).

[7] Osvald Demuth. *Some classes of arithmetical real numbers.* 1984 (cit. on p. 14).

[8] Osvald Demuth. "The differentiability of constructive functions of weakly bounded variation on pseudo numbers". In: *Comment. Math. Univ. Carolin* 16.3 (1975), pp. 583–599 (cit. on p. 18).

[9] Johanna N. Y. Franklin and Keng Meng Ng. "Difference randomness". In: *Proceedings of the American Mathematical Society* 139.1 (Jan. 1, 2011), pp. 345–345. ISSN: 0002-9939 (cit. on p. 31).

[10] Peter Gács. "Uniform test of algorithmic randomness over a general space". In: *Theoretical Computer Science* 341.1 (Sept. 2005), pp. 91–137. ISSN: 03043975 (cit. on p. 12).

[11] Mathieu Hoyrup and Jason Rute. "Computable measure theory and algorithmic randomness". In: *Handbook of Computability and Complexity in Analysis*. Springer, 2021, pp. 227–270 (cit. on p. 18).

[12] Marcus Hutter. *On the Existence and Convergence Computable Universal Priors*. May 29, 2003. arXiv: cs/0305052 (cit. on pp. 8, 10, 17).

[13] Marcus Hutter and Andrej Muchnik. "Universal Convergence of Semimeasures on Individual Random Sequences". In: *Algorithmic Learning Theory*. Ed. by Shoham Ben-David, John Case, and Akira Maruoka. Berlin, Heidelberg: Springer, 2004, pp. 234–248. ISBN: 978-3-540-30215-5 (cit. on pp. 2, 4, 20).

[14] Steven M Kautz. “A Dissertation Presented to the Faculty of the Graduate School of Cornell University in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy”. In: () (cit. on p. 14).

[15] Antonín Kucera and T. Slaman. “Randomness and Recursive Enumerability”. In: *SIAM Journal on Computing* 31.1 (Jan. 2001), pp. 199–211. ISSN: 0097-5397, 1095-7111 (cit. on p. 31).

[16] Antonín Kučera and André Nies. “Demuth randomness and computational complexity”. In: *Annals of Pure and Applied Logic* 162.7 (June 2011), pp. 504–513. ISSN: 01680072 (cit. on p. 14).

[17] Stuart Alan Kurtz. *Randomness and Genericity in the Degrees of Unsolvability*. Google-Books-ID: ZHfMHAAACAAJ. University of Illinois at Urbana-Champaign, 1981. 262 pp. (cit. on p. 14).

[18] Tor Lattimore and Marcus Hutter. “On Martin-Löf Convergence of Solomonoff’s Mixture”. In: *Theory and Applications of Models of Computation*. Ed. by T-H. Hubert Chan, Lap Chi Lau, and Luca Trevisan. Red. by David Hutchison et al. Vol. 7876. Series Title: Lecture Notes in Computer Science. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013, pp. 212–223. ISBN: 978-3-642-38235-2 978-3-642-38236-9 (cit. on p. 20).

[19] Jan Leike and Marcus Hutter. “On the computability of Solomonoff induction and AIXI”. In: *Theoretical Computer Science*. Special Issue on ALT 2015 716 (Mar. 15, 2018), pp. 28–49. ISSN: 0304-3975 (cit. on pp. 25, 26).

[20] Ming Li and Paul M.B. Vitányi. “Inductive reasoning and kolmogorov complexity”. In: *Journal of Computer and System Sciences* 44.2 (Apr. 1, 1992), pp. 343–384. ISSN: 0022-0000 (cit. on p. 17).

[21] Ming Li and Paul Vitányi. *An Introduction to Kolmogorov Complexity and Its Applications*. New York, NY: Springer, 1997. ISBN: 978-1-4757-2608-4 978-1-4757-2606-0 (cit. on pp. 8, 10, 17).

[22] Per Martin-Löf. “The definition of random sequences”. In: *Information and Control* 9.6 (Dec. 1, 1966), pp. 602–619. ISSN: 0019-9958 (cit. on p. 12).

[23] Kenshi Miyabe. “An Optimal Superfarthingale and Its Convergence over a Computable Topological Space”. In: *Algorithmic Probability and Friends. Bayesian Prediction and Artificial Intelligence: Papers from the Ray Solomonoff 85th Memorial Conference, Melbourne, VIC, Australia, November 30 – December 2, 2011*. Ed. by David L. Dowe. Berlin, Heidelberg: Springer, 2013, pp. 273–284. ISBN: 978-3-642-44958-1 (cit. on pp. 24, 32).

[24] Kenshi Miyabe, André Nies, and Jing Zhang. “USING ALMOST-EVERYWHERE THEOREMS FROM ANALYSIS TO STUDY RANDOMNESS”. In: *The Bulletin of Symbolic Logic* 22.3 (Sept. 2016), pp. 305–331. ISSN: 1079-8986, 1943-5894 (cit. on pp. 33, 38, 39).

[25] Shay Moran and Amir Yehudayoff. *Sample compression schemes for VC classes*. Apr. 14, 2015. arXiv: 1503.06960[cs] (cit. on p. 4).

[26] Emil L. Post. "Recursively enumerable sets of positive integers and their decision problems". In: *Mathematical Logic in the 20th Century*. CO-PUBLISHED WITH SINGAPORE UNIVERSITY PRESS, Aug. 2003, pp. 352–384. ISBN: 978-981-02-4736-2 (cit. on p. 7).

[27] C. P. Schnorr. "A unified approach to the definition of random sequences". In: *Mathematical systems theory* 5.3 (Sept. 1, 1971), pp. 246–258. ISSN: 1433-0490 (cit. on pp. 16, 19).

[28] C. P. Schnorr. "The process complexity and effective random tests." In: *Proceedings of the fourth annual ACM symposium on Theory of computing - STOC '72*. the fourth annual ACM symposium. Denver, Colorado, United States: ACM Press, 1972, pp. 168–176 (cit. on p. 17).

[29] Kevin Scott. "I Do Not Think It Means What You Think It Means: Artificial Intelligence, Cognitive Work & Scale". In: *Daedalus* 151.2 (May 1, 2022), pp. 75–84. ISSN: 0011-5266 (cit. on p. 4).

[30] R. Solomonoff. "Complexity-based induction systems: Comparisons and convergence theorems". In: *IEEE Transactions on Information Theory* 24.4 (July 1978). Conference Name: IEEE Transactions on Information Theory, pp. 422–432. ISSN: 1557-9654 (cit. on pp. 2, 4, 10).

[31] R. J. Solomonoff. "A formal theory of inductive inference. Part I". In: *Information and Control* 7.1 (Mar. 1, 1964), pp. 1–22. ISSN: 0019-9958 (cit. on p. 17).

[32] Tom F. Sterkenburg. *Solomonoff Prediction and Occam's Razor*. Philosophy of Science. Issue: 4 Number: 4 Pages: 459-479 Publisher: The University of Chicago Press Volume: 83. Oct. 2016. URL: http://www.journals.uchicago.edu/doi/pdfplus/10.1086/687257 (visited on 06/15/2025) (cit. on pp. 8, 39).

[33] Naftali Tishby, Fernando C. Pereira, and William Bialek. *The information bottleneck method*. Apr. 24, 2000. arXiv: physics/0004057 (cit. on p. 4).

[34] Francesca Zaffora Blando. "Bayesian merging of opinions and algorithmic randomness". In: *The British Journal for the Philosophy of Science* (July 19, 2022), p. 721758. ISSN: 0007-0882, 1464-3537 (cit. on pp. 19, 39).

[35] A. K. Zvonkin and L. A. Levin. "THE COMPLEXITY OF FINITE OBJECTS AND THE DEVELOPMENT OF THE CONCEPTS OF INFORMATION AND RANDOMNESS BY MEANS OF THE THEORY OF ALGORITHMS". In: *Russian Mathematical Surveys* 25.6 (Dec. 31, 1970). Publisher: IOP Publishing, p. 83. ISSN: 0036-0279 (cit. on p. 10).